Nonparametric Inference for Big-But-Biased Data

Laura Borrajo López

PhD Thesis

University of A Coruña
2021

UNIVERSIDADE DA CORUÑA

Nonparametric Inference for Big-But-Biased Data

Laura Borrajo López

PhD Thesis

2021

PhD advisor:
Ricardo Cao Abad

Doctoral Programme in Statistics and Operations Research
Department of Mathematics
University of A Coruña

UNIVERSIDADE DA CORUÑA

UNIVERSIDADE DA CORUÑA

The undersigned certify that he is the advisor of the Doctoral Thesis entitled "Nonparametric Inference for Big-But-Biased Data", developed by Laura Borrajo López at the University of A Coruña (Department of Mathematics), as part of the interuniversity PhD program (UDC, USC and UVigo) of Statistics and Operational Research, and hereby gives his consent to the author to proceed with the thesis presentation and the subsequent defense.

El abajo firmante hace constar que es el director de la Tesis Doctoral titulada "Nonparametric Inference for Big-But-Biased Data", realizada por Laura Borrajo López 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.

O abaixo asinante fai constar que é o director da Tese de Doutoramento titulada "Nonparametric Inference for Big-But-Biased Data", desenvolta por Laura Borrajo López 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 súa presentación e posterior defensa.

A Coruña, June 14th, 2021.

Advisor:

Dr. Ricardo Cao Abad

PhD student:

Laura Borrajo López

UNIVERSIDADE DA CORUÑA

The public defense of the Doctoral Thesis entitled "Nonparametric Inference for Big-But-Biased Data", developed by Laura Borrajo López and supervised by Dr. Ricardo Cao Abad, will be held on 10th September, 2021, at the Faculty of Computer Sciences at the University of A Coruña, with the examining committee:
Dr. Mario Francisco Fernández (President)
Dr. Catalina Bolancé Losilla (Secretary)
Dr. Mónica Fernández Bugallo (Board member)

A Coruña, September 10th, 2021.

PhD committee:

Dr. Mario Francisco Fernández

Dr. Catalina Bolancé Losilla

Dr. Mónica Fernández Bugallo

Advisor:

Dr. Ricardo Cao Abad

PhD student:

Laura Borrajo López

Agradecementos persoais

Esta tese é o resultado de cinco anos de traballo, o cal non tería saído adiante sen o apoio e colaboración, directa e indirecta, de moitas persoas.

En primeiro lugar, gustaríame agradecer o apoio recibido polo meu director de tese, Ricardo Cao, ao longo dos últimos anos. Grazas por darme a oportunidade de realizar esta tese e guiarme durante esta etapa, grazas pola constante dedicación e, sobre todo, polo gran entusiasmo e paixón que transmites pola investigación.

Gustaríame agradecer aos membros do tribunal do seminario de tese, Mario Francisco, Manuel Febrero e Mª Carmen Iglesias, polas súas aportacións e suxerencias para elaborar a versión final desta tese. Agradecer tamén a Sofia Olhede, por acollerme durante a miña estadía en Londres; a Swati Chandna, por ser un gran apoio durante a realización da mesma; a Alberto de Santos, pola oportunidade de colaborar con el e co seu equipo; aos profesores do programa de doutoramento, en especial a Silvia Lorenzo, quen me animou a realizar a tese; e a Luis Rodríguez, pola súa constante disposición, axuda e grata compaña no laboratorio.

Agradecer a confianza da miña familia ao longo desta etapa. En especial, grazas aos meus pais, Pili e Ramón, polo seu constante apoio e cariño e por ser un exemplo de superación e esforzo. Grazas aos meus irmáns, Marcos e Andrea, por estar sempre aí, en todos eses aspectos da vida cotiá que verdadeiramente importan. Non hai palabras suficientes para expresar o agradecida que estou e estarei sempre.

Por último, dar as grazas aos meus amigos, en especial a Marta e María, por escoitarme e ser o meu gran apoio nos momentos de maior frustración. Agradecer tamén a todas aquelas persoas que tiven o pracer de coñecer ao longo deste período e que formaron parte del; en especial, a Ana e Laura, polos bos consellos e a gran compañía e amizade, dentro e fóra do laboratorio, e a Pablo, Jorge, Pili e Patri, por acompañarme ao longo do proceso e polos bos recordos que me quedan del.

Institutional acknowledgments

The PhD student's research was sponsored by the Xunta de Galicia predoctoral grant (with reference ED481A-2016/367) for the universities of the Galician University System, public research organizations in Galicia and other entities of the Galician R&D&I System, whose funding comes from the European Social Fund (ESF) in 80% and in the remaining 20% from the General Secretary of Universities, belonging to the Ministry of Culture, Education and University Management of the Xunta de Galicia.

The doctoral student acknowledges partial support by MINECO Grant MTM2017-82724-R, by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015 and ED431C-2020-14, Centro Singular de Investigación de Galicia ED431G/01 and Centro de Investigación del Sistema Universitario de Galicia ED431G 2019/01), all of them through the European Regional Development Fund (ERDF), and by Vodafone Campus Lab grant.

The author is grateful to Guillermo Leira, responsible of environmental quality in the municipality of A Coruña (Galicia, NW Spain) for providing the air quality data set and would like to thank Vodafone ES for their collaboration and, especially, Alberto de Santos and the Big Data and Analytics team of this company for suggesting the customer churn problem and providing the data set used in the case study. The author would like to thank the Centro de Coordinación de Alertas y Emergencias Sanitarias (CCAES) of the Spanish Ministry for Health for providing the COVID-19 datasets used to study the age at infection. Thanks are also due to the Centro Nacional de Epidemiología (CNE), responsible of the sero-epidemiologic study for COVID-19 carried out in Spain (ENECOVID), whose staff helped the author to understand the insights of that study in the real data analysis considered in this thesis.

Á memoria da miña avoa Carme
e do meu avó Odilo

Abstract

It is often believed that in a Big Data context, given the large amount of data available, the data reflect precisely the underlying population. However, the data are often strongly biased due to the procedure used for obtaining them.

In order to reduce the significant bias that may appear in Big Data (Big-but-Biased Data, B3D), different testing methods for bias detection are used and completely nonparametric estimation methods for bias correction are proposed. Nonparametric estimators for the mean of a transformation of the random variable of interest are considered. When ignoring the biasing weight function, two different setups are proposed. In Setup 1 a small-sized simple random sample of the real population is assumed to be additionally observed, while in Setup 2 it is assumed that a twice biased sample of small size is observed. The asymptotic properties of the proposed estimators are extensively studied under suitable limit conditions on the small and the large sample sizes and standard and non-standard asymptotic conditions on the two bandwidths. The performance of the proposed nonparametric estimators is compared with the classical estimators based on the two samples involved in each setup through Monte Carlo simulation studies. Simulation results show that the new mean estimators outperform the classical empirical means for suitable choices of the two smoothing parameters involved. The influence of these smoothing parameters on the performance of the final estimators is also studied, exhibiting a striking limit behaviour of their optimal values. In addition, bootstrap bandwidth selection methods for each nonparametric mean estimator are introduced. Finally, the proposed techniques are applied to several real data sets from different areas.

Resumen

Se acostumbra a pensar que en un contexto de datos de gran volumen, el conjunto de datos refleja fielmente la población objeto de estudio, dada la gran cantidad de datos disponible. No obstante, en ocasiones estos datos están fuertemente sesgados debido, por lo general, al procedimiento de obtención de los mismos.

Con el objetivo de reducir el importante sesgo que puede aparecer en un contexto de datos de gran volumen, se propone el uso de métodos de contraste para la detección de sesgo y se desarrollan métodos de estimación para la corrección del mismo. Se consideran estimadores no paramétricos de la media de una transformación de la variable aleatoria de interés. Se proponen dos escenarios diferentes para abordar el problema de la estimación cuando la función peso que produce el sesgo es desconocida. En el escenario 1, se supone que se observa adicionalmente una muestra aleatoria simple de tamaño pequeño de la población verdadera, mientras que en el escenario 2 se asume que se observa una muestra de tamaño pequeño doblemente sesgada. Las propiedades asintóticas de los estimadores propuestos se estudian ampliamente bajo condiciones límite adecuadas en los tamaños muestrales y bajo condiciones asintóticas estándar y no estándar en los dos parámetros de suavizado. El comportamiento de los estimadores no paramétricos propuestos se compara con el de los estimadores clásicos basados en las dos muestras involucradas en cada escenario a través de estudios de simulación de Monte Carlo. Los resultados de la simulación muestran que los nuevos estimadores de la media mejoran a las medias empíricas clásicas para una elección adecuada de los dos parámetros de suavizado implicados. También se estudia la influencia de los parámetros de suavizado en el funcionamiento de los estimadores, los cuales exhiben un comportamiento límite llamativo en cuanto a sus valores óptimos. Además, se introducen métodos bootstrap para la selección automática de los parámetros de suavizado para cada estimador no paramétrico de la media. Finalmente, las técnicas propuestas se aplican a varios conjuntos de datos reales procedentes de diversas áreas.

Resumo

Adóitase pensar que nun contexto de datos de gran volume, o conxunto de datos reflicte fielmente a poboación obxecto de estudo, dada a gran cantidade de datos dos que se dispoñen. Non obstante, en moitas ocasións estes datos están fortemente nesgados debido, polo xeral, ao procedemento de obtención dos mesmos.

Co obxectivo de reducir o importante nesgo que pode aparecer nun contexto de datos de gran volume, proponse o uso de métodos de contraste para a detección do nesgo e desenvólvense métodos de estimación para a corrección do mesmo. Considéranse estimadores non paramétricos para a media dunha transformación da variable aleatoria de interese. Propóñense dous escenarios diferentes para abordar o problema da estimación cando a función peso que produce o nesgo é descoñecida. No escenario 1, suponse que se observa adicionalmente unha mostra aleatoria simple de tamaño pequeno da poboación verdadeira, mentres que no escenario 2 suponse que se observa unha mostra de tamaño pequeño dobremente nesgada. As propiedades asintóticas dos estimadores propostos son amplamente estudadas baixo condicións límite axeitadas sobre os tamaños mostrais e condicións asintóticas estándar e non estándar sobre os dous parámetros de suavizado. O comportamento dos estimadores non paramétricos propostos compárase co dos estimadores clásicos baseados nas dúas mostras implicadas en cada escenario por medio de estudos de simulación de Monte Carlo. Os resultados das simulacións amosan como os novos estimadores da media melloran ás medias empíricas clásicas para escollas axeitadas dos dous parámetros de suavizado implicados. Tamén se estuda a influencia dos parámetros de suavizado no funcionamento dos estimadores, amosando un comportamento límite sorprendente en canto os seus valores óptimos. Ademais, introdúcense métodos bootstrap para a selección automática dos parámetros de suavizado para cada estimador non paramétrico da media. Finalmente, as técnicas propostas aplícanse a varios conxuntos de datos reais procedentes de diversas áreas.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Methods 2

1.2.1 Setups 2

1.2.2 Nonparametric density estimation 3

1.2.3 Bias, variance and mean squared error 4

1.3 Content of the thesis 6

2 Mean estimation for Big-But-Biased Data **7**

2.1 Introduction 7

2.1.1 Sampling bias in big data 7

2.1.2 Related existing work 9

2.2 Basic inference in two different setups 10

2.2.1 Estimation procedure when the biasing function is known . . 10

2.2.2 Setup 1 11

2.2.3 Setup 2 12

3 Bias testing **15**

3.1 General bias detection 16

3.1.1 Kolmogorov-Smirnov test 16

3.1.2 Cramer-von Mises criterion 19

3.1.3 Mann–Whitney U test 22

3.2 Bias detection for mean estimation 23

3.3 Simulation 24

4 Nonparametric Estimation in Setup 1 **33**

4.1 Introduction 33

4.2 Estimation of the mean of a transformation for B3D 34

4.2.1 Sampling bias in big data 34

4.2.2 Density-based estimation in Setup 1 35
4.2.3 General weighted estimator 38
4.2.4 Asymptotic results 39
4.2.5 Further applicability 42
4.3 Simulations 42
4.3.1 Asymptotically-based simulations 46
4.4 Bootstrap algorithm 46
4.4.1 Simulations 47
4.5 Sketch of the proofs 49
4.5.1 Sketch of the proof of Theorem 4.2.1 49
4.5.2 Sketch of the proof of Theorem 4.2.2 55

5 Nonparametric Estimation in Setup 2 59
5.1 Introduction 59
5.2 Estimation for B3D in Setup 2 59
5.2.1 Asymptotic results 61
5.3 Simulations 63
5.3.1 Asymptotically-based simulations 65
5.4 Bootstrap algorithm 67
5.4.1 Simulations 69
5.5 Sketch of the proofs 70
5.5.1 Sketch of the proof of Theorem 5.2.1 70
5.5.2 Sketch of the proof of Theorem 5.2.2 76

6 Real data applications 81
6.1 Airline on-time performance 81
6.2 Air quality in smart cities 84
6.2.1 Motivation 84
6.2.2 Urban Air Pollution 85
6.2.3 Results 86
6.3 Vodafone 91
6.4 COVID-19 95
6.4.1 Setup 1 98
6.4.2 Setup 2 99

7 Conclusion and further research 103
7.1 Conclusion 103
7.2 Future work 104

A Appendix **105**

A.1 Proofs of the results in Chapter 4 105

A.1.1 Proof of Lemma 4.2.1 105

A.1.2 General weighted estimator in Subsection 4.2.3 107

A.1.3 Proof of Theorem 4.2.1 109

A.1.4 Proof of Theorem 4.2.2 146

A.2 Proofs of the results in Chapter 5 165

A.2.1 Proof of Theorem 5.2.1 165

A.2.2 Proof of Theorem 5.2.2 203

B Resumo en galego **225**

Bibliography **235**

Preface

This dissertation summarizes all the work developed during the PhD period. It mainly focuses on proposing different methods for detecting and correcting biases in a Big Data context. The proposed methodology is applied to various real data sets.

Chapter 1 introduces the context of Big-But-Biased data (B3D) in which this thesis is developed. It begins with a motivation based on different examples found in the literature. In Section 1.2, other works dealing with the problem of sampling bias are considered. In order to correct the bias present in a B3D sample, two different setups with additional information needed are proposed in Subsection 1.2.1. Since the study is performed in a nonparametric context and nonparametric density estimation becomes an important tool in the thesis, a review on this topic is included in Subsection 1.2.2.

Chapter 2 introduces the problem of mean estimation for Big-But-Biased data. Two different setups in which bias may be corrected are proposed in this chapter. In addition to the big and biased sample, in Setup 1 it is assumed that a simple random sample (SRS) of small size from the true population is observed; while in Setup 2, a twice biased sample is supposed to be observed. Different estimators for the mean are proposed in both setups, considering the unlikely case that the biasing weight function is known. The more realistic case of this function being unknown will be considered in Chapters 4 and 5, in which the general problem of estimating the mean of a transformation will also be addressed.

In Chapter 3, a procedure for bias testing is proposed. It consists of using two-sample existing methods to test the equality of distributions and the equality of means, but considering the distinctive feature that the ratio of both samples sizes does not tend to a constant, since the size of the B3D sample tends to infinity faster than that of the SRS. For the equality of distributions, the two-sample Kolmogorov-Smirnov test, the Cramer-von Mises criterion and the Mann-Whitney U-test are considered. In the case of testing the mean, the Welch's adaptation of the Student's t-test is used. This specific test is necessary since different distributions do not necessary imply different means. A comparative analysis between the different methods proposed is performed.

In Chapter 4, nonparametric estimation for a large-sized sample subject to sampling bias in Setup 1 is studied. The general parameter considered is the mean of a transformation of the random variable of interest. When ignoring the biasing weight function, a small-sized simple random sample of the real population is assumed to be additionally observed. A new nonparametric estimator that incorporates kernel density estimation is proposed. Asymptotic properties for this estimator are obtained under suitable limit conditions on the small and the large sample sizes and standard and non-standard asymptotic conditions on the two bandwidths. Explicit formulas are shown for the particular case of mean estimation. Simulation results show that the new mean estimator outperforms the classical empirical means of the two samples involved for suitable choices of the two smoothing parameters involved. The influence of two smoothing parameters on the performance of the final estimator is also studied, exhibiting a striking limit behaviour of their optimal values. A bootstrap algorithm is used to approximate the mean squared error of the proposed estimator. Its minimization leads to an automatic bandwidth selector.

Chapter 5 follows parallel lines to those of Chapter 4 but for Setup 2. The behavior of the nonparametric estimator proposed in this setup is analyzed under the standard and non-standard asymptotic conditions on the two smoothing parameters. The simulation study shows the good performance of the estimator under the non-standard conditions. A new bootstrap algorithm is used in this setup to approximate the mean squared error of the proposed estimator when a simple random sample of the true population is not observed.

In Chapter 6 the methods proposed in Chapters 4 and 5 are applied to several real data sets. Firstly, in Section 6.1, a data set concerning airline on-time performance of US flights is considered. It is a large data set with nearly 180 million records. The mean and the standard deviation of arrival delay of US flights in 2017 is estimated based on 2016 big-but-biased data, using the new approach. In Section 6.2 the issue of air pollution in smart cities is addressed. The estimation method for Setup 1 and the bootstrap algorithm are applied to a real data set concerning the levels of different pollutants in the urban air of the city of A Coruña (Galicia, NW

Spain). Estimations for the mean and the cumulative distribution function of the level of ozone and nitrogen dioxide when the temperature is greater than or equal to 30 °C based on 15 years of biased data are obtained. In Section 6.3, a real data set from the Telco Company Vodafone ES is considered. It consists of nearly 2.5 million records and 176 variables with information about Vodafone customers. A new variable, *index*, is constructed based on the 14 variables that best reflect the costumer's tendency to leave the company. The mean and the cumulative distribution function of that *index* is estimated based on the information of the target group and the universal control group for retention campaigns. Finally, in Section 6.4 the proposed methods in both setups are applied to the study of two COVID-19 data sets with information on asymptomatic, detected and hospitalized cases. Estimations of the mean age of people infected with COVID-19 are obtained when, in addition to the big-but-biased sample, a simple random sample of the true population or a doubly biased sample is observed.

Some comments about future work are given in Chapter 7. The possibility of extending the proposed methodology to categorical settings or to multidimensional variables, including covariate dependence in the biasing weight, is considered. Another idea is to apply the methods proposed to the estimation of the variance-covariance matrix and the correlation matrix, in order to perform principal component analysis and linear discriminant analysis.

In Appendix A the proofs for the theoretical results presented in Chapters 4 and 5 are collected.

Chapter 1

Introduction

1.1 Motivation

The sentence *with enough data, numbers speak for themselves* is often pronounced in this Big Data era. This reflects the doubtful notion that massive data sets always reflect objective and absolute truth. However, like any other human creation, data sets are not totally objective. Occasionally, a large sample is not completely representative of the population, but it is biased: Big-But-Biased Data (B3D). This includes partial vote counts on election nights, opinion polls carried out on social networks and large databases collected with self-selection mechanisms.

An interesting source of big-but-biased data is the StreetBump smartphone app mentioned by Crawford (2013). This app was created to help planning pothole patching in the city of Boston, where 20,000 of them are fixed every year. The app passively detects bumps by recording the accelerometers of the phone and GPS data while driving, instantly reporting them to the traffic department of the city. Thus, the city could plan their repair and the management of resources in the most efficient possible way. However, an important problem observed when using StreetBump was that people with lower income have a low rate of smartphone use. This rate is even lower for older residents, where smartphone penetration is as low as 16%. Therefore, these data provide a big but very biased sample of the population of potholes in Boston. As a consequence, the number of potholes in certain neighborhoods are underestimated, which causes a skewed management of resources.

The database of tweets generated by Hurricane Sandy is another interesting example cited by Crawford (2013). The data consists of more than 20 million tweets published between October 27 and November 1, 2012. A combined analysis of Twitter and Foursquare data produced some expected findings, such as an increase in grocery shopping the night before the storm, and other more surprising, such as an increase in nightlife the day after the hurricane. However, these data do not represent an unbiased sample of the population. It is well known that the greatest number of tweets about Sandy came from Manhattan. This was due to the high level of smartphone owners and Twitter use in New York. Not many messages were originated in the most affected areas by the catastrophe, since the lack of electricity caused many problems with internet access and many devices run out of battery in the hours after the storm.

In other examples, such as those cited in Hargittai (2015), survey data show that people do not select into the use of sites randomly; instead, use is biased in certain ways yielding samples that limit the generalizability of findings. Some of the problems coming from ignoring sampling bias in big data statistical analysis have been reported by Cao (2015). Calissano et al. (2018) also point out sampling bias problems for Twitter data sets.

1.2 Methods

Sampling bias is not a specific feature of big data. It has been widely considered in the statistical literature over the past few decades. Length bias problems were described in 1963 by C.R. Rao in the First International Symposium on Classical and Contagious Discrete Distribution (Patil & Rao, 1978). But preliminary ideas about sampling bias were already considered in the seminal paper by Fisher (1934) when studying albinism in genetics. In the nonparametric framework, Lloyd & Jones (2000) considered nonparametric density estimation for length biased data with unknown biasing weight function, whereas Cristóbal & Alcalá (2001) gave an overview of existing methods for dealing with length bias in nonparametric curve estimation.

In order to reduce the significant bias that may appear in Big Data, the main objective of this thesis is to develop estimation methods for bias correction as well as to adapt testing methods for bias detection. To solve the bias correction problem, there are at least two different setups we can think of.

1.2.1 Setups

Bias correction is considered in two possible setups, which are detailed in Chapters 4 and 5, respectively. Basically, in Setup 1 we assume that, in addition to the

B3D sample, a small-sized simple random sample (SRS) of the real population is observed. While in Setup 2, a twice biased sample of small size of the population is assumed to be additionally observed.

Setup 1 covers situations when the natural way of observing the variable of interest is biased; one can collect a big amount of data in that way; but there exists an alternative way to proceed (probably much more expensive or time-consuming) for which a simple random sample of the underlying population can be drawn. A real life example that can be covered by Setup 1 is estimating the pothole location density in the city of Boston using the StreetBump smartphone app data mentioned by Crawford (2013). The B3D sample of pothole locations coming from StreetBump exhibits an important sampling bias due to different smartphone penetration all over city areas and age groups. However, one could think of generating a simple random sample by just driving using randomly generated routes, inspecting at randomly chosen times and collecting locations of the potholes observed. Of course this procedure would be much more expensive and time-consuming, so it is reasonable that this simple random sample would be of a much smaller size.

Setup 2 is plausible when the sampling bias is coming from an acceptance/rejection method not controlled by the data collector. In such a case, the weigth function is proportional to the acceptance probability in the B3D sample. If the acceptance probability is assumed to remain the same but the second sample is obtained from a population already obtained after performing an acceptance/rejection procedure, then the second one is a simple random sample from a twice biased population. A motivating practical example for this setup refers to social network polls, where there is an important bias of acceptance/rejection nature: millions of people are offered to complete a survey but ‘only’ a few hundreds of thousands of persons answer the survey. If a second poll is performed within the list of users who already replied to the first one, and the acceptance probability function is assumed to remain the same, then the sample of people who also replied to the second survey is a twice biased sample as considered in Setup 2.

1.2.2 Nonparametric density estimation

The proposed estimators for bias correction proposed in Chapters 4 and 5 use nonparametric density estimation techniques.

Nonparametric density estimation has been one of the most studied fields in recent decades in statistics. The methods proposed in this area allow to analyze data without any prior parametric assumption about the distribution of the underlying variables.

Let $(X_1, X_2, \ldots, X_n)$ be an independent and identically distributed (i.i.d.) sample drawn from some unknown distribution function F, with density function f. If we are interested in estimating the density function, f, then the most common nonparametric estimator is the kernel density estimator proposed by Parzen (1962) and Rosenblatt (1956), which is:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \tag{1.1}$$

where $K_h(u) = (1/h)K(u/h)$, being K a kernel function and $h > 0$ the smoothing parameter, also known as bandwidth.

It is usual to request that the kernel function is non-negative and its integral is one:

$$K(u) \geq 0, \quad \forall u, \quad \int_{-\infty}^{\infty} K(u)du = 1,$$

and it is also frequent to demand that K is a symmetric function:

$$K(-u) = K(u).$$

Estimator (1.1) will be used in Chapters 4 and 5 to estimate the density functions involved in each setup. Below we will analyze the properties of this estimator, which are collected in Rosenblatt (1956), Parzen (1962) and Silverman (1986).

1.2.3 Bias, variance and mean squared error

The bias of the Parzen-Rosenblatt estimator (1.1) can be easily obtained:

$$\begin{aligned} Bias\left(\hat{f}_h(x)\right) &= E\left(\hat{f}_h(x)\right) - f(x) = \int \frac{1}{h}K\left(\frac{x-y}{h}\right) f(y)dy - f(x) \\ &= (K_h * f)(x) - f(x) = \frac{\mu_2(K)}{2}h^2 f''(x) + O(h^4), \end{aligned}$$

being $\mu_2(K) = \int t^2 K(t)dt$ the second central moment and $*$ the convolution operator, defined by:

$$(f * g)(x) = \int f(x-y)g(y)dy.$$

The variance can be handled similarly:

$$\begin{aligned}
Var\left(\hat{f}_h\left(x\right)\right) &= \frac{1}{nh^2}Var\left(K\left(\frac{x-X_1}{h}\right)\right)\\
&= \frac{1}{nh^2}\left[\int K\left(\frac{x-y}{h}\right)^2 f(y)dy-\left(\int K\left(\frac{x-y}{h}\right)f(y)dy\right)^2\right]\\
&= \frac{1}{n}\left[\left((K_h)^2 * f\right)(x)-((K_h * f)(x))^2\right],
\end{aligned}$$

being its asymptotic expression

$$Var\left(\hat{f}_h\left(x\right)\right)=\frac{\mu_0(K^2)}{nh}f(x)-\frac{1}{n}f(x)^2+O\left(\frac{h}{n}\right),$$

where $\mu_0(K^2)=\int K(t)^2dt$.

Consequently, the mean squared error (MSE) of the estimator is:

$$\begin{aligned}
MSE\left(\hat{f}_h\left(x\right)\right) &= E\left(\hat{f}_h\left(x\right)-f(x)\right)^2=Bias\left(\hat{f}_h\left(x\right)\right)^2+Var\left(\hat{f}_h\left(x\right)\right)\\
&= [(K_h * f)(x)-f(x)]^2+\frac{1}{n}\left[\left((K_h)^2 * f\right)(x)-((K_h * f)(x))^2\right]\\
&= AMSE\left(\hat{f}_h\left(x\right)\right)+O(h^6)+O\left(\frac{h}{n}\right),
\end{aligned}$$

where the asymptotic expression is:

$$AMSE\left(\hat{f}_h\left(x\right)\right)=\frac{\mu_2(K)^2}{4}h^4f''(x)^2+\frac{\mu_0(K^2)}{nh}f(x)-\frac{1}{n}f(x)^2.$$

A global measure of the error made when using the estimator is the mean integrated squared error (MISE):

$$\begin{aligned}
MISE\left(\hat{f}_h\left(x\right)\right) &= \int E\left[\left(\hat{f}_h\left(x\right)-f(x)\right)^2\right]dx=\int MSE\left(\hat{f}_h\left(x\right)\right)dx\\
&= \int\left[(K_h * f)(x)-f(x)\right]^2dx\\
&+ \frac{1}{n}\int\left[\left((K_h)^2 * f\right)(x)-((K_h * f)(x))^2\right]dx\\
&= AMISE\left(\hat{f}_h\left(x\right)\right)+O(h^6)+O\left(\frac{h}{n}\right),
\end{aligned}$$

where the asymptotic expression is:

$$AMISE\left(\hat{f}_h\left(x\right)\right) = \frac{\mu_2(K)^2}{4}h^4\int f''(x)^2dx+\frac{\mu_0(K^2)}{nh}-\frac{1}{n}\int f(x)^2dx. \quad (1.2)$$

In equation (1.2) the negative effect of choosing too large or too small bandwidths is observed.

It is easy to obtain the optimal smoothing parameter that minimizes the AMISE:

$$h_{AMISE} = \left(\frac{\mu_0(K^2)}{\mu_2(K)^2 \int f''(x)^2 dx} \right)^{1/5} n^{-1/5}.$$

1.3 Content of the thesis

The rest of the thesis is organized as follows. In Chapter 2, two different setups in which bias can be corrected are introduced. In Chapter 3 a comparative analysis of several existing methods for bias detection is performed. These are nonparametric tests of the equality of distributions and the equality of means, which include the distinctive feature assumed for the asymptotics under the B3D context. The nonparametric method for bias correction in Setup 1 is presented in Chapter 4. The general parameter considered is the mean of a transformation of the random variable of interest. A new nonparametric estimator that incorporates kernel density estimation is deeply studied, obtaining its asymptotic properties under standard and non-standard conditions. Simulation results show the good performance of the proposed estimator. Moreover, a bootstrap bandwidth selection method is also proposed. In Chapter 5 the nonparametric estimator for bias correction in Setup 2 is presented and studied in an analogous way to Chapter 4. In Chapter 6 the methods proposed in Chapters 4 and 5 are applied to several real data sets. The proofs of the theoretical results in Chapters 4 and 5 are included in Appendices A.1 and A.2.

Chapter 2

Mean estimation for Big-But-Biased Data

2.1 Introduction

The present chapter focuses on a rather simple but fundamental problem in the context of big-but-biased data, the estimation of the mean of a continuous population. For this purpose, two different setups will be considered.

In particular, this chapter deals with the unrealistic case where the biasing function is known, which has been presented by Cao & Borrajo (2018). To the best of our knowledge, at that time, there was no existing published work considering sampling bias when the sample size is very large (large-sized biased data or big-but-biased data).

In Chapters 4 and 5 this study will be extended to the general case of nonparametric estimation of the mean of a transformation of a continuous population when the biasing function is unknown. This includes as special cases the mean of the population and any other moment, the cumulative distribution function at a given point and also the characteristic function evaluated at a given value.

2.1.1 Sampling bias in big data

In many practical situations, e.g. in machine learning, a large amount of data can be collected but the sampling mechanism cannot be controlled by the data scientist. As a consequence, although the sample size can be very large, the distribution of this sample needs not to be the same as the one of the population of interest. In this section this idea is formalized mathematically.

Let us consider a continuous population with cumulative distribution function F (density f) and let us denote by

$$\mathbf{X} = (X_1, \ldots, X_n)$$

a simple random sample of size n from this population. Let us assume that we are not able to observe this sample but we observe, instead, another sample,

$$\mathbf{Y} = (Y_1, \ldots, Y_N)$$

of a much larger size ($N >> n$) from another distribution G (called the biased distribution, with density g) different from F but with a common support, $\mathcal{D}$. This condition can be formulated assuming a positive biasing function, $w(x), \forall x \in \mathcal{D}$, such that

$$g(x) = w(x)f(x) \ \forall x \in \mathcal{D}. \tag{2.1}$$

Equation (2.1) implies that $w(x)$ is the likelihood ratio or importance function.

Remark 2.1.1. *More general models to (2.1) have been considered by Vardi (1985) and Gill et al. (1988), but for the case where the biasing weight function $w(x)$ is known. These papers consider conditions on the functions w, f and g for the model to be identifiable and for f to be estimable. As these authors point out, there is no hope to estimate f off the support of g. So it is clear that the support of f has to be included in the support of g, i.e., $w(x) > 0$ for every x in the support of f.*

If the sampling bias is caused by an acceptance sampling method, the function w is, up to a constant factor, the probability of acceptance in the sampling mechanism. This sample mechanism is useful when sampling from a density f is difficult or time-consuming but $f(x) \leq cg(x)$ for some constant $c > 0$ and an easy-to-sample density g. In fact, if we consider the following acceptance sampling: draw X from f and keep $X = Y$ with probability $\pi(X)$, otherwise draw X again; the density function g becomes

$$g(x) = \frac{\pi(x)f(x)}{\int_{\mathcal{D}} \pi(y)f(y)dy} \ \forall x \in \mathcal{D},$$

and the biasing function is

$$w(x) = \frac{\pi(x)}{\int_{\mathcal{D}} \pi(y)f(y)dy} \ \forall x \in \mathcal{D}. \tag{2.2}$$

If we consider two proportional acceptance probabilities $\pi_1(x) = d\pi_2(x) \ \forall x \in \mathcal{D}$ for some $d > 0$ in an acceptance sampling, then (2.2) shows that the two biasing functions are the same, i.e., $w_1(x) = w_2(x) \ \forall x \in \mathcal{D}$.

2.1.2 Related existing work

A fundamental relationship within this chapter is Equation (2.1) in Section 2.2 below. It relates the density of the underlying population, f, with the density of the biased population, g, via some biasing function, w. This equation is present in many statistical papers, although most of them are not related to sampling bias nor to big data.

Equation (2.1) plays an important role in survey sampling, when defining calibration estimators. The paper by Kott (2016) contains an overview on calibration weighting in survey sampling. Just to give two examples, optimal calibration estimators have been proposed by Deville & Särndal (1992), while nonparametric model calibration estimation has been considered by Montanari & Ranalli (2005). For calibration in survey sampling, the biasing function, w, appearing in Equation (2.1), is typically known and, in fact, often chosen by the researcher. This is, of course, not the case in B3D. On the other hand the finite population in survey sampling implies a discrete nature of w, while the biasing function in our setting is defined on a real interval.

Equation (2.1) also appears in acceptance-rejection methods in simulation (see, for instance, the book by Devroye (1986)). In that case, the density g is difficult to simulate and the density f is easy to simulate. Then w is directly related to the acceptance probability in the simulation algorithm, which determines its efficiency. Although there are no inference issues concerning acceptance-rejection algorithms in simulation, this method motivates natural acceptance-rejection sampling procedures as a possible source for sampling bias, also in big data.

Relation (2.1) is also present when random sampling is too difficult or too costly and the probability density function, f, is distorted by some multiplicative nonnegative weight function, w, often specified up to a finite number of parameters. Estimation of these parameters is an important problem that has been considered by Ma et al. (2005) for generalized skew-elliptical multivariate distributions. Semiparametric estimation methods when a representative sample of the population is unavailable due to selection bias, have been studied by Genton et al. (2012) and Ma et al. (2013), among other. These authors proved asymptotic properties for the estimators of location parameters when the underlying (unobservable) distribution is symmetric.

A very special case, related to the previous situation when random sampling is too difficult, appears when sampling can only be carried out in a paired way and only one of the components of the pair can be observed. This is the case when the random variable, X, cannot be observed but we are able to observe $Y = \max\{X_1, X_2\}$ the maximum of two independent random copies, X_1 and X_2, of X. In this case the distribution of Y becomes $G(x) = F(x)^2$, where F is the distribution function of X. This gives $g(x) = 2F(x)f(x)$, i.e. $w(x) = 2F(x)$. In this very particular setting, the function w depends on f and estimators for μ can be based on the empirical cdf estimator for G.

The problem is introduced in Subsection 2.2.1, in which two unrealistic estimators for the known w case are proposed. Two possible setups for bias correction are considered in Subsections 2.2.2 and 2.2.3.

2.2 Basic inference in two different setups

We focus on the problem of estimating the mean of a continuous random variable, $\mu = \int xf(x)dx$, in a B3D context, i.e. using a sample of a large size generated from a distribution which is not the one we are interested in, but some biased version of it. In the next subsection we look at some unrealistic version of the estimation procedure that is only feasible when the bias function, $w(x)$, is known.

2.2.1 Estimation procedure when the biasing function is known

In the B3D context presented in Subsection 2.1.1, it is clear that, if the sample $\mathbf{X}$ were observed, a classical estimator for μ would be available: the X-sample mean,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

However, if $\mathbf{X}$ is not available and we only observe $\mathbf{Y}$, it is reasonable to use relation (2.1) in order to estimate μ. In fact, using (2.1), the following equation holds:

$$E\left(\frac{Y}{w(Y)}\right) = \int \frac{y}{w(y)} g(y)\,dy = \int yf(y)\,dy = \mu. \tag{2.3}$$

Equation (2.3) motivates the definition of an unrealistic estimator which can be only used in practice when the function w is known (Cao & Borrajo, 2018):

$$\tilde{\mu}^{(1)} = \frac{1}{N}\sum_{i=1}^{N} \frac{Y_i}{w(Y_i)} = \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)\,Y_i}{g(Y_i)}, \tag{2.4}$$

as it would happen, for instance, in acceptance sampling. Since $\tilde{\mu}^{(1)}$ is the sample mean of the simple random sample $Z_i = Y_i/w(Y_i),\ i = 1, \ldots, N$, its properties as a good estimator of μ are straightforward:

$$E\left(\tilde{\mu}^{(1)}\right) = \mu,$$

$$Var\left(\tilde{\mu}^{(1)}\right) = \sigma_Z^2/N$$

and

$$\sqrt{N}\left(\tilde{\mu}^{(1)} - \mu\right)/\sigma_Z \rightarrow N(0,1),$$

where $\sigma_Z^2 = \int x^2 f(x)^2 g(x)^{-1} dx - \mu^2$.

The estimator in (2.4) is a weighted average of the Y-sample, but the sample weights $f\left(Y_i\right)/g\left(Y_i\right)$ do not sum up to 1. So a reasonable modification of $\tilde{\mu}^{(1)}$ is the following convex linear combination version:

$$\tilde{\mu}^{(2)} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{w\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w\left(Y_i\right)}} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)Y_i}{g\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{g\left(Y_i\right)}}. \tag{2.5}$$

The estimator in (2.5) can also be regarded as an empirical analogue of the expectation ratio

$$\frac{E\left(\frac{Y}{w(Y)}\right)}{E\left(\frac{1}{w(Y)}\right)},$$

which is equal to μ, by just recalling equation (2.3) and

$$E\left(\frac{1}{w\left(Y\right)}\right) \quad = \quad \int \frac{1}{w\left(y\right)} g\left(y\right) dy = \int f\left(y\right) dy = 1.$$

In general, the biasing function, w, is not known (because the underlying densities, f and g, are unknown), so the estimators in (2.4) and (2.5) are useless. Thinking in the more realistic unknown w case, which will be developed in Chapters 4 and 5, we can think of two different setups.

2.2.2 Setup 1

Since the Y-sample is available, it is certainly possible to estimate the density g and then plug it into (2.4) or (2.5). To end up with completely observable versions of $\tilde{\mu}^{(1)}$ and $\tilde{\mu}^{(2)}$, we also need to estimate the density f. This can be done if we are able to collect a simple random sample, $\mathbf{X}$, from the real population with density f.

Of course, when having the sample $\mathbf{X}$, it is certainly possible to estimate μ based on it. However, when the sample size of $\mathbf{X}$ is small, the quality of estimators based on it may be poor, while estimating μ using $\mathbf{Y}$ will have a much smaller variance (although some bias) due to its much larger sample size.

The formulation of this first setup is the following:

Setup 1. Let us assume that we observe the big data sample of a large size N,

$$\mathbf{Y} = (Y_1, \ldots, Y_N),$$

from the biased distribution G (density g). Suppose that we also observe a simple random sample,

$$\mathbf{X} = (X_1, \ldots, X_n),$$

of a much smaller sample size n ($n << N$) of the real population F (density f).

The two resulting estimators for the unknown w case in Setup 1 will be studied in Chapter 4.

2.2.3 Setup 2

Another possible setup in which we can think of consists in assuming that the biasing mechanism can be replicated a second time for an already first biased population.

Setup 2. Let us suppose that we are in a B3D situation, assuming that we observe the big data sample of a large size N,

$$\mathbf{Y} = (Y_1, \ldots, Y_N),$$

from the biased distribution G (density g).

Suppose that we are not able to observe a simple random sample, $\mathbf{X}$, from the real population, but instead, we can mimic the biasing mechanism and apply it to the population G to get a biased sample of it,

$$\mathbf{Z} = (Z_1, \ldots, Z_n),$$

i.e. a simple random sample from a twice biased population, M (density m), of a much smaller sample size n ($n << N$).

To study the relation between Setup 1 and Setup 2, let us denote the biasing functions involved by $w_1(x) = g(x)/f(x)$ and $w_2(x) = m(x)/g(x)$. If the sampling bias is caused by an acceptance/rejection sampling method and we denote the probability of acceptance in the sampling mechanism by w_0, then the two biasing weight functions w_1 and w_2 are proportional to w_0:

$$w_1(x) = c_1 \cdot w_0(x),$$

$$w_2(x) = c_2 \cdot w_0(x).$$

Therefore, one of them, for example w_2, is proportional to the other, w_1, i.e., $w_2(x) = c \cdot w_1(x)$:

$$w_2(x) = c_2 \cdot \frac{w_1(x)}{c_1} = c \cdot w_1(x) \tag{2.6}$$

for some constant $c = \dfrac{c_2}{c_1}$.

Since

$$w_1(x) = \frac{g(x)}{f(x)} \Rightarrow g(x) = c_1 \cdot w_0(x) \cdot f(x) \Rightarrow 1 = \int g(x)dx = c_1 \int w_0(x)f(x)dx$$

and

$$w_2(x) = \frac{m(x)}{g(x)} \Rightarrow m(x) = c_2 \cdot w_0(x) \cdot g(x) \Rightarrow 1 = \int m(x)dx = c_2 \int w_0(x)g(x)dx,$$

we could express

$$c_1 = \left(\int w_0(x)f(x)dx \right)^{-1} \text{ and } c_2 = \left(\int w_0(x)g(x)dx \right)^{-1},$$

obtaining that the constant c is

$$c = \frac{c_2}{c_1} = \frac{\int w_0(x)f(x)dx}{\int w_0(x)g(x)dx}.$$

When the simple random sample, $\mathbf{X}$, from the true population is not available and we only observe $\mathbf{Y}$, it is reasonable to use the following relation:

$$m(x) = w_2(x)g(x) \ \forall x \in \mathcal{D} \tag{2.7}$$

in order to estimate μ.

In fact, using (2.7), the following equation holds:

$$\begin{aligned} E\left(\frac{Y}{w_2(Y)}\right) &= \int \frac{y}{w_2(y)} g(y)\, dy = \int \frac{y}{m(y)/g(y)} g(y)\, dy \\ &= \int \frac{y}{cg(y)/f(y)} g(y)\, dy = \frac{1}{c} \int y f(y)\, dy = \frac{1}{c}\mu. \end{aligned} \tag{2.8}$$

Equation (2.8) motivates the definition of an unrealistic estimator of $\frac{\mu}{c}$ which can be only used in practice when the function w_2 is known:

$$\tilde{\tilde{\mu}}^{(1)} = \frac{1}{N}\sum_{i=1}^{N} \frac{Y_i}{w_2\left(Y_i\right)} = \frac{1}{N}\sum_{i=1}^{N} \frac{g\left(Y_i\right)Y_i}{m\left(Y_i\right)} = \frac{1}{c}\cdot\frac{1}{N}\sum_{i=1}^{N} T_i. \tag{2.9}$$

Since $\tilde{\tilde{\mu}}^{(1)}$ is the sample mean of the simple random sample $T_i = Y_i/w_1(Y_i),\ i = 1,\ldots,N$, its properties as a good estimator of $\frac{\mu}{c}$ are straightforward:

$$E\left(\tilde{\tilde{\mu}}^{(1)}\right) = \frac{1}{c}\cdot\mu,$$

$$Var\left(\tilde{\tilde{\mu}}^{(1)}\right) = \frac{\sigma_T^2}{c^2\cdot N}$$

and

$$\sqrt{N}\quad\frac{\tilde{\tilde{\mu}}^{(1)} - \frac{\mu}{c}}{\frac{\sigma_T}{c}} \to N(0,1),$$

where $\sigma_T^2 = \int x^2 f(x)^2 g(x)^{-1} dx - \mu^2$.

The estimator in (2.9) is a weighted average of the Y-sample, but the sample weights $g\left(Y_i\right)/m\left(Y_i\right)$ do not sum up to 1. So a reasonable modification of $\tilde{\tilde{\mu}}^{(1)}$ is the following convex linear combination version:

$$\tilde{\tilde{\mu}}^{(2)} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{w_2\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_2\left(Y_i\right)}} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{g\left(Y_i\right)Y_i}{m\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N}\frac{g\left(Y_i\right)}{m\left(Y_i\right)}}. \tag{2.10}$$

The estimator in (2.10) can also be regarded as an empirical analogue of the expectation ratio

$$\frac{E\left(\frac{Y}{w_2(Y)}\right)}{E\left(\frac{1}{w_2(Y)}\right)},$$

which is equal to μ, by just recalling equation (2.8) and

$$E\left(\frac{1}{w_2\left(Y\right)}\right) \quad = \quad \int \frac{1}{w_2\left(y\right)} g\left(y\right)dy = \frac{1}{c}\int f\left(y\right)dy = \frac{1}{c}.$$

In general, as in Setup 1, the biasing function, w_2, is not known (because the underlying densities, g and m, are unknown), so the estimators in (2.9) and (2.10) are useless. Equations (2.9) and (2.10) require the estimation of the densities g and m (or the estimation of the biasing function, w_2) to obtain observable estimators for the population mean. This is possible if one is able to collect information on g and m or indirect information on w_2, which will be discused in Chapter 5.

Chapter 3

Bias testing

When working with a large database, a logical first step is to check if we are in a context of biased data. In order to detect if bias exist, we can use several existing methods that allow to test if the two distributions involved (F and G in Setup 1 and G and M in Setup 2) come from the same distribution (unbiased situation) or not (biased situation). This is tantamount to using tests for the null hypothesis:

$$H_0 : F = G$$

against the alternative

$$H_1 : F \neq G,$$

like, for instance, the Kolmogorov-Smirnov test or the Cramer-von Mises criterion.

For simplicity, in this chapter we will focus on Setup 1, in which we observe the simple random sample X, coming from the unbiased distribution F and the B3D sample Y from the biased distribution G. The adaptation to Setup 2 of the following methods is straightforward. It would be enough to consider the sample Z from the twice biased distribution M, instead of the sample X.

Despite the fact that the following tests for bias detection are widely known methods, it is important to consider the distinctive feature of our B3D context: we will assume that the ratio of both samples sizes involved does not tend to a constant but to infinity, i.e., the size of the B3D sample tends to infinity faster than that of the SRS.

3.1 General bias detection

3.1.1 Kolmogorov-Smirnov test

The Kolmogorov–Smirnov test (KS test) proposed by Kolmogorov (1933) and Smirnov (1939) is a nonparametric test of equality of distributions used to compare the population distribution with a reference probability distribution

$$H_0 : F = F_0,$$

or to compare two populations and conclude if they have the same distribution

$$H_0 : F = G.$$

Let F_n be the empirical cumulative distribution function (ecdf) for the sample $(X_1, X_2, \dots, X_n)$ and G_N the ecdf for the sample $(Y_1, Y_2, \dots, Y_N)$, defined as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}, \tag{3.1}$$

$$G_N(x) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{Y_j \leq x\}}, \tag{3.2}$$

where **1** denotes de indicator function.

It is well known that F_n and G_N are the nonparametric maximum likelihood estimators of F and G, respectively. Therefore, the proximity of F_n and G_N will be indicative of the veracity of H_0, while a large distance between both ecdf will evidence that H_0 is probably false.

The two-sample Kolmogorov–Smirnov statistic quantifies the distance between the two ecdf involved:

$$D_{N,n} = \sup_{x \in \mathbb{R}} |G_N(x) - F_n(x)|,$$

where sup denotes the supremum over all $x \in \mathbb{R}$, while the one-sample test statistic computes the distance between the ecdf of the sample and the cdf of the reference distribution:

$$D_n^F = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|.$$

Many authors have studied the distribution of $D_n^{F_0}$ under the assumption that $F_0(x)$ is continuous. Kolmogorov (1933), Smirnov (1939), Feller (1948), Doob (1949) and Smirnov (1948) considered the limiting distribution of $D_n^{F_0}$ and Massey (1951) showed that the exact distribution of $D_n^{F_0}$ under H_0 is independent of F_0 if F_0 is continuous.

As it happens with the one-sample test statistic with continuous $F = G$ under the null hypothesis, it can be proven that the exact distribution of the two-sample test statistic does not depend on the distributions involved, it is distribution-free:

Proposition 3.1.1. *(Smirnov, 1939) The two-sample Kolmogorov-Smirnov test is a distribution-free test under H_0 if $F = G$ is continuous.*

Proof. Let us define the inverse of F by

$$F^{-1}(t) = \min\{x : F(x) \geq t\}.$$

Taking into account the change of variables $t = F(x)$ or $x = F^{-1}(t)$, we can write the statistic as

$$D_{N,n} = \sup_{x\in\mathbb{R}} |G_N(x) - F_n(x)| = \sup_{0<t<1} |G_N(F^{-1}(t)) - F_n(F^{-1}(t))|.$$

Using the definitions of the ecdfs (3.1) and (3.2), under the null hypothesis H_0, we obtain:

$$\begin{aligned} F_n(F^{-1}(t)) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i\leq F^{-1}(t)\}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{F(X_i)\leq t\}}, \\ G_N(F^{-1}(t)) &= \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{Y_j\leq F^{-1}(t)\}} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{F(Y_j)\leq t\}}, \end{aligned}$$

and therefore,

$$\sup_{0<t<1} |G_N(F^{-1}(t)) - F_n(F^{-1}(t))| = \sup_{0<t<1} \left| \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{F(Y_j)\leq t\}} - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{F(X_i)\leq t\}} \right|.$$

The distributions of $F(X_i)$ and $F(Y_j)$ are uniform on the interval $[0,1]$ since

$$P(F(X_i) \leq t) \;=\; P(X_i \leq F^{-1}(t)) = F(F^{-1}(t)) = t$$

and

$$P(F(Y_j) \leq t) \;=\; P(Y_j \leq F^{-1}(t)) = F(F^{-1}(t)) = t.$$

Therefore, the random variables $U_i = F(X_i), i = 1, \ldots, n$ and $V_j = F(Y_j), j = 1, \ldots, N$ are independent and have uniform distribution on $[0,1]$, so:

$$D_{N,n} = \sup_{x\in\mathbb{R}} |G_N(x) - F_n(x)| = \sup_{0<t<1} \left| \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{V_j\leq t\}} - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{U_i\leq t\}} \right|,$$

which clearly does not depend on F. □

Smirnov (1939) proved that, for n and N sufficiently large, under H_0 the statistic

$$\sqrt{\frac{N \cdot n}{N+n}} D_{N,n}$$

has the same asymptotic distribution that the Kolmogorov distribution:

$$\begin{aligned} P\left(\sqrt{\frac{N \cdot n}{N+n}} D_{N,n} \leq t\right) \xrightarrow{d} K(t) &= 1 - 2\sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2t^2} \\ &= \frac{\sqrt{2\pi}}{t} \sum_{i=1}^{\infty} e^{-(2i-1)^2\pi^2/(8t^2)}, \end{aligned}$$

where $K(t)$ denotes the Kolmogorov-Smirnov cdf.

Let's see now what happens when considering the distinctive feature of our B3D context, i.e., $N/n \to \infty$:

Proposition 3.1.2. *Under H_0, when $N/n \to \infty$, the statistic $\sqrt{\frac{N \cdot n}{N+n}} D_{N,n}$ has the same asymptotic distribution that the statistics $\sqrt{\frac{Nn}{N+n}} D_n^F$ and $\sqrt{n} D_n^F$ under $H_0 : F = F_0$.*

Proof. On the one hand, under the null hypothesis H_0, and defining

$$D_N^G = \sup_{x \in \mathbb{R}} |G_N(x) - G(x)|$$

and

$$D_n^F = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|,$$

we obtain that:

$$\begin{aligned} \sqrt{\frac{N \cdot n}{N+n}} D_{N,n} &= \sqrt{\frac{N \cdot n}{N+n}} \sup_{x \in \mathbb{R}} |G_N(x) - G(x) + F(x) - F_n(x)| \\ &\leq \sqrt{\frac{N \cdot n}{N+n}} \sup_{x \in \mathbb{R}} |G_N(x) - G(x)| + \sqrt{\frac{N \cdot n}{N+n}} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \\ &= \sqrt{\frac{n}{N+n}} \sqrt{N} D_N^G + \sqrt{\frac{N}{N+n}} \sqrt{n} D_n^F \xrightarrow{d} K, \qquad (3.3) \end{aligned}$$

since when $N/n \to \infty$:

$$\sqrt{\frac{n}{N+n}} \approx \sqrt{\frac{n}{N}} = o(1),$$

$$\sqrt{N} D_N^G \xrightarrow{d} K,$$

$$\sqrt{\frac{N}{N+n}} \approx \sqrt{\frac{N}{N}} = 1$$

and

$$\sqrt{n}D_n^F \xrightarrow{d} K.$$

On the other hand, under the null hypothesis H_0, the one-sample test statistic, D_n^F, satisfies:

$$\begin{aligned}\sqrt{n}D_n^F &= \sqrt{n}\sup_{x\in\mathbb{R}}|F_n(x)-F(x)| = \sqrt{n}\sup_{x\in\mathbb{R}}|F_n(x)-G_N(x)+G_N(x)-G(x)| \\ &\leq \sqrt{n}D_{N,n}+\sqrt{n}D_N^G = \sqrt{n}D_{N,n}+\sqrt{\frac{n}{N}}\sqrt{N}D_N^G,\end{aligned}$$

which implies that

$$\sqrt{\frac{N}{N+n}}\sqrt{n}D_n^F \leq \sqrt{\frac{Nn}{N+n}}D_{N,n}+\sqrt{\frac{n}{N+n}}\sqrt{N}D_N^G$$

and therefore

$$\sqrt{\frac{N}{N+n}}\sqrt{n}D_n^F - \sqrt{\frac{n}{N+n}}\sqrt{N}D_N^G \leq \sqrt{\frac{Nn}{N+n}}D_{N,n}. \tag{3.4}$$

Considering (3.3) and (3.4), we obtain:

$$\sqrt{\frac{N}{N+n}}\sqrt{n}D_n^F - \sqrt{\frac{n}{N+n}}\sqrt{N}D_N^G \leq \sqrt{\frac{N\cdot n}{N+n}}D_{N,n} \leq \sqrt{\frac{N}{N+n}}\sqrt{n}D_n^F + \sqrt{\frac{n}{N+n}}\sqrt{N}D_N^G$$

and since $\sqrt{\frac{n}{N+n}}\sqrt{N}D_N^G \approx o_p(1)$, we conclude that the asymptotic distribution of $\sqrt{\frac{N\cdot n}{N+n}}D_{N,n}$ is the same as that of $\sqrt{\frac{N}{N+n}}\sqrt{n}D_n^F$, which, in our particular situation of the ratio of the two sample sizes tending to infinity, is the same as the asymptotic distribution of $\sqrt{n}D_n^F$. □

As a consequence, when $N/n \to \infty$, the two-sample statistic, $D_{N,n}$, will be used but callibrating it with the asymptotic distribution of the one-sample test.

Apart from this test, we could also consider other tests or criterions. In Subsections 3.1.2 and 3.1.3 the Cramer-von Mises criterion and the Mann-Whitney U test will be used for bias detection.

3.1.2 Cramer-von Mises criterion

The Cramer-von Mises criterion (Cramer, 1928; von Mises, 1928), like the Kolmogorov-Smirnov test, is used to judge the goodness of fit of a theoretical cumulative distribution function, F_0, compared to a given empirical distribution function, F_n. As F is unknown, the two following statistics could be used:

$$\tilde{\omega}^2 = \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_n(x),$$

$$\tilde{\omega}^2 = \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x)$$

For comparing two empirical distributions, the generalization to the two-sample case is given by Anderson (1962):

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - G_N(x)]^2 dH_{N+n}(x)$$

which compares the two empirical cdf. In our context, F_n and G_N denote the empirical distribution functions of the SRS and the B3D sample respectively, being H_{N+n} the empirical distribution function of the two samples together, i.e., $(N+n)H_{N+n}(x) = nF_n(x) + NG_N(x)$.

The statistic for the one-sample case is

$$T_n = n\tilde{\omega}^2 = n\int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x) = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - F_0(x_i)\right]^2$$

and for the two-sample case:

$$T_{N,n} = \frac{Nn}{N+n}\omega^2 = \frac{Nn}{N+n}\int_{-\infty}^{\infty} [F_n(x) - G_N(x)]^2 dH_{N+n}(x) = \frac{V}{Nn(N+n)} - \frac{4Nn-1}{6(N+n)},$$

where V is defined by

$$V = n\sum_{i=1}^{n}(r_i - i)^2 + N\sum_{j=1}^{N}(s_j - j)^2$$

being $r_i, i = 1, 2, \ldots, n$, the ranks of the SRS in the combined sample and $s_j, j = 1, 2, \ldots, N$, the ranks of the B3D sample in the combined one.

In Rosenblatt (1952) it has been proved that, under the null hypotheses, $T_{N,n}$ has the same limiting distribution as T_n when $n \to \infty$, $N \to \infty$ and $N/n \to \lambda$, being λ a positive constant. For moderate sample sizes, the limiting distribution is a good approximation to the exact distribution.

Proposition 3.1.3. *Under H_0, when $N/n \to \infty$, the statistic $T_{N,n}$ has the same asymptotic distribution that the statistic T_n under $H_0 : F = F_0$.*

Proof. Defining

$$T_N^G = N \int_{-\infty}^{\infty} [G_N(x) - G(x)]^2 dG(x),$$

$$T_n^F = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)$$

and

$$T_{N,n}^F = \frac{Nn}{N+n} \int_{-\infty}^{\infty} [F_n(x) - G_N(x)]^2 dF(x),$$

it can be proven that $T_{N,n}$ has the same asymptotic distribution that $T_{N,n}^F$.

On the one hand, under the null hypothesis H_0 and using the triangular inequality, we obtain that:

$$\begin{aligned} &\left(\int_{-\infty}^{\infty} [F_n(x) - F(x) + G(x) - G_N(x)]^2 dF(x)\right)^{1/2} \\ &\quad\leq \left(\int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)\right)^{1/2} + \left(\int_{-\infty}^{\infty} [G_N(x) - G(x)]^2 dG(x)\right)^{1/2}, \end{aligned}$$

which implies that

$$\begin{aligned} T_{N,n}^F &= \frac{Nn}{N+n} \int_{-\infty}^{\infty} [F_n(x) - F(x) + G(x) - G_N(x)]^2 dF(x) \\ &\leq \frac{Nn}{N+n} \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x) + \frac{Nn}{N+n} \int_{-\infty}^{\infty} [G_N(x) - G(x)]^2 dG(x) \quad (3.5) \\ &+ \frac{2Nn}{N+n} \left(\int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)\right)^{1/2} \left(\int_{-\infty}^{\infty} [G_N(x) - G(x)]^2 dG(x)\right)^{1/2}. \end{aligned}$$

On the other hand, under the null hypothesis H_0 and using again the triangular inequality, we obtain that:

$$\begin{aligned} &\left(\int_{-\infty}^{\infty} [F_n(x) - G_N(x) + G_N(x) - G(x)]^2 dF(x)\right)^{1/2} \\ &\quad\leq \left(\int_{-\infty}^{\infty} [F_n(x) - G_N(x)]^2 dF(x)\right)^{1/2} + \left(\int_{-\infty}^{\infty} [G_N(x) - G(x)]^2 dG(x)\right)^{1/2}, \end{aligned}$$

then, the one-sample test statistic, T_n^F, satisfies:

$$\begin{aligned} T_n^F &= n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x) = n \int_{-\infty}^{\infty} [F_n(x) - G_N(x) + G_N(x) - G(x)]^2 dF(x) \\ &\leq n \int_{-\infty}^{\infty} [F_n(x) - G_N(x)]^2 dF(x) + n \int_{-\infty}^{\infty} [G_N(x) - G(x)]^2 dG(x) \\ &+ 2n \left(\int_{-\infty}^{\infty} [F_n(x) - G_N(x)]^2 dF(x)\right)^{1/2} \left(\int_{-\infty}^{\infty} [G_N(x) - G(x)]^2 dG(x)\right)^{1/2}, \end{aligned}$$

which implies that

$$\frac{N}{N+n}T_n^F - \frac{n}{N+n}T_N^G - \frac{2\sqrt{Nn}}{N+n}\left(T_n^F\right)^{1/2}\left(T_N^G\right)^{1/2} \le T_{N,n}^F. \tag{3.6}$$

Considering (3.5) and (3.6), we obtain:

$$\begin{aligned}\frac{N}{N+n}T_n^F - \frac{n}{N+n}T_N^G - \frac{2\sqrt{Nn}}{N+n}\left(T_n^F\right)^{1/2}\left(T_N^G\right)^{1/2} &\le T_{N,n}^F \\ \le \quad \frac{N}{N+n}T_n^F + \frac{n}{N+n}T_N^G + \frac{2\sqrt{Nn}}{N+n}\left(T_n^F\right)^{1/2}\left(T_N^G\right)^{1/2}\end{aligned}$$

and since when $N/n \to \infty$:

$$\frac{n}{N+n} \approx \frac{n}{N} = o(1),$$

$$\frac{\sqrt{Nn}}{N+n} \approx \frac{\sqrt{n}}{\sqrt{N+n}} = o(1),$$

$$T_n^F = O_P(1)$$

and

$$T_N^G = O_P(1),$$

we conclude that the asymptotic distribution of $T_{N,n}^F$ is the same as that of $\frac{N}{N+n}T_n^F$, which, in our particular situation of the ratio of the two sample sizes tending to infinity, is the same as the asymptotic distribution of T_n^F, since

$$\frac{N}{N+n} \approx \frac{N}{N} = 1.$$

□

3.1.3 Mann–Whitney U test

The Mann–Whitney U test, also called the Mann–Whitney–Wilcoxon (MWW) or Wilcoxon rank-sum test (Wilcoxon, 1945; Mann & Whitney, 1947), is a nonparametric test of the null hypothesis that, for randomly selected values X and Y from two populations, the probability of X being greater than Y is equal to the probability of Y being greater than X:

$$H_0 : P(X > Y) = P(Y > X).$$

The U test is weaker than that of Kolmogorov-Smirnov, since it does not test the equality of distributions, but a condition that is verified in that case.

To compute the statistic, U, the two combined samples are ranked and each of the values of the two samples is assigned to its rank (i.e., rank 1 is assigned to the smallest observation, rank 2 to the second smallest observation, and so on). If two or more observations are equal, the mean rank is assigned to the two observations. Finally, R_X and R_Y, the adjusted rank-sums, i.e. the sum of the ranks in each of the samples X and Y, respectively), are computed. This allows to construct:

$$U_X = Nn + \frac{n(n+1)}{2} - R_X$$

$$U_Y = Nn + \frac{N(N+1)}{2} - R_Y.$$

Knowing that

$$R_X + R_Y = \frac{(N+n)(N+n+1)}{2},$$

the sum of two values is given by

$$U_X + U_Y = Nn.$$

The U statistic is defined as the minimum between U_X and U_Y:

$$U = \min\{U_X, U_Y\}.$$

For large sized samples, U is approximately normally distributed under the null hypothesis. In that case, the standardized value is

$$z = \frac{U - m_U}{\sigma_U} \to N(0,1),$$

where m_U and σ_U are the mean and the standard deviation of U, under H_0, which are given by

$$m_U = \frac{Nn}{2} \quad \text{and} \quad \sigma_U = \sqrt{\frac{Nn(N+n+1)}{12}}.$$

In the note written by Kasuya (2001), the author warns about using the Mann-Whitney U-test when the distributions of the two samples are very different, since it can lead to a misinterpretation of the results. In that case, it is recommended to use the unequal variances version of the t-test (Welch's t-test), which provides more reliable results.

3.2 Bias detection for mean estimation

Since the non-equality of distributions does not have to imply the non-equality of means, it would be reasonable to use a specific test for the equality of means when addressing a mean estimation problem.

To see the effect of bias on the estimation of the mean, the Student's t-test of equality of means will be used. In particular, according with our specific context, the Welch's adaptation of the two sample t-test (Welch, 1947, 1951) will be considered. Welch's t-test is a more reliable version of the test when the two samples have unequal variances and/or unequal sample sizes.

Welch's t-test defines the statistic as follows:

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{N}}},$$

where $\frac{S_X^2}{n}$ and $\frac{S_Y^2}{N}$ denote the estimated variances of $\overline{X}$ and $\overline{Y}$, respectively; being the degrees of freedom:

$$d.f. = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{N}\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{N^2(N-1)}}.$$

This test will not be affected by the condition $N/n \to \infty$ since, in that case, the variance of $\overline{X} - \overline{Y}$,

$$\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{N}$$

tends to the variance of $\overline{X}$. Therefore, the statistic will be approximated by:

$$t \approx \frac{\overline{X} - \mu}{\sqrt{\frac{S_X^2}{n}}},$$

which has asymptotic distribution $N(0, 1)$.

3.3 Simulation

The performance of the tests proposed in Section 3.1 and 3.2 is studied via simulation. We generated 10^3 pairs of datasets, each with sample size $n = 10^3$ in the case of the sample $\mathbf{X}$, sample size $N = 10^6$ for the sample $\mathbf{Y}$ and sample size $n = 10^3$ in the case of the sample $\mathbf{Z}$.

The model presented below has been designed to easily simulate the samples $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ involved in the study. Furthermore, taking into account the hypotheses considered in the theoretical results of Chapters 4 and 5, we opted for a model in which the densities are bounded away from zero.

Let us consider a population with density f,

$$f(x) = \frac{3}{14}(x^2+1)\,\mathbf{1}_{[0,2]}(x), \tag{3.7}$$

from which the sample $\mathbf{X}$ is generated and the following class of weight functions,

$$w(x) = \varepsilon^k\,\mathbf{1}_{[0,\varepsilon]}(x) + x^k\,\mathbf{1}_{(\varepsilon,2]}(x), \tag{3.8}$$

with different choices of $k > 0$ and $\varepsilon > 0$.

The biased density is

$$g(x) = \frac{3}{14c}\varepsilon^k(x^2+1)\,\mathbf{1}_{[0,\varepsilon]}(x) + \frac{3}{14c}(x^{k+2}+x^k)\,\mathbf{1}_{(\varepsilon,2]}(x), \tag{3.9}$$

being

$$c = \frac{1}{14}\left[\frac{k\cdot\varepsilon^{k+3}+3\cdot 2^{k+3}}{k+3} + \frac{3(k\cdot\varepsilon^{k+1}+2^{k+1})}{k+1}\right], \tag{3.10}$$

from which we simulate the sample $\mathbf{Y}$.

The twice biased density is

$$m(x) = \frac{3}{14c_2}\varepsilon^{2k}(x^2+1)\,\mathbf{1}_{[0,\varepsilon]}(x) + \frac{3}{14c_2}(x^{2k+2}+x^{2k})\,\mathbf{1}_{(\varepsilon,2]}(x), \tag{3.11}$$

being

$$c_2 = \frac{1}{14}\left[\frac{2k\cdot\varepsilon^{2k+3}+3\cdot 2^{2k+3}}{2k+3} + \frac{3(2k\cdot\varepsilon^{2k+1}+2^{2k+1})}{2k+1}\right]. \tag{3.12}$$

from which the sample $\mathbf{Z}$ can be easily simulated.

Different combinations of k and ε are considered in this simulation study, providing very biased situations ($k = 2$, $\varepsilon = 0.1$) and others in which bias is quite significant (see Figure 3.1), decreasing the degree of bias by decreasing k and increasing ε (see Figure 3.2), until reaching situations in which bias is almost imperceptible (see Figure 3.3) or it does not exist ($k = 0$, $\varepsilon = 2$).

Several indices to measure the amount of bias in Setup 1 are defined below. All of them are invariant under location and scale transformations. This means that if we consider any positive constant $a > 0$ and any real number b, the index defined for the two new random variables, $X' = aX + b$ and $Y' = aY + b$, has the same value as for the original random variables, X and Y. This is a very convenient property since the value of the index does not depend on the measure units used. All the indices except i_1 are defined in such a way that they all lie within the interval [0,1]. The value 0 for all those indices corresponds to no bias, while the value 1 is the maximal possible value of them.

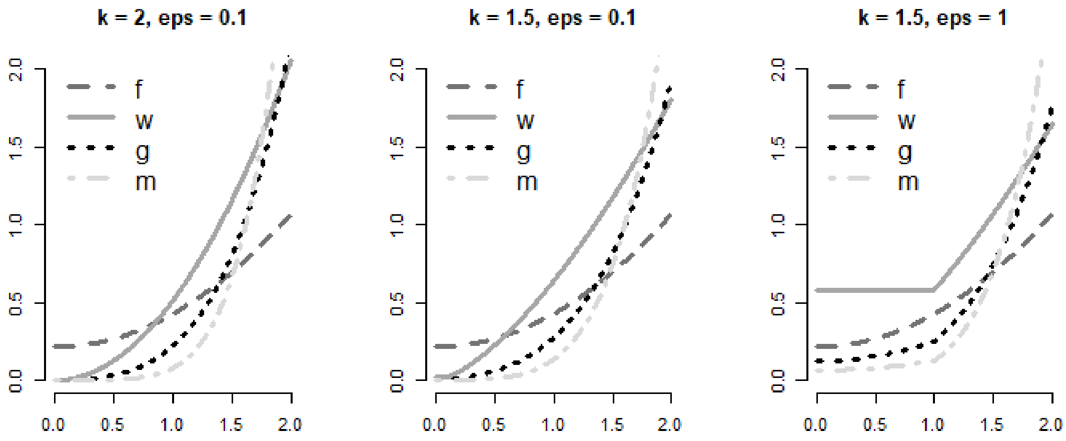

Figure 3.1: Densities f (dashed dark gray line), g (dotted black line) and m (dashed light gray line) involved in the simulated models for different values of k and ε for the biasing function, w (solid line).

The indices for Setup 2 could be obtained in an analogous way, simply considering the sample Z instead of X and density m and distribution M instead of f and F.

The first index considers the absolute value of the difference of the population means of the distributions involved in each case. The average of the standard deviations in the denominator is necessary in order to obtain an scale-invariant index. For Setup 1:

$$i_1 = \frac{|\mu_Y - \mu_X|}{\frac{\sigma_X + \sigma_Y}{2}}.$$

The following two indices are based, respectively, on the L_1 and L_2 distances

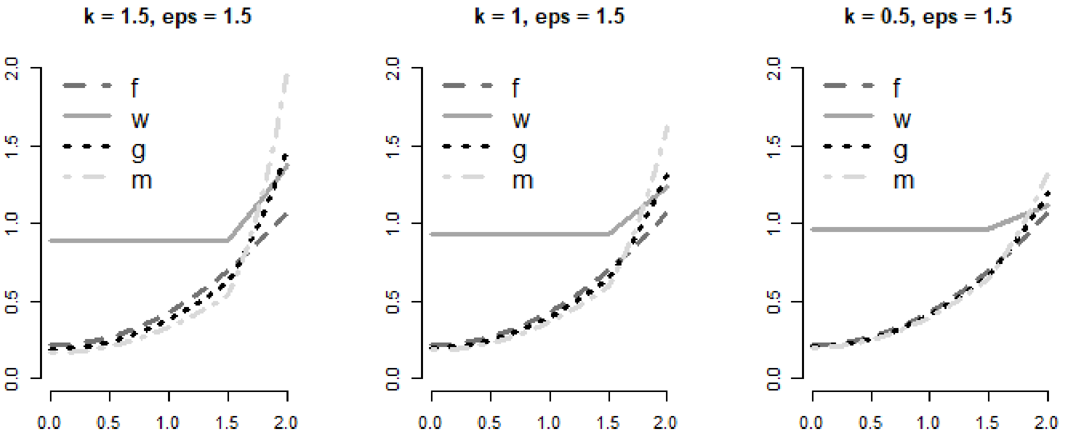

Figure 3.2: Densities f (dashed dark gray line), g (dotted black line) and m (dashed light gray line) involved in the simulated models for different values of k and ε for the biasing function, w (solid line).

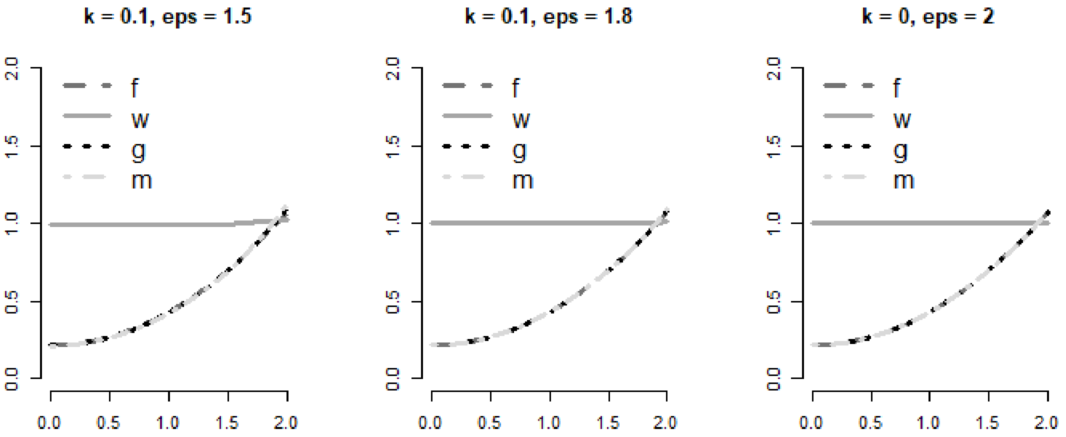

Figure 3.3: Densities f (dashed dark gray line), g (dotted black line) and m (dashed light gray line) involved in the simulated models for different values of k and ε for the biasing function, w (solid line).

between the density functions:

$$\begin{aligned} d_{L_1} &= \|f-g\|_1 = \int_a^b |f(x)-g(x)|dx, \\ d_{L_2} &= \|f-g\|_2 = \left[\int_a^b (f(x)-g(x))^2 dx\right]^{1/2}, \end{aligned}$$

where $[a, b]$ is the common support of F and G.

Since the distance d_{L_1} takes values between 0 and 2, the second index is divided by 2 in order to be in the range $[0, 1]$:

$$i_2 = \frac{1}{2}\|f - g\|_1.$$

Since the distance d_{L_2} is not an scale-invariant measure, it is transformed to obtain the third relative index in $[0, 1]$:

$$i_3 = \frac{\|f - g\|_2}{\|f\|_2 + \|g\|_2}.$$

The fourth and fifth indices consider the Kolmogorov-Smirnov and the Cramer-von Mises distances between the distribution functions, respectively:

$$\begin{aligned} d_{KS} &= \sup_{x \in \mathbb{R}} |F(x) - G(x)|, \\ d_{CvM} &= \int_{-\infty}^{\infty} (F(x) - G(x))^2 \frac{1}{2} d\,(F + G)\,(x). \end{aligned}$$

The Kolmogorov-Smirnov distance is already a location and scale invariant measure that takes values in $[0, 1]$, therefore it does not require any modification to obtain the fourth index:

$$i_4 = d_{KS} = \sup_{x \in \mathbb{R}} |F(x) - G(x)|.$$

Since the Cramer-von Mises distance is location and scale invariant but takes values between 0 and 1/3, the fifth index is:

$$i_5 = 3 \int_{-\infty}^{\infty} (F(x) - G(x))^2 \frac{1}{2} d\,(F + G)\,(x).$$

Finally, an index that measures the proximity of the weight function, w, to its nearest constant is considered:

$$i_6 = \frac{\|w - c_w\|_2}{\|w\|_2 + \|c_w\|_2} = \frac{\left[\int_a^b (w(x) - c_w)^2 dx\right]^{1/2}}{\left[\int_a^b w(x)^2 dx\right]^{1/2} + (b - a)^{1/2} \cdot c_w},$$

where

$$c_w = \frac{1}{b - a} \int_a^b w(x) dx$$

and the correction in the denominator is introduced to get a location and scale invariant index with values in $[0, 1]$.

Table 3.1: Comparison of different relative bias indices in Setup 1.

k	ε	i_1	i_2	i_3	i_4	i_5	i_6
2	0.1	0.661448	0.276439	0.263611	0.276439	0.120925	0.381806
1.5	0.1	0.572910	0.231033	0.222505	0.231033	0.085240	0.332718
1.5	1	0.344384	0.165341	0.173335	0.165339	0.040523	0.192136
1.5	1.5	0.114274	0.066758	0.085891	0.066758	0.005741	0.064524
1	1.5	0.074342	0.043826	0.056667	0.043826	0.002465	0.041561
0.5	1.5	0.036569	0.021553	0.027965	0.021553	0.000594	0.020051
0.1	1.5	0.007302	0.004250	0.005513	0.004250	0.000023	0.003894
0.1	1.8	0.001315	0.000885	0.001595	0.000885	0.000001	0.000940
0	2	$4 \cdot 10^{-16}$	$1 \cdot 10^{-16}$	$1 \cdot 10^{-16}$	$4 \cdot 10^{-16}$	$7 \cdot 10^{-32}$	$4 \cdot 10^{-17}$

Table 3.2: Comparison of different relative bias indices in Setup 2.

k	ε	i_1	i_2	i_3	i_4	i_5	i_6
2	0.1	0.228873	0.143197	0.136703	0.143201	0.033393	0.381806
1.5	0.1	0.236722	0.128118	0.122137	0.128118	0.026777	0.332718
1.5	1	0.276367	0.138238	0.134581	0.138238	0.029754	0.192136
1.5	1.5	0.122380	0.072108	0.089893	0.072108	0.006874	0.064524
1	1.5	0.080052	0.046485	0.059155	0.046485	0.002819	0.041561
0.5	1.5	0.037777	0.022276	0.028752	0.022276	0.000640	0.020051
0.1	1.5	0.007173	0.004281	0.005549	0.004281	0.000023	0.003894
0.1	1.8	0.001321	0.000889	0.001604	0.000889	0.000001	0.000940
0	2	0	0	0	0	0	0

Tables 3.1 and 3.2 show the values of the different bias indices considered in Setup 1 and Setup 2, respectively. In Setup 1, it is clearly observed that in the most biased situation the value of all indices is greater, decreasing as the bias decreases. While in Setup 2, although this relationship is not so evident for the three first considerations of k and ε, something similar happens: the values of the different indices allow us to distinguish the more biased situations (the first ones) from the less biased ones (the last ones).

For the implementation of the two sample KS test in R we will use the *pkolmim* function of the *kolmim* package, which is an improved version of the *ks.test* routine presented in Carvalho (2015). The reason for using this package is that the *ks.test* function returns approximated values in case of ties, being the *pkolmim* function more efficient since it returns the exact values. For the implementation of the two-sample Cramer-von Mises test we use the *cvm_test* function of the *twosamples* package with 1000 bootstrap iterations and for the Mann-Whitney test the *wilcox.test* function of the *stats* package. As for the *ks.test* function, the *wilcox.test* returns approximated values due to the presence of ties. For the equality of means, we use the *t.test* with unequal variances.

Table 3.3: Comparison of the rejection probability of the equality of the distributions F and G in Setup 1 using the-two sample KS test, through the *ks.test* and the *pkolmim* functions, the two-sample Cramer-von Mises criterion and the Mann-Whitney-Wilcoxon U-test and comparison of the rejection probability of the equality of means using the Welch's t-test for different values of k and ε ($n = 10^3$, $N = 10^6$, trials=10^3, $\alpha = 0.05$).

k	ε	*ks.test*	*pkolmim*	*cvm_test*	*wilcox.test*	*t.test*
2	0.1	1	1	1	1	1
1.5	0.1	1	1	1	1	1
1.5	1	1	1	1	1	1
1.5	1.5	0.990	0.991	1	0.990	0.956
1	1.5	0.781	0.786	0.915	0.774	0.651
0.5	1.5	0.237	0.247	0.468	0.273	0.209
0.1	1.5	0.053	0.055	0.161	0.059	0.050
0.1	1.8	0.048	0.050	0.150	0.052	0.052
0	2	0.049	0.050	0.154	0.051	0.050

Table 3.4: Comparison of the rejection probability of the equality of the distributions G and M in Setup 2 using the-two sample KS test, through the *ks.test* and the *pkolmim* functions, the two-sample Cramer-von Mises criterion and the Mann-Whitney-Wilcoxon U-test and comparison of the rejection probability of the equality of means using the Welch's t-test for different values of k and ε ($n = 10^3$, $N = 10^6$, trials=10^3, $\alpha = 0.05$).

k	ε	*ks.test*	*pkolmim*	*cvm_test*	*wilcox.test*	*t.test*
2	0.1	1	1	1	1	1
1.5	0.1	1	1	1	1	1
1.5	1	1	1	1	1	1
1.5	1.5	0.997	0.997	0.998	0.995	0.981
1	1.5	0.848	0.851	0.936	0.847	0.715
0.5	1.5	0.282	0.285	0.486	0.300	0.234
0.1	1.5	0.051	0.056	0.166	0.061	0.058
0.1	1.8	0.045	0.046	0.164	0.048	0.041
0	2	0.041	0.044	0.166	0.044	0.039

Tables 3.3 and 3.4 show the rejection probability obtained in the implementation of the different test proposed for bias testing in Setup 1 and Setup 2, respectively. Except for the Cramer-von Mises criterion, whose bad results apparently come from a malfunction of the *twosamples* package, the rest of the methods considered to test the equality of distributions offer similar conclusions in both setups. As the indices

considered in the Tables 3.1 and 3.2 showed, for the first combinations of k and ε the absence of bias is totally rejected with probability 1, while in the last cases, the fact that we only reject H_0 5% of the times is observed. Regarding the equality of means, the conclusions are similar.

Table 3.5 shows the CPU times of the different methods considered for bias testing. These times correspond to a single combination of ε and k with $n = 10^3$, $N = 10^6$, trials=10^3. The Cramer-von Mises criterion, in addition to moving away from the behavior of the other methods, is the slowest one. Both functions used in the implementation of the Kolmogorov-Smirnov test (*ks.test* and *pkolmim*) offer similar simulation times; however, it is preferable to use the *pkolmim* function since, as we have already mentioned, it is the only one that returns exact p-values. The Mann-Whitney U-test throught the *wilcox.test* function is also quite fast, but it returns approximated p-values like the *ks.test* function, instead of exact ones. Finally, the t-test for the equality of means is the fastest of the tests considered.

Table 3.5: Comparison of CPU times (minutes) of different methods for bias detection. The times correspond to a single combination of ε and k with $n = 10^3$, $N = 10^6$, trials=10^3.

ks.test	*pkolmim*	*cvm_test*	*wilcox.test*	*t.test*
16.8	15.7	3513.6	63.6	8

Chapter 4

Nonparametric Estimation in Setup 1

4.1 Introduction

The content of the present chapter corresponds to the work published in Borrajo & Cao (2021). It is an extension of Setup 1 introduced in Chapter 2, which deals with the mean estimaton problem in a B3D context. This chapter addresses a more general problem: the nonparametric estimation of the mean of a transformation of a continuous population. This includes as special cases the mean of the population (Chapter 2) and any other moment, the cumulative distribution function at a given point and also the characteristic function evaluated at a given value.

The estimators proposed in Chapter 2 only work in the unrealistic case where the biasing function is known. In this chapter, estimators for the unknown w case in Setup 1 are proposed.

The rest of the chapter proceeds as follows. Section 4.2 presents a density-based nonparametric estimator, as well as a general weighted estimator and includes some asymptotic results. A simulation study is included in Section 4.3, which shows that the proposed method outperforms the classical sample means, even the one computed with a simple random sample obtained without sampling bias. The simulations show striking results for the new estimator in terms of the optimal smoothing parameters. Section 4.4 presents a bootstrap algorithm to estimate the mean squared error of the estimator. Its minimization leads to a method for automatic bandwidth selection, which is a relevant practical problem. Finally, Section 4.5 contains sketches of the proofs. Detailed proofs are long and tedious and can be found in Appendix A.1.

4.2 Estimation of the mean of a transformation for B3D

We focus on the problem of estimating the mean of a transformation, v, of a continuous random variable, $\mu_v = \int v(x)f(x)dx$, where v is a known function, in the context of B3D, i.e. using a sample of a large size generated from a distribution which is not the one we are interested in, but some biased version of it. This general parameter μ_v includes as special cases: the k-th moment of the random variable (considering $v(x) = x^k$) which includes the mean ($k = 1$), already studied in Chapter 2; the cumulative distribution function $F(t)$ (considering $v(x) = \mathbf{1}_{\{x \leq t\}}$); and the characteristic function $\varphi(t)$ (taking $v(x) = \exp(itx)$), among many other. In the next subsection we look at the mathematical notation to formulate the problem.

4.2.1 Sampling bias in big data

Let us consider the continuous population and the B3D context presented in Section 2.2.2 in Chapter 2. Following parallel steps to those of that chapter, it is possible to define analogous estimators to $\tilde{\mu}^{(1)}$ and $\tilde{\mu}^{(2)}$, but for the general case of the mean of a transformation.

In case of $\mathbf{X}$ being observed, we could use the classical estimator for μ_v: the $v(X)$-sample mean,

$$\overline{v(X)} = \frac{1}{n}\sum_{i=1}^{n} v(X_i).$$

If $\mathbf{X}$ is not available and we only observe $\mathbf{Y}$, the relationship between f and g (see Equation (2.1)) motivates the definition of the following unrealistic estimator for μ_v in the known w case (Cao & Borrajo, 2018):

$$\tilde{\mu}_v^{(1)} = \frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)}{w\left(Y_i\right)} = \frac{1}{N}\sum_{i=1}^{N} \frac{f\left(Y_i\right)v(Y_i)}{g\left(Y_i\right)}, \tag{4.1}$$

since

$$E\left(\frac{v(Y)}{w\left(Y\right)}\right) = \int \frac{v(y)}{w\left(y\right)} g\left(y\right)dy = \int v(y)f\left(y\right)dy = \mu_v.$$

Since $\tilde{\mu}_v^{(1)}$ is the sample mean of the simple random sample $Z_i = v(Y_i)/w(Y_i),\ i = 1, \ldots, N$, it can be proven that is an unbiased estimator of μ_v, with normal asymptotic distribution and variance σ_Z^2/N, where $\sigma_Z^2 = \int v(x)^2 f(x)^2 g(x)^{-1}dx - \mu_v^2$.

The estimator in (4.1) is a weighted average of the $v(Y)$-sample, but the sample weights $f\left(Y_i\right)/g\left(Y_i\right)$ do not sum up to 1. So, we can consider the empirical version of the expectation ratio

$$\frac{E\left(\frac{v(Y)}{w(Y)}\right)}{E\left(\frac{1}{w(Y)}\right)},$$

which is equal to μ_v, since

$$E\left(\frac{1}{w\left(Y\right)}\right) = \int \frac{1}{w\left(y\right)} g\left(y\right) dy = \int f\left(y\right) dy = 1.$$

Thus, a reasonable modification of $\tilde{\mu}_v^{(1)}$ is given by:

$$\tilde{\mu}_v^{(2)} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{v(Y_i)}{w\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w\left(Y_i\right)}} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)v(Y_i)}{g\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{g\left(Y_i\right)}}. \tag{4.2}$$

Note that the estimators $\tilde{\mu}^{(1)}$ and $\tilde{\mu}^{(2)}$ proposed in Chapter 2 are particular cases of the estimators in (4.1) and (4.2) when choosing v as the identity function.

In general, the biasing function, w, is not known (because the underlying densities, f and g, are unknown), so the estimators in (4.1) and (4.2) are useless. However, since the Y-sample is available, it is certainly possible to estimate the density g and then plug it into (4.1) or (4.2). To end up with completely observable versions of $\tilde{\mu}_v^{(1)}$ and $\tilde{\mu}_v^{(2)}$, we also need to estimate the density f. This can be done if we are able to collect a simple random sample, $\mathbf{X}$, from the real population with density f.

Of course, when having the sample $\mathbf{X}$, it is certainly possible to estimate μ_v based on it. However, when the sample size of $\mathbf{X}$ is small, the quality of estimators based on it may be poor, while estimating μ_v using $\mathbf{Y}$ will have a much smaller variance (although some bias) due to its much larger sample size. The two resulting estimators will be presented in the following subsection.

4.2.2 Density-based estimation in Setup 1

Equations (4.1) and (4.2) require the estimation of the densities f and g (or the estimation of the biasing function, w) to obtain real (observable) estimators for the population mean. This is possible if one is able to collect information on f and g or indirect information on w.

Let us assume that we observe the big data sample of a large size N, $\mathbf{Y} = (Y_1, \dots, Y_N)$, from the biased distribution G (density g). In Setup 1, we suppose that we also observe a simple random sample, $\mathbf{X} = (X_1, \dots, X_n)$, of a much smaller sample size n ($n << N$) of the real population F (density f).

We will now discuss how we can build estimators for μ_v in this setup.

The Parzen-Rosenblatt kernel density estimators (Parzen, 1962; Rosenblatt, 1956) based on the samples $\mathbf{X}$ and $\mathbf{Y}$ can be used to estimate $f(x)$ and $g(x)$:

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \\ \hat{g}_b(x) &= \frac{1}{N}\sum_{i=1}^{N} K_b(x - Y_i), \end{aligned}$$

where $K_h(u) = (1/h)K(u/h)$, being K a kernel function and h and b two suitable bandwidths. The biasing function, w, can be easily estimated as the ratio of both estimated densities: $\hat{w}_{h,b}(x) = \hat{g}_b(x)/\hat{f}_h(x)$.

Plugging these estimators into (4.1) and (4.2) leads to observable versions of $\tilde{\mu}_v^{(1)}$ and $\tilde{\mu}_v^{(2)}$ defined as follows:

$$\hat{\mu}_v^{1,h,b} = \frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)}{\hat{w}_{h,b}(Y_i)} = \frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}, \tag{4.3}$$

$$\hat{\mu}_v^{2,h,b} = \frac{\frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)}{\hat{w}_{h,b}(Y_i)}}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{\hat{w}_{h,b}(Y_i)}} = \frac{\frac{1}{N}\sum_{i=1}^{N} v(Y_i)\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}}{\frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}}. \tag{4.4}$$

In fact $\hat{\mu}_v^{1,h,b}$ and $\hat{\mu}_v^{2,h,b}$ depend on the two bandwidths, h and b, so the role of these smoothing parameters is a relevant issue.

From now on, $\hat{\mu}^{1,h,b}$ and $\hat{\mu}^{2,h,b}$ denote the estimators in (4.3) and (4.4) when considering the mean estimation problem, i.e. $v(x) = x$. In the particular case of the mean estimation and $h = b$, the estimators $\hat{\mu}^{1,h,b}$ and $\hat{\mu}^{2,h,b}$ turn out to be much simpler when $h \to \infty$ and when $h \to 0^+$ or for very small h. This is stated in the next lemma, that requires the following assumptions on the kernel function and the two observed samples.

A1. K is a continuous function and $K(0) > 0$.
A2. $\lim_{u\to\pm\infty} K(u) = 0$.
A3. For any $a > 1$ we have $\lim_{v\to\pm\infty} K(av)/K(v) = 0$.
A4. There is a unique pair (i_0, j_0) such that $|Y_{i_0} - X_{j_0}| = \min B$, with $B = \{|Y_i - X_j|, i = 1, \ldots, N, j = 1, \ldots, n\}$.
A5. The support of the kernel K is $[-1, 1]$.

Conditions A1 and A2 are fulfilled by most of the classical kernels. Condition A3 is satisfied by kernels with exponentially decaying tails (e.g. the Gaussian kernel). Condition A5 is not compatible with A3 and it is fulfilled by the uniform, the triangular and the Epanechnikov kernel, among many other. Condition A4 is a technical assumption needed to identify which one of the terms $K((Y_i - X_j)/h)$ dominates when h tends to zero and Condition A3 holds.

Lemma 4.2.1. *Consider two fixed samples, $\boldsymbol{X}$ and $\boldsymbol{Y}$, with sizes n and N, and equal smoothing parameters, $h = b$. Then the extreme undersmoothing and oversmoothing versions of the estimators in (4.3) and (4.4) is reduced as follows. If Condition A1 holds, then*

$$\lim_{h\to\infty} \hat{\mu}^{1,h,h} = \overline{Y}, \ \lim_{h\to\infty} \hat{\mu}^{2,h,h} = \overline{Y}.$$

Under Condition A2, and assuming there are no ties between any Y_i and any X_j, we have

$$\lim_{h\to 0^+} \hat{\mu}^{1,h,h} = 0.$$

Assuming there are no ties in the union of the X and Y samples, Conditions A2, A3 and A4 imply

$$\lim_{h\to 0^+} \hat{\mu}^{2,h,h} = \arg\min_{y\in\{Y_1,\ldots Y_N\}} \min_{j\in\{1,\ldots,n\}} |y - X_j|.$$

Finally, either assuming Condition A3 or A5 and defining $b_N = (\log N)/N$ we have

$$\hat{\mu}^{1,b_N,b_N} \simeq \overline{X}, \ \hat{\mu}^{2,b_N,b_N} \simeq \overline{X}.$$

The two sample means $\overline{X}$ and $\overline{Y}$ are extreme cases in the family of estimators $\hat{\mu}^{1,h,b}$ and $\hat{\mu}^{2,h,b}$ for $h = b$. Other type of estimators with those two as extreme cases is the convex linear combination of $\overline{X}$ and $\overline{Y}$: $\hat{\mu}^{3,\lambda} = \lambda\overline{X} + (1-\lambda)\overline{Y}$, for $\lambda \in [0, 1]$. Simulation results, not reported in this chapter, showed that the estimator with minimal mean squared error within this convex linear combinations family exhibits a worse performance than the best estimator in the family $\hat{\mu}^{2,h,b}$, proposed here, since

$$MSE(\hat{\mu}^{3,\lambda}) = (1-\lambda)^2(\mu_g - \mu)^2 + \lambda^2\frac{\sigma_X^2}{n} + (1-\lambda^2)\frac{\sigma_Y^2}{N},$$

where μ_g is the biased population mean.

From now on, only the estimator $\hat{\mu}_v^{2,h,b}$ for Setup 1 and its analogous version for the mean, $\hat{\mu}^{2,h,b}$, will be considered. For the sake of brevity these estimators are denoted by $\hat{\mu}_v$ and $\hat{\mu}$, respectively, in the rest of the chapter, except when a more explicit notation is needed.

4.2.3 General weighted estimator

When the function w is known (unrealistic situation) the estimator $\tilde{\mu}_v^{(2)}$ in (4.2) is a particular case of a general weighted estimator of the form

$$\hat{\mu}_v^{\tau} = \frac{\frac{1}{N}\sum_{i=1}^{N}\tau\left(Y_i\right)v(Y_i)}{\frac{1}{N}\sum_{i=1}^{N}\tau\left(Y_i\right)},$$

where $\tau : \mathbb{R} \to \mathbb{R}^+$ is a general weight function. For this estimator, it can be proven (see Appendix A.1.2) that

$$\begin{aligned} MSE(\hat{\mu}_v^{\tau}) &\simeq \left[\int \tau(y)(v(y)-\mu_v)g(y)dy\right]^2 \\ &+ \frac{1}{N}\int\left[\tau(y)(v(y)-\mu_v) - \int \tau(x)(v(x)-\mu_v)g(x)dx\right]^2 g(y)dy =: \Psi(\tau) \geq 0. \end{aligned}$$

Observe that if $\tau_1(y)(v(y)-\mu_v) = 0, \forall y \in \mathbb{R}$, then $\Psi(\tau_1) = 0$. But the only way to fulfill this is that $\tau_1(y) \equiv 0$ which gives the undefined value $\hat{\mu}_v^{\tau_1} = 0/0$.

On the other hand, if $\tau_2(y)(v(y)-\mu_v) = \lambda$ (constant), with $\lambda \neq 0$ then

$$\Psi(\tau_2) = \left[\int \lambda g(y)dy\right]^2 + \frac{1}{N}\int\left[\lambda - \int \lambda g(x)dx\right]^2 g(y)dy = \lambda^2. \qquad (4.5)$$

So by picking a small value of $\lambda > 0$ we obtain $\Psi(\tau_2) = \lambda^2$ as close to zero as desired by just considering $\tau_2(y) = \lambda/(v(y)-\mu_v)$.

For the function $\tau_3(y) = f(y)/g(y) = 1/w(x)$ we get

$$\Psi(\tau_3) = \frac{1}{N}\int \frac{f(y)^2}{g(y)}(v(y)-\mu_v)^2 dy. \qquad (4.6)$$

By examining (4.5) and (4.6) we conclude that $MSE(\tilde{\mu}_v^{(2)}) = MSE(\hat{\mu}_v^{\tau_3}) \simeq \Psi(\tau_3)$ tends to zero at the rate $1/N$ and $MSE(\hat{\mu}_v^{\tau_2}) \simeq \Psi(\tau_2)$ can be arbitrary small (λ^2) by taking λ small. In fact, a suitable choice of λ (a very small one) will give $AMSE(\tilde{\mu}_v^{(2)}) = AMSE(\hat{\mu}_v^{\tau_2}) < AMSE(\hat{\mu}_v^{\tau_3})$ using $\tau_2(y) = \lambda/(v(y) - \mu_v)$. Of course this choice of τ_2 is not available in practice (since it depends on the true μ_v) but we may get very close to λ^2 using other feasible functions $\tau(y)$ which are not far from $\tau_2(y)$.

4.2.4 Asymptotic results

To obtain the asymptotic mean squared error of $\hat{\mu}_v$ we need the following two assumptions:
A6. The density functions f and g are six times differentiable and all their first six derivatives are bounded. Additionally f is bounded and g is bounded away from zero in the support of f. The function v is six times differentiable.
A7. The following integrals are finite:

$$\begin{aligned}
&\int v(y)^2 f(y)dy < \infty \\
&\int v^{(k)}(y)^2 f(y)dy < \infty \quad \text{for} \quad k = 2, 4, 6 \\
&\int (v(y) - \mu_v)^2 |f^{(k)}(y)| dy < \infty \quad \text{for} \quad k = 2, 4, 6.
\end{aligned}$$

The fact that the sample size of the big-but-biased sample is much larger than the one of the simple random sample is translated into the asymptotic condition $N/n \to \infty$.

Theorem 4.2.1. *Under the classical conditions on the bandwiths and the sample sizes, i.e. $h \to 0$, $b \to 0$, $nh \to \infty$, $Nb \to \infty$ and $N/n \to \infty$, if Conditions A1, A6 and A7 are fulfilled, then the asymptotic mean squared error of $\hat{\mu}_v$ is*

$$\begin{aligned}
AMSE(\hat{\mu}_v) &= \left(C_1 b^2 + \frac{C_2}{Nb} + C_3 h^2\right)^2 + \frac{C_4}{n} + \frac{C_5}{Nn} + \frac{C_6}{N^2} + \frac{C_7}{Nnh} \\
&+ \frac{C_8}{N^2 b} + \frac{C_9 h^2}{n} + \frac{C_{10} h^2}{N^2 b} + \frac{C_{11} h^4}{N} + \frac{C_{12} h^2 b^2}{N}, \qquad (4.7)
\end{aligned}$$

where the first three terms come from the squared bias and the rest of them from the variance of the estimator. The constants $C_1, \ldots, C_{12}$ are defined in the sketch of the proofs (Subsection 4.5.1).

Corollary 4.2.1. *Under the same conditions in Theorem 4.2.1 and choosing v as the identity function, since the constants C_3 and C_9 in (4.7) become zero, the asymptotic mean squared error of $\hat{\mu}$ is*

$$\begin{aligned} AMSE\left(\hat{\mu}\right) &= \left(C_1 b^2 + \frac{C_2}{Nb}\right)^2 + \frac{C_4}{n} + \frac{C_5}{Nn} + \frac{C_6}{N^2} + \frac{C_7}{Nnh} + \frac{C_8}{N^2 b} \\ &+ \frac{C_{10}h^2}{N^2 b} + \frac{C_{11}h^4}{N} + \frac{C_{12}h^2 b^2}{N}. \end{aligned}$$

Consequently, the expression for the optimal bandwidth h is:

$$h_{AMSE} = \tilde{h}_0 = \left(\frac{C_7}{4C_{11}}\right)^{1/5} n^{-1/5} \tag{4.8}$$

and the one for the optimal bandwidth b:

$$b_{AMSE} = \tilde{b}_0 = \begin{cases} \left(\frac{-C_2}{C_1}\right)^{1/3} N^{-1/3} & \text{if} \quad C_1 C_2 < 0 \\ \left(\frac{C_2}{2C_1}\right)^{1/3} N^{-1/3} & \text{if} \quad C_1 C_2 > 0. \end{cases} \tag{4.9}$$

Remark 4.2.1. *The dominant part of $AMSE\left(\hat{\mu}\right)$ in Corollary 4.2.1 is*

$$AMSE_0\left(\hat{\mu}\right) = \left(C_1 b^2 + \frac{C_2}{Nb}\right)^2 + \frac{C_4}{n}.$$

Since $C_4 = Var(X)$ for $v(x) = x$, we have $MSE(\overline{X}) = C_4/n < AMSE_0\left(\hat{\mu}\right)$, which implies that, in the asymptotic sense of Theorem 4.2.1, $\hat{\mu}$ cannot beat the classical estimator $\overline{X}$.

The simulations results in Section 4.3 below, show a very good performance of the proposed mean estimator, which contradicts somehow Remark 4.2.1. Moreover, they are very striking with respect to the theoretical asymptotically optimal bandwidths derived from Corollary 4.2.1. More specifically, Table 4.3 motivates to study theoretically the asymptotic properties of $\hat{\mu}_v$ under the non-standard conditions that the two bandwidths, h and b, tend to positive constants when the sample size tends to infinity. The asymptotic mean squared error for the estimator under these non-standard conditions is presented in Theorem 4.2.2 below. Three assumptions are needed:

A8. The kernel K is bounded.

A9. The density function g is bounded away from zero.

A10. The integral $\int (v(y) - \mu_v)^2 g(y) dy$ is finite.

Theorem 4.2.2. *Let us assume $h \to h_0 > 0$, $b \to b_0 > 0$, $n \to \infty$, $N/n \to \infty$, and Conditions A1 and A8-A10. The asymptotic mean squared error for the estimator $\hat{\mu}_v$ in (4.4) is given by*

$$AMSE\left(\hat{\mu}_v\right) = C_1^* + \frac{C_2^*}{n} + \frac{C_3^*}{N} + \frac{C_4^*}{Nn} + \frac{C_5^*}{N^2} + \frac{C_6^*}{N^3},$$

where the first two constants are

$$\begin{aligned} C_1^* &= \left(\int \frac{K_h * f(y)}{K_b * g(y)}(v(y) - \mu_v)g(y)dy\right)^2 \\ C_2^* &= \int \left(\int \frac{K_h(y-z)}{K_b * g(y)}(v(y) - \mu_v)g(y)dy - C_1^{*1/2}\right)^2 f(z)dz \end{aligned}$$

and C_3^*, C_4^*, C_5^* and C_6^* are constants depending on populational functions reported in the sketch of the proofs (Subsection 4.5.2).

Remark 4.2.2. *In general, the integral in C_1^* can be made equal to zero for $h \in I$ (a suitable subinterval of $[0, \infty)$) by choosing an adequate $b = \lambda(h)$, as a function of h. This relationship still gives us the freedom to choose $h \in I$. In this case $AMSE\left(\hat{\mu}_v\right) = C_2^*/n + C_3^*/N + C_4^*/(Nn) + C_5^*/N^2 + C_6^*/N^3$, where the constant $C_2^* \geq 0$ can be simplified to:*

$$C_2^* = C_2^*(h, b) = \int \left(\int \frac{K_h(y-z)}{K_b * g(y)}(v(y) - \mu_v)g(y)dy\right)^2 f(z)dz.$$

In practice we can choose $h_0 \in I$ and $b_0 = \lambda(h_0)$ such that $C_2^*(h_0, b_0)$ can be minimized. The resulting constant is very close to zero (see Table 4.2 below).

For the mean case and the models used in the simulations in Section 4.3, the constant $C_2^*(h_0, b_0)$ is much smaller than C_4 in Corollary 4.2.1 (see Table 4.2). This implies that the dominant part of $AMSE\left(\hat{\mu}^{2,h_0,b_0}\right)$ in Theorem 4.2.2, $AMSE_1\left(\hat{\mu}^{2,h_0,b_0}\right)$, is

$$AMSE_1\left(\hat{\mu}^{2,h_0,b_0}\right) = C_2^*(h_0, b_0)/n < AMSE_0\left(\hat{\mu}^{2,\tilde{h}_0,\tilde{b}_0}\right)$$

for $\tilde{h}_0$ and $\tilde{b}_0$ in (4.8) and (4.9). This means that the optimal choice for the bandwidths when they are considered tending to some positive constants, (h_0, b_0), gives a smaller MSE than the optimal bandwidths, $(\tilde{h}_0, \tilde{b}_0)$, tending to zero. This striking result resembles the one obtained by Delyon & Portier (2016) when estimating $\int \varphi(x)dx$ using an average of ratios of the form $\varphi(X_i)/\hat{f}_h(X_i)$ for a known function φ.

Remark 4.2.3. *Since $C_4 = Var(X)$, using parallel arguments to those in Remark 4.2.2, we have $AMSE_1\left(\hat{\mu}^{2,h_0,b_0}\right) < MSE\left(\overline{X}\right)$. This implies that, using the optimal constant bandwidths h_0 and b_0, $\hat{\mu}^{2,h_0,b_0}$ beats the classical mean estimator $\overline{X}$.*

4.2.5 Further applicability

When working with B3D, it is common to deal with extremely asymmetric or heavy tailed variables. Logarithmic and other types of transformations are often used when X is a positive random variable. This situations is covered in this chapter by just considering $v(x) = \log x$, giving $\mu_v = E(\log X)$. This parameter can be estimated using (4.3) or (4.4) with $v(x) = \log x$.

Data with outliers is a common feature also in big data. In such a situation the underlying density, $f(x)$, is not directly observed. A large sized sample from a mixture density $g(x) = (1-\alpha)f(x) + \alpha h(x)$ is observed instead. Here the density $h(x)$ denotes the outlier generation density. This contaminated version of $f(x)$ is covered by Equation (2.1) by just considering $w(x) = 1 - \alpha + \alpha h(x)/f(x)$ and that the support of $h(x)$ is contained in the support of $f(x)$. If, for instance, the outlier generator density, $h(x)$ is proportional to the underlying density, $f(x)$, in some region, then the biasing weight function $w(x)$ will be constant in that region. This means that the bias is only present in the region where outliers are expected. In general, if α is very small, the bias is small too, although it heavily depends on the values x for which $h(x)$ is very different from $f(x)$. If one has qualitative information about the shape of the outlier generator density, $h(x)$, it could be incorporated in the estimation process. Otherwise, the estimators in (4.3) and (4.4) can be used to estimate any moment of X, or the expected value of any transformation of interest.

4.3 Simulations

The performance of the mean estimator, $\hat{\mu}$, proposed in Section 4.2, is studied via simulation. We generated 10^3 pairs of datasets, each with sample size $n = 10^3$ in the case of the sample $\mathbf{X}$ and sample size $N = 10^6$ for the sample $\mathbf{Y}$.

Let us consider the population presented in Subsection 3.3. We suppose that the sample $\mathbf{X}$ is generated from the density f defined in (3.7). Considering the class of weight functions described in (3.8), we simulate the sample $\mathbf{Y}$ from the biased density characterized in (3.9) and (3.10).

Different combinations of k and ε are considered in this simulation study (see Figure 4.1), providing very biased situations ($k = 1.5$ and $\varepsilon = 0.1$) and others in which bias is almost imperceptible ($k = 0.1$ and $\varepsilon = 1.8$).

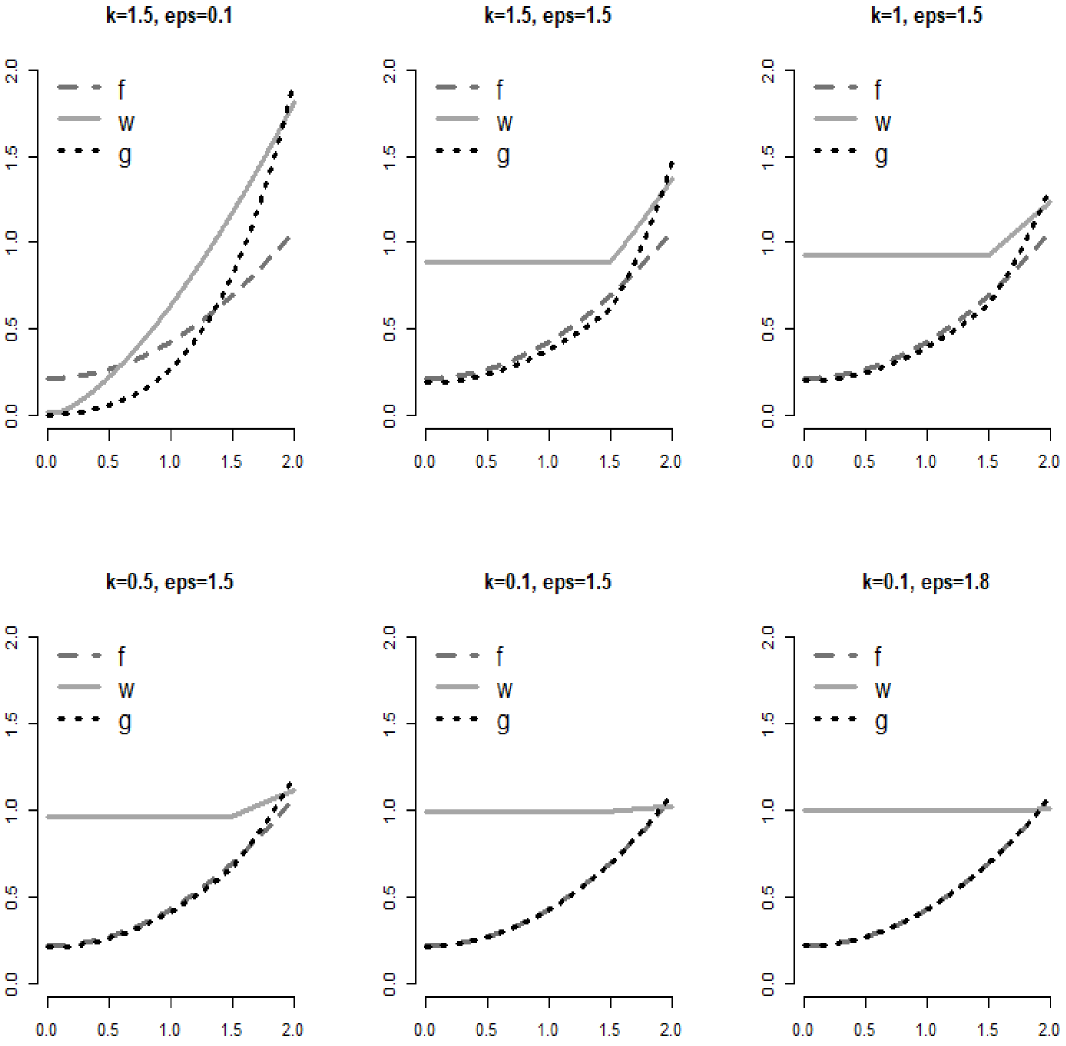

Figure 4.1: Densities f (dashed line) and g (dotted line) involved in the simulated models for different values of k and ε for the biasing function, w (solid line).

Table 4.1 shows the good performance of $\hat{\mu}$ when compared with the classical sample means. As expected, in the most biased situations, it is preferable to use the mean of the simple random sample $\mathbf{X}$ rather than the $\mathbf{Y}$ sample mean; while in the opposite case, the mean of the biased sample works better due to the larger sample size. Anyhow, $\hat{\mu}$ outperforms both estimators even in the extreme cases. It is also observed how the optimal bandwidths obtained by simulation, h_{opt} and b_{opt}, widely differ from the asymptotic optimal ones, $\tilde{h}_0$ and $\tilde{b}_0$ derived from Theorem 4.2.1. The former are not only larger, but they also seem to contradict the classical condition

Table 4.1: Comparison of the MSE of $\overline{X}$, $\overline{Y}$ and $\hat{\mu}^{2,h_{opt},b_{opt}}$ for different values of k and ε ($n = 10^3$, $N = 10^6$, trials=10^3). $\tilde{h}_0$ and $\tilde{b}_0$ refer to the asymptotic optimal bandwidths, in expressions (4.8) and (4.9), and h_{opt} and b_{opt} to the optimal ones obtained by simulation.

k	ε	$MSE(\overline{X})$	$MSE(\overline{Y})$	$MSE(\hat{\mu})$	h_{opt}	b_{opt}	$\tilde{h}_0$	$\tilde{b}_0$
1.5	0.1	$2.7 \cdot 10^{-4}$	$8.2 \cdot 10^{-2}$	$1.2 \cdot 10^{-6}$	4.22	0.32	0.19	0.012
1.5	1.5	$2.7 \cdot 10^{-4}$	$3.8 \cdot 10^{-3}$	$2.7 \cdot 10^{-7}$	12.52	1.05	0.20	0.017
1	1.5	$2.7 \cdot 10^{-4}$	$1.6 \cdot 10^{-3}$	$2.7 \cdot 10^{-7}$	14.41	1.24	0.20	0.017
0.5	1.5	$2.7 \cdot 10^{-4}$	$4.0 \cdot 10^{-4}$	$2.9 \cdot 10^{-7}$	15.16	1.68	0.20	0.012
0.1	1.5	$2.7 \cdot 10^{-4}$	$1.6 \cdot 10^{-5}$	$3.0 \cdot 10^{-7}$	16.00	3.54	0.21	0.010
0.1	1.8	$2.7 \cdot 10^{-4}$	$8.1 \cdot 10^{-7}$	$3.1 \cdot 10^{-7}$	16.19	7.48	0.21	0.009

Table 4.2: Comparison of the value of constant C_4 with the values of constant $C_2^*(h, b)$ for $h = h_{opt}$ and $b = b_{opt}$ the optimal bandwidths obtained by simulation ($n = 10^3$, $N = 10^6$, trials=10^3).

k	ε	$MSE(\overline{X})$	C_4	h_{opt}	b_{opt}	$C_2^*(h_{opt}, b_{opt})$
1.5	0.1	$2.7 \cdot 10^{-4}$	0.27	4.22	0.32	$2.3 \cdot 10^{-6}$
1.5	1.5	$2.7 \cdot 10^{-4}$	0.27	12.52	1.05	$1.4 \cdot 10^{-8}$
1	1.5	$2.7 \cdot 10^{-4}$	0.27	14.41	1.24	$1.0 \cdot 10^{-8}$
0.5	1.5	$2.7 \cdot 10^{-4}$	0.27	15.16	1.68	$1.4 \cdot 10^{-8}$
0.1	1.5	$2.7 \cdot 10^{-4}$	0.27	16.00	3.54	$3.7 \cdot 10^{-8}$
0.1	1.8	$2.7 \cdot 10^{-4}$	0.27	16.19	7.48	$1.6 \cdot 10^{-7}$

that they tend to zero when the sample sizes tend to infinity. Table 4.2 contains the constant C_4 and C_2^* mentioned in Remark 4.2.2.

We explore the behavior of our method for smaller sample sizes. For the particular choice of $k = 1.5$, $\varepsilon = 1.5$, $n = 10^2$ and $N = 10^4$, Figure 4.2 represents the MSE of our estimator as a function of the smoothing parameters h and b. The color code on the right side of this figure refers to the MSE values of the estimator and indicates whether its value is very small (purple) or very high (yellow). The MSE of the classical estimators are also depicted in the figure, showing the good performance of our proposed estimator, in comparison with them, for a very wide range of values for the smoothing parameter h and a still wide range in the case of b. This figure also provides some approximation of the optimal values for the two bandwidths, those that minimize the MSE (in red). Figure 4.3 shows similar information for $k = 1$, $\varepsilon = 1.5$ and the samples sizes, $n = 10^3$ and $N = 10^6$, considered in the simulation study.

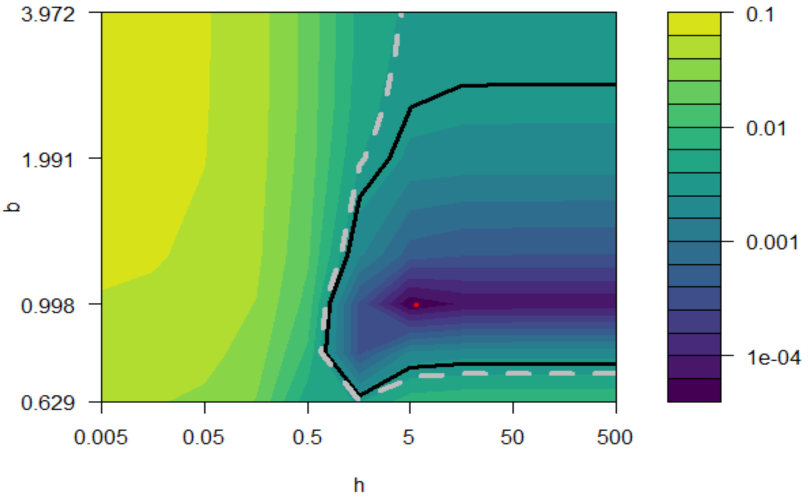

Figure 4.2: Comparison of the MSE of the proposed estimator as a function of h and b with the MSE of $\overline{X}$ (solid black line) and the MSE of $\overline{Y}$ (dashed gray line) for the particular choice of $k = 1.5$ and $\varepsilon = 1.5$, with $n = 10^2$ and $N = 10^4$. The red dot represents the minimal value of the MSE.

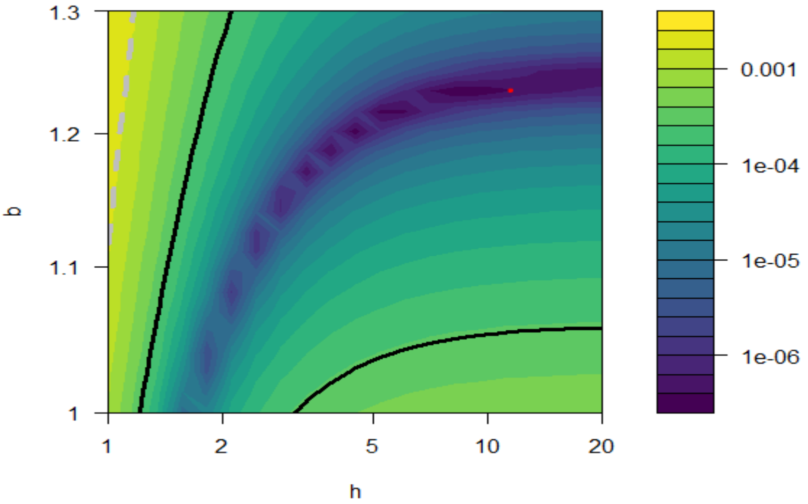

Figure 4.3: Comparison of the MSE of the proposed estimator as a function of h and b with the MSE of $\overline{X}$ (solid black line) and the MSE of $\overline{Y}$ (dashed gray line) for the particular choice of $k = 1$ and $\varepsilon = 1.5$, with $n = 10^3$ and $N = 10^6$. The red dot represents the minimal value of the MSE.

4.3.1 Asymptotically-based simulations

To further study the properties of the proposed estimator, $\hat{\mu}$, we carried out another simulation study. In order to explore the optimal values of the bandwidths obtained by simulation, we focus on the particular setting with $k = 1$ and $\varepsilon = 1.5$ and we analyze how the optimal bandwidths behave when progressively increasing the sample sizes.

The results in Table 4.3 show how the optimal values obtained by simulation do not tend to zero, but tend to constant values. This motivates studying the asymptotic behavior of the estimator under the non-standard conditions that the two bandwidths (h and b) tend to positive constants when n tends to infinity (see Theorem 4.2.2 above). Anyhow, it is worth mentioning that $MSE(\hat{\mu})$ is a very flat function of (h, b) within a large region that contains the optimal bandwidths (see Figures 4.2 and 4.3). Thus the proposed estimator is rather stable for a wide range of values for the two smoothing parameters.

Table 4.3: Comparison of the MSE of $\overline{X}$, $\overline{Y}$ and $\hat{\mu}^{2,h_{opt},b_{opt}}$ for different sample sizes and the choice of $k = 1$ and $\varepsilon = 1.5$. h_{opt} and b_{opt} refer to the optimal values obtained by simulation.

n	N	$MSE(\overline{X})$	$MSE(\overline{Y})$	$MSE(\hat{\mu})$	h_{opt}	b_{opt}
10	100	$2.9 \cdot 10^{-2}$	$4.4 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$	2.04	1.05
20	400	$1.6 \cdot 10^{-2}$	$2.2 \cdot 10^{-3}$	$6.0 \cdot 10^{-4}$	3.07	1.15
50	$2,500$	$5.6 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$9.9 \cdot 10^{-5}$	3.89	1.18
100	$10,000$	$2.9 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$2.6 \cdot 10^{-5}$	5.55	1.22
200	$40,000$	$1.5 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$7.0 \cdot 10^{-6}$	10.92	1.24
500	$250,000$	$5.8 \cdot 10^{-4}$	$1.6 \cdot 10^{-3}$	$1.0 \cdot 10^{-6}$	15.28	1.24
1000	$1,000,000$	$2.7 \cdot 10^{-4}$	$1.6 \cdot 10^{-3}$	$2.7 \cdot 10^{-7}$	14.41	1.24

4.4 Bootstrap algorithm

Minimizing in the bandwidths h and b some estimator of the MSE is a reasonable way to obtain an automatic bandwidth selection method. To do this, the following bootstrap algorithm for MSE estimation is proposed:

1. Based on the original $\mathbf{X}$ and $\mathbf{Y}$ samples, the estimated densities $\hat{f}_{h_{pil}}$ and $\hat{g}_{b_{pil}}$, where h_{pil} and b_{pil} denote the pilot bandwidths obtained from the Seather-Jones method, are considered as the true population densities in the bootstrap world.

2. Bootstrap resamples, $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ and $\mathbf{Y}^* = (Y_1^*, \ldots, Y_N^*)$, of sizes n and N respectively, are obtained from the estimated densities $\hat{f}_{h_{pil}}$ and $\hat{g}_{b_{pil}}$ as follows:

   (a) $X_i^* = \psi_i^* + h_{pil} \cdot U_i$, where $\psi^* = (\psi_1^*, \ldots, \psi_n^*)$ is a simple random sample obtained from the empirical distribution computed with the values $\mathbf{X} = (X_1, \ldots, X_n)$ and $U = (U_1, \ldots, U_n)$, with U_i simulated from the density K (a $N(0,1)$ when considering a Gaussian kernel), for $i = 1, \ldots, n$.

   (b) $Y_i^* = \eta_i^* + b_{pil} \cdot V_i$, where $\eta^* = (\eta_1^*, \ldots, \eta_N^*)$ is a simple random sample obtained from the empirical distribution computed with the values $\mathbf{Y} = (Y_1, \ldots, Y_N)$ and $V = (V_1, \ldots, V_N)$, with V_i simulated from the density K (a $N(0,1)$ when considering a Gaussian kernel), for $i = 1, \ldots, N$.

3. The estimator $\hat{\mu}^{2,h,b,*}$ is computed using the resamples $\mathbf{X}^*$ and $\mathbf{Y}^*$ and considering a very wide range of values for the smoothing parameters h and b.

4. Steps 2 and 3 are repeated a large number of times, B, in order to obtain an approximation of the bootstrap mean squared error (MSE^*) of the estimator,

$$MSE^*(h,b) = \frac{1}{B}\sum_{j=1}^{B}\left(\hat{\mu}_j^{2,h,b,*} - \overline{X}\right)^2.$$

5. The bandwidths h^* and b^* that minimize the function $MSE^*(h,b)$ are considered as bootstrap bandwidth selectors.

Since the MSE^* is not a robust measure, the presence of outliers could affect its value. In that case, other error measures could be considered, such as the bootstrap trimmed mean squared error ($TMSE^*$), i.e. the trimmed mean to a certain proportion α (the mean excluding the proportion α of the highest values) of the squared errors; or the bootstrap median of the squared errors:

$$MeSE^*(h,b) = Median\left(\hat{\mu}_j^{2,h,b,*} - \overline{X}\right)^2.$$

4.4.1 Simulations

The performance of the bootstrap algorithm is studied via simulation. We generated 100 pairs of datasets, each with sample size $n = 10^3$ in the case of the sample $\mathbf{X}$ and sample size $N = 10^6$ for the sample $\mathbf{Y}$. For each pair of generated samples, an iterative search of the bootstrap selectors is performed. This search consists of applying the bootstrap algorithm 3 times using 100 resamples $\mathbf{X}^*$ and 100 resamples $\mathbf{Y}^*$ each time. In each search, a grid of size 5 is used for each bandwidth (25 combinations of bandwidths), progressively reducing the area of this grid in each iteration of the method.

The choice of the pilot bandwidths has some relevance. In this case, we have used the bandwidths obtained by the Seather-Jones method, sometimes multiplied by a factor between 0.5 and 2.

Table 4.4: Comparison of the MSE of $\overline{X}$, $\overline{Y}$ and $\hat{\mu}^{2,h^*,b^*}$ for different values of k and ε ($n = 10^3$, $N = 10^6$, trials=100, $B = 100$). $Median(h^*)$ and $Median(b^*)$ refer to the mean of the bandwidths obtained using the bootstrap algorithm.

k	ε	$MSE(\overline{X})$	$MSE(\overline{Y})$	$MSE(\hat{\mu}^{2,h^*,b^*})$	$Median(h^*)$	$Median(b^*)$
1.5	0.1	$2.90 \cdot 10^{-4}$	$8.17 \cdot 10^{-2}$	$1.88 \cdot 10^{-4}$	17.79	0.32
1.5	1.5	$2.90 \cdot 10^{-4}$	$3.81 \cdot 10^{-3}$	$2.89 \cdot 10^{-4}$	$2.11 \cdot 10^3$	1.05
1	1.5	$2.90 \cdot 10^{-4}$	$1.64 \cdot 10^{-3}$	$2.91 \cdot 10^{-4}$	3.42	1.24
0.5	1.5	$2.90 \cdot 10^{-4}$	$3.97 \cdot 10^{-4}$	$2.23 \cdot 10^{-4}$	$3.14 \cdot 10^3$	1.68
0.1	1.5	$2.90 \cdot 10^{-4}$	$1.61 \cdot 10^{-5}$	$1.28 \cdot 10^{-4}$	67.47	8.39
0.1	1.8	$2.90 \cdot 10^{-4}$	$7.92 \cdot 10^{-7}$	$1.19 \cdot 10^{-4}$	28.80	23.64

Table 4.4 contains the true values of the MSE of the classical estimators, since they can be computed explicitely, the median of the bootstrap bandwidths, h^* and b^*, obtained from the algorithm presented above, and the MSE of the proposed mean estimator. The results in this table show how, in the first and the fourth setting considered, the more biased setting and other in which bias is quite significant, the estimator obtained with the bootstrap bandwidths outperforms the two classical sample means, $\overline{X}$ and $\overline{Y}$. In the last two cases (less biased settings), our estimator for the mean has a slightly worse behavior than $\overline{Y}$, which is logical since the bias in these situations is practically imperceptible and the size of sample $\mathbf{Y}$ is very large ($N = 10^6$). In the second and the third setting, it works better than $\overline{Y}$ but it has a behavior quite similar to $\overline{X}$. Increasing the number of resamples and trials would give a more precise picture of the real difference in MSE between the new estimator and the classical ones. However, this was discarded at this point given the high computational cost (see Table 4.5).

In conclusion, in situations where we confirm the existence of bias, we will use the bootstrap algorithm to obtain the bandwidth selectors that provide a good behavior of our estimator. On the contrary, when we reject the presence of bias, we will directly use the estimator based on the sample $\mathbf{Y}$.

Table 4.5: Comparison of CPU times (minutes) of the iterative bootstrap method applied in a grid of smoothing parameters for different combinations in the number of trials and resamples (B), considering a single combination of ε and k with $n = 10^3$ and $N = 10^6$.

B / Trials	1	50	100
1	0.55	38.62	71.16
50	39.50	1583.91	3033.50
100	72.12	2974.25	5413.41

4.5 Sketch of the proofs

4.5.1 Sketch of the proof of Theorem 4.2.1

A set of lemmas needed to proof Theorem 4.2.1 is listed below. Their detailed proofs can be found in Appendix A.1. Along this subsection, Conditions A1, A6 and A7 are assumed for Lemmas 4.5.1-4.5.10.

Lemma 4.5.1. *The difference $\hat{\mu}_v - \mu_v$ can be expressed as follows*

$$\hat{\mu}_v - \mu_v = \widehat{A} + \widehat{A}\left(1 - \widehat{B}\right) + \frac{\widehat{A}\left(1 - \widehat{B}\right)^2}{\widehat{B}} \simeq \widehat{A}, \tag{4.10}$$

where

$$\widehat{A} = \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}(v(Y_i) - \mu_v) \tag{4.11}$$

and

$$\widehat{B} = \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}. \tag{4.12}$$

The term in (4.11) can be splitted into different terms, $\widehat{A} = \widehat{A}_1 + \widehat{A}_2 - \widehat{A}_3 - \widehat{A}_4 + \widehat{A}_5$, where

$$\begin{aligned}
\widehat{A}_1 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)}{g(Y_i)}(v(Y_i) - \mu_v), \\
\widehat{A}_2 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i) - f(Y_i)}{g(Y_i)}(v(Y_i) - \mu_v), \\
\widehat{A}_3 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)(\hat{g}_b(Y_i) - g(Y_i))}{g(Y_i)^2}(v(Y_i) - \mu_v), \\
\widehat{A}_4 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{(\hat{f}_h(Y_i) - f(Y_i))(\hat{g}_b(Y_i) - g(Y_i))}{g(Y_i)^2}(v(Y_i) - \mu_v),
\end{aligned}$$

$$\widehat{A}_5 \ := \ \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i)-g(Y_i)}{g(Y_i)}\right)^2(v(Y_i)-\mu_v).$$

Since the terms $\widehat{A}_4$ and $\widehat{A}_5$ have some factors of quadratic nature inside the sum (i.e. $(\hat{f}_h(Y_i)-f(Y_i))(\hat{g}_b(Y_i)-g(Y_i))$ and $(\hat{g}_b(Y_i)-g(Y_i))^2$) they are negligible with respect to other terms. Consequently: $\widehat{A}\simeq\widehat{A}_1+\widehat{A}_2-\widehat{A}_3$.

Lemma 4.5.2. *The expectation and variance of $\widehat{A}$ can be approximated by*

$$\begin{aligned} E\left(\widehat{A}\right) &\simeq E\left(\widehat{A}_1\right)+E\left(\widehat{A}_2\right)-E\left(\widehat{A}_3\right), && (4.13)\\ Var\left(\widehat{A}\right) &\simeq Var\left(\widehat{A}_1\right)+Var\left(\widehat{A}_2\right)+Var\left(\widehat{A}_3\right) \\ &+ 2Cov\left(\widehat{A}_1,\widehat{A}_2\right)-2Cov\left(\widehat{A}_1,\widehat{A}_3\right)-2Cov\left(\widehat{A}_2,\widehat{A}_3\right). && (4.14)\end{aligned}$$

The proof of Theorem 4.2.1 proceeds by analyzing the expectations and variances involved.

Lemma 4.5.3. *The expectation of $\widehat{A}$ is*

$$\begin{aligned} E(\widehat{A}) &\simeq D_1\frac{1}{Nb}+D_2b^2+D_3b^4-D_2\frac{b^2}{N}-D_3\frac{b^4}{N} \\ &+ D_4h^2+D_5h^4+O(b^6)+O(h^6), && (4.15)\end{aligned}$$

where

$$\begin{aligned} D_1 &:= -K(0)\int\gamma(y)dy,\\ D_2 &:= -\frac{\mu_2(K)}{2}\int\gamma(y)g''(y)dy,\\ D_3 &:= -\frac{\mu_4(K)}{24}\int\gamma(y)g^{(4)}(y)dy,\\ D_4 &:= \frac{\mu_2(K)}{2}\int v''(y)f(y)dy,\\ D_5 &:= \frac{\mu_4(K)}{24}\int v^{(4)}(y)f(y)dy,\end{aligned}$$

with

$$\gamma(y):=\frac{f(y)}{g(y)}(v(y)-\mu_v).$$

Lemma 4.5.4. *The variance of $\widehat{A}_1$ is*

$$Var\left(\widehat{A}_1\right) \ = \ \frac{D_6}{N},$$

where

$$D_6:=\int\beta(y)dy,$$

with

$$\beta(y) := \frac{f(y)^2}{g(y)}(v(y)-\mu_v)^2.$$

Lemma 4.5.5. *The variance of* $\widehat{A}_2$ *is*

$$\begin{aligned} Var\left(\widehat{A}_2\right) &= D_7\frac{1}{n} + D_8\frac{1}{Nn} + D_9\frac{1}{Nnh} + D_{10}\frac{h^2}{n} + D_{11}\frac{h^4}{n} + D_{12}\frac{h^4}{N} + D_{13}\frac{h}{Nn} \\ &+ D_{14}\frac{h^2}{Nn} + D_{15}\frac{h^3}{Nn} + D_{16}\frac{h^6}{n} + D_{17}\frac{h^6}{N} + O\left(\frac{h^8}{n}\right) + O\left(\frac{h^4}{Nn}\right), \end{aligned} \tag{4.16}$$

where

$$\begin{aligned} D_7 &:= B(v^2) - \mu_v^2, \\ D_8 &:= -D_6 - B(v^2) + \mu_v^2, \\ D_9 &:= \mu_0(K^2)\int \theta(y)dy, \\ D_{10} &:= \mu_2(K)\left[B(v\cdot v'') - \mu_v B(v'')\right], \\ D_{11} &:= \frac{\mu_2(K)^2}{4}\left[B(v''^2) - B(v'')^2\right] + \frac{\mu_4(K)}{12}\left[B(v\cdot v^{(4)}) - \mu_v B(v^{(4)})\right], \\ D_{12} &:= \frac{\mu_2(K)^2}{4}\left[\int \frac{f''(y)^2}{g(y)}(v(y)-\mu_v)^2dy - B(v'')^2\right], \\ D_{13} &:= \frac{\mu_2(K^2)}{2}\int \frac{f''(y)}{g(y)}(v(y)-\mu_v)^2dy, \\ D_{14} &:= -\mu_2(K)\left[\int \theta(y)f''(y)dy + B(v\cdot v'') - \mu_v B(v'')\right], \\ D_{15} &:= \frac{\mu_4(K^2)}{24}\int \frac{f^{(4)}(y)}{g(y)}(v(y)-\mu_v)^2dy, \\ D_{16} &:= \frac{\mu_2(K)\mu_4(K)}{24}\left[B(v''\cdot v^{(4)}) - B(v'')B(v^{(4)})\right] \\ &+ \frac{\mu_6(K)}{360}\left[B(v\cdot v^{(6)}) - \mu_v B(v^{(6)})\right], \\ D_{17} &:= \frac{\mu_2(K)\mu_4(K)}{24}\left[\int \frac{f''(y)f^{(4)}(y)}{g(y)}(v(y)-\mu_v)^2dy - B(v'')B(v^{(4)})\right], \end{aligned}$$

where the operator B *is defined by*

$$B(\phi) := \int \phi(x)f(x)dx$$

and

$$\theta(y) := \frac{f(y)}{g(y)}(v(y)-\mu_v)^2.$$

Lemma 4.5.6. *The variance of $\widehat{A}_3$ is*

$$\begin{aligned} Var\left(\widehat{A}_3\right) &= D_{18}\frac{1}{N}+D_{19}\frac{b^2}{N}+D_{20}\frac{b^4}{N}+D_{21}\frac{1}{N^2b}+D_{22}\frac{1}{N^2}+D_{23}\frac{b}{N^2} \\ &+ D_{24}\frac{1}{N^3b^2}+D_{25}\frac{1}{N^3b}+D_{26}\frac{1}{N^3}+O\left(\frac{b^6}{N}\right)+O\left(\frac{b^2}{N^2}\right), \quad (4.17)\end{aligned}$$

where

$$\begin{aligned} D_{18} &:= \int \beta(y)dy, \\ D_{19} &:= \mu_2(K)\left[\int \alpha(y)g''(y)dy+\int \gamma''(y)f(y)(v(y)-\mu_v)dy\right], \\ D_{20} &:= \frac{\mu_4(K)}{12}\left[\int \alpha(y)g^{(4)}(y)dy+\int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy\right] \\ &+ \frac{\mu_2(K)^2}{4}\left[\int \delta(y)g''(y)^2dy-4\left(\int \gamma(y)g''(y)dy\right)^2\right. \\ &+ \left.\int \gamma''(y)^2g(y)dz+2\int \gamma(y)\gamma''(y)g''(y)dy\right], \\ D_{21} &:= 2\left[\mu_0(K^2)+K(0)\right]\int \alpha(y)dy, \\ D_{22} &:= -8\int \beta(y)dy=-8D_{18}, \\ D_{23} &:= \mu_2(K)K(0)\left[\int \delta(y)g''(y)dy+\int \gamma(y)\gamma''(y)dy\right. \\ &- \left.\left(\int \gamma(y)g''(y)dy+\int \gamma''(y)g(y)dy\right)\left(\int \gamma(y)dy\right)\right] \\ &+ \frac{\mu_2(K^2)}{2}\left[\int \delta(y)g''(y)dy+\int \gamma(y)\gamma''(y)dy\right], \end{aligned}$$

$$\begin{aligned} D_{24} &:= K(0)^2\left[\int \delta(y)dy-\left(\int \gamma(y)dy\right)^2\right], \\ D_{25} &:= -2\left[\mu_0(K^2)+2K(0)\right]\int \alpha(y)dy, \\ D_{26} &:= 8\int \beta(y)dy=8D_{18}, \end{aligned}$$

with

$$\begin{aligned} \alpha(y) &:= \frac{f(y)^2}{g(y)^2}(v(y)-\mu_v)^2, \\ \delta(y) &:= \frac{f(y)^2}{g(y)^3}(v(y)-\mu_v)^2. \end{aligned}$$

Lemma 4.5.7. *The covariance between* $\widehat{A}_1$ *and* $\widehat{A}_2$ *is*

$$Cov\left(\widehat{A}_1, \widehat{A}_2\right) = D_{27}\frac{h^2}{N} + D_{28}\frac{h^4}{N} + O\left(\frac{h^6}{N}\right), \tag{4.18}$$

where

$$\begin{aligned} D_{27} &:= \frac{\mu_2(K)}{2}\int \theta(y) f''(y)dy, \\ D_{28} &:= \frac{\mu_4(K)}{24}\int \theta(y) f^{(4)}(y)dy. \end{aligned}$$

Lemma 4.5.8. *The covariance between* $\widehat{A}_1$ *and* $\widehat{A}_3$ *is*

$$\begin{aligned} Cov\left(\widehat{A}_1, \widehat{A}_3\right) &= D_{29}\frac{1}{N} + D_{30}\frac{1}{N^2 b} + D_{31}\frac{1}{N^2} + D_{32}\frac{b^2}{N} \\ &+ D_{33}\frac{b^4}{N} + O\left(\frac{b^2}{N^2}\right) + O\left(\frac{b^6}{N}\right), \end{aligned} \tag{4.19}$$

where

$$\begin{aligned} D_{29} &:= \int \beta(y)dy = D_{18}, \\ D_{30} &:= K(0)\int \alpha(y)dy, \\ D_{31} &:= -2\int \beta(y)dy = -2D_{18}, \\ D_{32} &:= \frac{\mu_2(K)}{2}\left[\int \alpha(y)g''(y)dy + \int \gamma''(y)f(y)(v(y)-\mu_v)dy\right], \\ D_{33} &:= \frac{\mu_4(K)}{24}\left[\int \alpha(y)g^{(4)}(y)dy + \int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy\right]. \end{aligned}$$

Lemma 4.5.9. *The covariance between* $\widehat{A}_2$ *and* $\widehat{A}_3$ *is*

$$\begin{aligned} Cov\left(\widehat{A}_2, \widehat{A}_3\right) &= D_{34}\frac{h^2}{N} + D_{35}\frac{h^2b^2}{N} + D_{36}\frac{h^4}{N} + D_{37}\frac{h^4b^2}{N} + D_{38}\frac{h^2b^4}{N} \\ &+ D_{39}\frac{h^2}{N^2 b} + D_{40}\frac{h^2}{N^2} + D_{41}\frac{h^4}{N^2 b} + D_{42}\frac{h^4}{N^2} + D_{43}\frac{h^2b^2}{N^2} + O\left(\frac{h^6}{N}\right) \\ &+ O\left(\frac{h^2b^6}{N}\right) + O\left(\frac{h^4b^4}{N}\right) + O\left(\frac{h^6}{N^2 b}\right) + O\left(\frac{h^2b^4}{N^2}\right) + O\left(\frac{h^4b^2}{N^2}\right), \end{aligned} \tag{4.20}$$

where

$$\begin{aligned}
D_{34} &:= \frac{\mu_2(K)}{2}\int \theta(y)f''(y)dy,\\
D_{35} &:= \frac{\mu_2(K)^2}{4}\left[\int \rho(y)f''(y)g''(y)dy + \int f''(y)\gamma''(y)(v(y)-\mu_v)dy\right.\\
&- \left. 2\int v''(y)f(y)dy\int \gamma(y)g''(y)dy\right],\\
D_{36} &:= \frac{\mu_4(K)}{24}\int \theta(y)f^{(4)}(y)dy,\\
D_{37} &:= \frac{\mu_2(K)\mu_4(K)}{48}\left[\int \rho(y)f^{(4)}(y)g''(y)dy + \int f^{(4)}(y)\gamma''(y)(v(y)-\mu_v)dy\right.\\
&- \left. 2\int v^{(4)}(y)f(y)dy\int \gamma(y)g''(y)dy\right],\\
D_{38} &:= \frac{\mu_2(K)\mu_4(K)}{48}\left[\int \rho(y)f''(y)g^{(4)}(y)dy + \int f''(y)\gamma^{(4)}(y)(v(y)-\mu_v)dy\right.\\
&- \left. 2\int v''(y)f(y)dy\int \gamma(y)g^{(4)}(y)dy\right],\\
D_{39} &:= \frac{\mu_2(K)K(0)}{2}\left[\int \rho(y)f''(y)dy - \int v''(y)f(y)dy\int \gamma(y)dy\right],\\
D_{40} &:= -\mu_2(K)\int \theta(y)f''(y)dy = -2D_{34},\\
D_{41} &:= \frac{\mu_4(K)K(0)}{24}\left[\int \rho(y)f^{(4)}(y)dy - \int v^{(4)}(y)f(y)dy\int \gamma(y)dy\right],\\
D_{42} &:= -\frac{\mu_4(K)}{12}\int \theta(y)f^{(4)}(y)dy = -2D_{36},\\
D_{43} &:= -D_{35}.
\end{aligned}$$

Lemma 4.5.10. *The variance of $\widehat{A}$ is*

$$\begin{aligned}
Var\left(\widehat{A}\right) &\simeq D_7\frac{1}{n} + D_8\frac{1}{Nn} - 4D_6\frac{1}{N^2} + D_{26}\frac{1}{N^3} + D_9\frac{1}{Nnh} + D_{44}\frac{1}{N^2b} + D_{24}\frac{1}{N^3b^2}\\
&+ D_{25}\frac{1}{N^3b} - 2D_{39}\frac{h^2}{N^2b} - 2D_{41}\frac{h^4}{N^2b} + D_{10}\frac{h^2}{n} + D_{11}\frac{h^4}{n} + D_{12}\frac{h^4}{N}\\
&+ D_{45}\frac{b^4}{N} - 2D_{35}\frac{h^2b^2}{N} - 2D_{37}\frac{h^4b^2}{N} - 2D_{38}\frac{h^2b^4}{N} + D_{13}\frac{h}{Nn} + D_{14}\frac{h^2}{Nn}\\
&+ D_{15}\frac{h^3}{Nn} + D_{23}\frac{b}{N^2} + 4D_{27}\frac{h^2}{N^2} + 4D_{28}\frac{h^4}{N^2} + 2D_{35}\frac{h^2b^2}{N^2} + O\left(\frac{h^6}{n}\right)\\
&+ O\left(\frac{b^6}{N}\right) + O\left(\frac{h^4b^4}{N}\right) + O\left(\frac{h^4}{Nn}\right) + O\left(\frac{b^2}{N^2}\right) + O\left(\frac{h^6}{N^2b}\right),
\end{aligned}$$

being

$$\begin{aligned} D_{44} &:= 2\mu_0(K^2)\int \alpha(y)dy, \\ D_{45} &:= \frac{\mu_2(K)^2}{4}\left[\int \delta(y)g''(y)^2dy - 4\left(\int \gamma(y)g''(y)dy\right)^2\right. \\ &+ \left.\int \gamma''(y)^2 g(y)dz + 2\int \gamma(y)\gamma''(y)g''(y)dy\right]. \end{aligned}$$

The proof of Theorem 4.2.1 is a consequence of Lemmas from 4.5.3 to 4.5.10, considering $C_1 := D_2$, $C_2 := D_1$, $C_3 := D_4$, $C_4 := D_7$, $C_5 := D_8$, $C_6 := -4D_6$, $C_7 := D_9$, $C_8 := D_{44}$, $C_9 := D_{10}$, $C_{10} := -2D_{39}$, $C_{11} := D_{12}$ and $C_{12} := -2D_{35}$.

4.5.2 Sketch of the proof of Theorem 4.2.2

The proof of Theorem 4.2.2 follows parallel lines to that of Theorem 4.2.1. The proofs of the following lemmas are available in Appendix A.1. Along this subsection, Conditions A1 and A8-A10 are assumed for Lemmas 4.5.11-4.5.18.

Using Lemma 4.5.1, in this case $\widehat{A}$ can be expressed as:

$$\widehat{A} = \widehat{A}_1^* + \widehat{A}_2^* - \widehat{A}_3^* - \widehat{A}_4^* + \widehat{A}_5^*,$$

where

$$\begin{aligned} \widehat{A}_1^* &:= \frac{1}{N}\sum_{i=1}^N \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v), \\ \widehat{A}_2^* &:= \frac{1}{N}\sum_{i=1}^N \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v), \\ \widehat{A}_3^* &:= \frac{1}{N}\sum_{i=1}^N \frac{(K_h * f)(Y_i)(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v), \\ \widehat{A}_4^* &:= \frac{1}{N}\sum_{i=1}^N \frac{(\hat{f}_h(Y_i) - (K_h * f)(Y_i))(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v), \\ \widehat{A}_5^* &:= \frac{1}{N}\sum_{i=1}^N \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_b * g)(Y_i)}\right)^2 (v(Y_i) - \mu_v), \end{aligned}$$

being $\widehat{A}_4^*$ and $\widehat{A}_5^*$ negligible terms. Thus we will consider $\widehat{A}^* := \widehat{A}_1^* + \widehat{A}_2^* - \widehat{A}_3^*$.

Lemma 4.5.11. *The expectation of* $\widehat{A}^*$ *is*

$$E(\widehat{A}^*) = D_1^* + D_2^*\frac{1}{N}, \tag{4.21}$$

where

$$\begin{aligned} D_1^* &:= \int \gamma^*(y)g(y)dy, \\ D_2^* &:= D_1^* - \frac{K(0)}{b}\int \varphi^*(y)g(y)dy, \end{aligned}$$

with

$$\begin{aligned} \gamma^*(y) &:= \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y)-\mu_v), \\ \varphi^*(y) &:= \frac{(K_h * f)(y)}{(K_b * g)(y)^2}(v(y)-\mu_v). \end{aligned}$$

Lemma 4.5.12. *The variance of* $\widehat{A}_1^*$ *is*

$$Var\left(\widehat{A}_1^*\right) = \frac{D_3^*}{N}, \tag{4.22}$$

where

$$D_3^* := \int \alpha^*(y)g(y)dy - D_1^{*2},$$

with

$$\alpha^*(y) := \frac{(K_h * f)(y)^2}{(K_b * g)(y)^2}(v(y)-\mu_v)^2.$$

Lemma 4.5.13. *The variance of* $\widehat{A}_2^*$ *is*

$$Var\left(\widehat{A}_2^*\right) = D_4^*\frac{1}{n} + D_5^*\frac{1}{Nn}, \tag{4.23}$$

where

$$\begin{aligned} D_4^* &:= \int\left(\int \frac{K_h(y-z)}{(K_b * g)(y)}(v(y)-\mu_v)g(y)dy - \int \gamma^*(y)g(y)dy\right)^2 f(z)dz, \\ D_5^* &:= \int \frac{((K_h)^2 * f)(y)}{(K_b * g)(y)^2}(v(y)-\mu_v)^2 g(y)dy - \int \alpha^*(y)g(y)dy - D_4^*. \end{aligned}$$

Lemma 4.5.14. *The variance of* $\widehat{A}_3^*$ *is*

$$Var\left(\widehat{A}_3^*\right) = D_6^*\frac{1}{N} + D_7^*\frac{1}{N^2} + D_8^*\frac{1}{N^3}, \tag{4.24}$$

where

$$D_6^* := \int\left[\int \varphi^*(y)K_b(y-z)g(y)dy - \int \gamma^*(y)g(y)dy\right]^2 g(z)dz,$$

$$\begin{aligned}
D_7^* &:= \frac{2K(0)}{b}\left[\int\int \varphi^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz\right.\\
&- \left.\int \gamma^*(y)g(y)dy\int \varphi^*(y)g(y)dy\right]\\
&- 4\int\int \gamma^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz + 3\left(\int \gamma^*(y)g(y)dy\right)^2\\
&- \int \alpha^*(y)g(y)dy + \int \varphi^*(y)^2((K_b)^2 * g)(y)g(y)dy\\
&+ \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)^2 g(y)g(z)dydz\\
&- 3\int\left[\int \varphi^*(y)K_b(y-z)g(y)dy - \int \gamma^*(y)g(y)dy\right]^2 g(z)dz,
\end{aligned}$$

$$\begin{aligned}
D_8^* &:= \frac{K(0)^2}{b^2}\left[\int \varphi^*(y)^2 g(y)dy - \left(\int \varphi^*(y)g(y)dy\right)^2\right]\\
&+ \frac{2K(0)}{b}\left[-\int\int \varphi^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz\right.\\
&+ \left.2\int \gamma^*(y)g(y)dy\int \varphi^*(y)g(y)dy - \int \delta^*(y)g(y)dy\right]\\
&+ 4\int\int \gamma^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz - 4\left(\int \gamma^*(y)g(y)dy\right)^2\\
&+ 2\int \alpha^*(y)g(y)dy - \int \varphi^*(y)^2((K_b)^2 * g)(y)g(y)dy\\
&- \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)^2 g(y)g(z)dydz\\
&+ 2\int\left[\int \varphi^*(y)K_b(y-z)g(y)dy - \int \gamma^*(y)g(y)dy\right]^2 g(z)dz.
\end{aligned}$$

Lemma 4.5.15. *The covariance between $\widehat{A}_1^*$ and $\widehat{A}_2^*$ is*

$$Cov\left(\widehat{A}_1^*, \widehat{A}_2^*\right) = 0.$$

Lemma 4.5.16. *The covariance between $\widehat{A}_1^*$ and $\widehat{A}_3^*$ is*

$$Cov\left(\widehat{A}_1^*, \widehat{A}_3^*\right) = D_9^*\frac{1}{N} + D_{10}^*\frac{1}{N^2}, \tag{4.25}$$

where

$$
\begin{aligned}
D_9^* &:= \int\int \gamma^*(z)\varphi^*(y)K_b(y-z)g(y)g(z)dydz - \left(\int \gamma^*(y)g(y)dy\right)^2, \\
D_{10}^* &:= \frac{K(0)}{b}\left[\int \delta^*(y)g(y)dy - \left(\int \varphi^*(y)g(y)dy\right)\left(\int \gamma^*(y)g(y)dy\right)\right] \\
&- \int \alpha^*(y)g(y)dy + 2\left(\int \gamma^*(y)g(y)dy\right)^2 \\
&- \int\int \gamma^*(z)\varphi^*(y)K_b(y-z)g(y)g(z)dydz.
\end{aligned}
$$

Lemma 4.5.17. *The covariance between $\widehat{A}_2^*$ and $\widehat{A}_3^*$ is*

$$
Cov\left(\widehat{A}_2^*, \widehat{A}_3^*\right) = 0.
$$

Lemma 4.5.18. *The variance of $\widehat{A}$ is*

$$
Var\left(\widehat{A}^*\right) = D_4^*\frac{1}{n} + D_{11}^*\frac{1}{N} + D_5^*\frac{1}{Nn} + D_{12}^*\frac{1}{N^2} + D_8^*\frac{1}{N^3},
$$

where

$$
\begin{aligned}
D_{11}^* &:= D_3^* + D_6^* - 2D_9^*, \\
D_{12}^* &:= D_7^* - 2D_{10}^*.
\end{aligned}
$$

The proof of Theorem 4.2.2 is a consequence of Lemmas from 4.5.11 to 4.5.18, considering $C_1^* := D_1^{*2}$, $C_2^* := D_4^*$, $C_3^* := 2D_1^*D_2^* + D_{11}^*$, $C_4^* := D_5^*$, $C_5^* := D_2^{*2} + D_{12}^*$ and $C_6^* := D_8^*$.

Chapter 5

Nonparametric Estimation in Setup 2

5.1 Introduction

In this chapter, the nonparametric estimation of the mean of a transformation of a continuous population in Setup 2 is considered. Following an analogous procedure to that presented in Chapter 2 for the known w case and in Chapter 4 for Setup 1, we present in Section 5.2 the density-based nonparametric estimator proposed for bias correction in Setup 2 and some asymptotic results. In Section 5.3, a simulation study is included.

5.2 Estimation for B3D in Setup 2

Let us consider the continuous population and the B3D context presented in Section 2.2.3 in Chapter 2.

Following parallel steps to those of that chapter, it is possible to define analogous estimators to $\tilde{\tilde{\mu}}^{(1)}$ and $\tilde{\tilde{\mu}}^{(2)}$, but for the general case of the mean of a transformation.

Using (2.7), we can observe that:

$$\begin{aligned} E\left(\frac{v(Y)}{w_2\,(Y)}\right) &= \int \frac{v(y)}{w_2\,(y)} g\,(y)\,dy = \int \frac{v(y)}{m(y)/g(y)} g\,(y)\,dy \\ &= \int \frac{v(y)}{cg(y)/f(y)} g\,(y)\,dy = \frac{1}{c}\int v(y) f\,(y)\,dy = \frac{1}{c}\mu_v. \end{aligned} \tag{5.1}$$

Equation (5.1) motivates the definition of a general estimator in Setup 2 for the known w_2 case:

$$\tilde{\tilde{\mu}}_v^{(1)} = \frac{1}{N}\sum_{i=1}^{N}\frac{v(Y_i)}{w_2\left(Y_i\right)} = \frac{1}{N}\sum_{i=1}^{N}\frac{g\left(Y_i\right)v(Y_i)}{m\left(Y_i\right)} = \frac{1}{c}\frac{1}{N}\sum_{i=1}^{N}T_i. \tag{5.2}$$

Since $\tilde{\tilde{\mu}}_v^{(1)}$ is the sample mean of the simple random sample $T_i = v(Y_i)/w_1(Y_i),\ i = 1,\ldots,N$, its properties as a good estimator of μ_v are straightforward:

$$E\left(\tilde{\tilde{\mu}}_v^{(1)}\right) = \frac{1}{c}\cdot\mu_v,$$

$$Var\left(\tilde{\tilde{\mu}}_v^{(1)}\right) = \frac{\sigma_T^2}{c^2\cdot N}$$

and

$$\sqrt{N}\ \ \frac{\tilde{\tilde{\mu}}_v^{(1)} - \frac{\mu_v}{c}}{\frac{\sigma_T}{c}} \to N(0,1),$$

where $\sigma_T^2 = \int v(x)^2 f(x)^2 g(x)^{-1}dx - \mu_v^2$.

Since the sample weights $g\left(Y_i\right)/m\left(Y_i\right)$ do not sum up to 1, a reasonable modification of $\tilde{\tilde{\mu}}_v^{(1)}$ is the following convex linear combination version:

$$\tilde{\tilde{\mu}}_v^{(2)} = \frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\frac{v(Y_i)}{w_2\left(Y_i\right)}}{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\frac{1}{w_2\left(Y_i\right)}} = \frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\frac{g\left(Y_i\right)v(Y_i)}{m\left(Y_i\right)}}{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\frac{g\left(Y_i\right)}{m\left(Y_i\right)}}. \tag{5.3}$$

The estimator in (5.3) can also be regarded as an empirical analogue of the expectation ratio

$$\frac{E\left(\frac{v(Y)}{w_2(Y)}\right)}{E\left(\frac{1}{w_2(Y)}\right)},$$

which is equal to μ_v, by just recalling equation (5.1) and

$$E\left(\frac{1}{w_2\left(Y\right)}\right) = \int\frac{1}{w_2\left(y\right)}g\left(y\right)dy = \frac{1}{c}\int f\left(y\right)dy = \frac{1}{c}.$$

Note that the estimators $\tilde{\tilde{\mu}}^{(1)}$ and $\tilde{\tilde{\mu}}^{(2)}$ proposed in Chapter 2 are particular cases of the estimators in (5.2) and (5.3) when considering v the identity function.

In general, the estimators in (5.2) and (5.3) are useless when the biasing function, w_2, is unknown. However, we can obtain completely observable versions of these estimators by estimating the densities involved, g and m.

The Parzen-Rosenblatt kernel density estimators (Parzen, 1962; Rosenblatt, 1956) based on the samples $\mathbf{Y}$ and $\mathbf{Z}$ can be used to estimate $g(x)$ and $m(x)$:

$$\begin{aligned} \hat{g}_b(x) &= \frac{1}{N}\sum_{i=1}^{N} K_b(x - Y_i) \\ \hat{m}_h(x) &= \frac{1}{n}\sum_{i=1}^{n} K_h(x - Z_i) \end{aligned}$$

where $K_h(u) = (1/h)K(u/h)$, being K a kernel function and h and b two bandwidths. The biasing function, w_2, can be easily estimated as the ratio of both estimated densities: $\hat{w}_{2,h,b}(x) = \hat{m}_h(x)/\hat{g}_b(x)$.

Plugging these estimators into (5.2) and (5.3) leads to observable versions of $\tilde{\hat{\mu}}_v^{(1)}$ and $\tilde{\hat{\mu}}_v^{(2)}$:

$$\hat{\hat{\mu}}_v^{1,h,b} = \frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)}{\hat{w}_{2,h,b}(Y_i)} = \frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)},$$

$$\hat{\hat{\mu}}_v^{2,h,b} = \frac{\frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)}{\hat{w}_{2,h,b}(Y_i)}}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{\hat{w}_{2,h,b}(Y_i)}} = \frac{\frac{1}{N}\sum_{i=1}^{N} v(Y_i)\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}}{\frac{1}{N}\sum_{i=1}^{N} \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}}. \tag{5.4}$$

From now on, only the estimator $\hat{\hat{\mu}}_v^{2,h,b}$ for Setup 2 and its analogous version for the mean, $\hat{\hat{\mu}}^{2,h,b}$, will be considered. For the sake of brevity these estimators are denoted by $\hat{\hat{\mu}}_v$ and $\hat{\hat{\mu}}$, respectively, in the rest of the chapter, except when a more explicit notation is needed.

5.2.1 Asymptotic results

To obtain the asymptotic mean squared error of $\hat{\hat{\mu}}_v$ in Setup 2 we need the following two assumptions:

A11. The density functions f, g and m are six times differentiable and all their first six derivatives are bounded. Additionally f is bounded and g and m are bounded away from zero.

A12. The following integrals are finite:

$$\int \Omega(y)^2 m(y)dy < \infty$$
$$\int \Omega^{(k)}(y)^2 m(y)dy < \infty \quad \text{for} \quad k = 2, 4, 6$$
$$\int (v(y) - \mu_v)^2 |m^{(k)}(y)|dy < \infty \quad \text{for} \quad k = 2, 4, 6,$$

with

$$\Omega(y) := \frac{f(y)^2}{g(y)^2}(v(y) - \mu_v).$$

Theorem 5.2.1. *Under the classical conditions on the bandwiths and the sample sizes, i.e. $h \to 0$, $b \to 0$, $nh \to \infty$, $Nb \to \infty$ and $N/n \to \infty$, if Conditions A1, A11 and A12 are fulfilled, then the asymptotic mean squared error of $\hat{\hat{\mu}}_v$ is*

$$AMSE\left(\hat{\hat{\mu}}_v^{2,h,b}\right) = \left(C_1^{\bullet} b^2 + \frac{C_2^{\bullet}}{Nb} + C_3^{\bullet} h^2\right)^2 + \frac{C_4^{\bullet}}{n} + \frac{C_5^{\bullet}}{Nn} + \frac{C_{13}^{\bullet}}{N} + \frac{C_6^{\bullet}}{N^2} + \frac{C_7^{\bullet}}{Nnh}$$
$$+ \frac{C_8^{\bullet}}{N^2 b} + \frac{C_9^{\bullet} h^2}{n} + \frac{C_{14}^{\bullet} h^2}{N} + \frac{C_{10}^{\bullet} h^2}{N^2 b} + \frac{C_{11}^{\bullet} h^4}{N} + \frac{C_{15}^{\bullet} b^2}{N} + \frac{C_{12}^{\bullet} h^2 b^2}{N},$$

where the first three terms come from the squared bias and the rest of them from the variance of the estimator. The constants $C_1^{\bullet}, \ldots, C_{15}^{\bullet}$ are defined in the sketch of the proofs (Subsection 5.5.1).

The simulations results in Section 5.3 below, show a very good performance of the proposed mean estimator in Setup 2. However, Table 5.1 motivates to study theoretically the asymptotic properties of $\hat{\hat{\mu}}_v$ under the non-standard conditions that the two bandwidths, h and b, tend to positive constants when the sample size tends to infinity.

The asymptotic mean squared error for the estimator under these non-standard conditions is presented in Theorem 5.2.2 below. Three assumptions are needed.
A13. The kernel K is bounded.
A14. The density function m is bounded away from zero.
A15. The integral $\int (v(y) - \mu_v)^2 g(y)dy$ is finite.

Theorem 5.2.2. *Let us assume $h \to h_0 > 0$, $b \to b_0 > 0$, $n \to \infty$, $N/n \to \infty$, and Conditions A1 and A13-A15. The asymptotic mean squared error for the estimator $\hat{\hat{\mu}}_v$ in (5.4) is given by*

$$AMSE\left(\hat{\hat{\mu}}_v^{2,h,b}\right) = C_1^{*\bullet} + \frac{C_2^{*\bullet}}{n} + \frac{C_3^{*\bullet}}{N} + \frac{C_4^{*\bullet}}{Nn} + \frac{C_5^{*\bullet}}{N^2} + \frac{C_6^{*\bullet}}{N^3},$$

where the first two constants are

$$
\begin{aligned}
C_1^{*\bullet} &= \left(\frac{1}{c} \int \frac{K_b * g(y)}{K_h * m(y)} (v(y) - \mu_v) g(y) dy \right)^2 , \\
C_2^{*\bullet} &= \frac{1}{c^2} \int \left(\int \frac{K_b * g(y)^2 (v(y) - \mu_v)^2}{K_h * m(y)^4} K_h(y - z) g(y) dy - C_1^{*\bullet 1/2} \right)^2 m(z) dz
\end{aligned}
$$

and $C_3^{*\bullet}$, $C_4^{*\bullet}$, $C_5^{*\bullet}$ and $C_6^{*\bullet}$ are defined in the sketch of the proofs (Subsection 5.5.2)

5.3 Simulations

The performance of the mean estimator, $\hat{\hat{\mu}}$, proposed in Section 5.2, is studied via simulation.

We consider again the population presented in Subsection 3.3 and the class of weight functions described in (3.8). We simulate the sample $\mathbf{Y}$ from the biased density g characterized in (3.9) and (3.10) and the sample $\mathbf{Z}$ from the twice biased density m defined in (3.11) and (3.12).

We generated 10^3 pairs of datasets, each with sample size $n = 10^3$ in the case of the sample $\mathbf{Z}$ and sample size $N = 10^6$ for the sample $\mathbf{Y}$. Different combinations of k and ε are considered (see Figure 5.1).

Table 5.1: Comparison of the MSE of $\overline{Z}$, $\overline{Y}$ and $\hat{\hat{\mu}}$ for different values of k and ε ($n = 10^3$, $N = 10^6$, trials=10^3). h_{opt} and b_{opt} refer to the optimal bandwidths obtained by simulation.

k	ε	$MSE(\overline{Z})$	$MSE(\overline{Y})$	$MSE(\hat{\hat{\mu}})$	h_{opt}	b_{opt}
1.5	0.1	$1.6 \cdot 10^{-1}$	$8.1 \cdot 10^{-2}$	$1.8 \cdot 10^{-4}$	0.49	12.33
1.5	1.5	$1.6 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$	$1.4 \cdot 10^{-5}$	1.19	15.28
1	1.5	$7.2 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$8.0 \cdot 10^{-6}$	1.36	16.86
0.5	1.5	$1.9 \cdot 10^{-3}$	$3.9 \cdot 10^{-4}$	$2.8 \cdot 10^{-6}$	1.78	30.18
0.1	1.5	$3.4 \cdot 10^{-4}$	$1.5 \cdot 10^{-5}$	$4.4 \cdot 10^{-7}$	3.66	33.81
0.1	1.8	$2.9 \cdot 10^{-4}$	$8.2 \cdot 10^{-7}$	$3.1 \cdot 10^{-7}$	8.30	44.04

Table 5.1 shows the good performance of $\hat{\hat{\mu}}$ when compared with the classical sample means. As expected, it is preferable to use the mean of the biased sample $\mathbf{Y}$ due to the larger sample size, rather than the $\mathbf{Z}$ sample mean since it is twice biased. Anyhow, $\hat{\hat{\mu}}$ outperforms $\mathbf{Y}$ in all the situations. It is also observed how the optimal bandwidths obtained by simulation, h_{opt} and b_{opt}, seem to contradict the classical condition considered in Theorem 5.2.1 that they tend to zero when the sample sizes

tend to infinity.

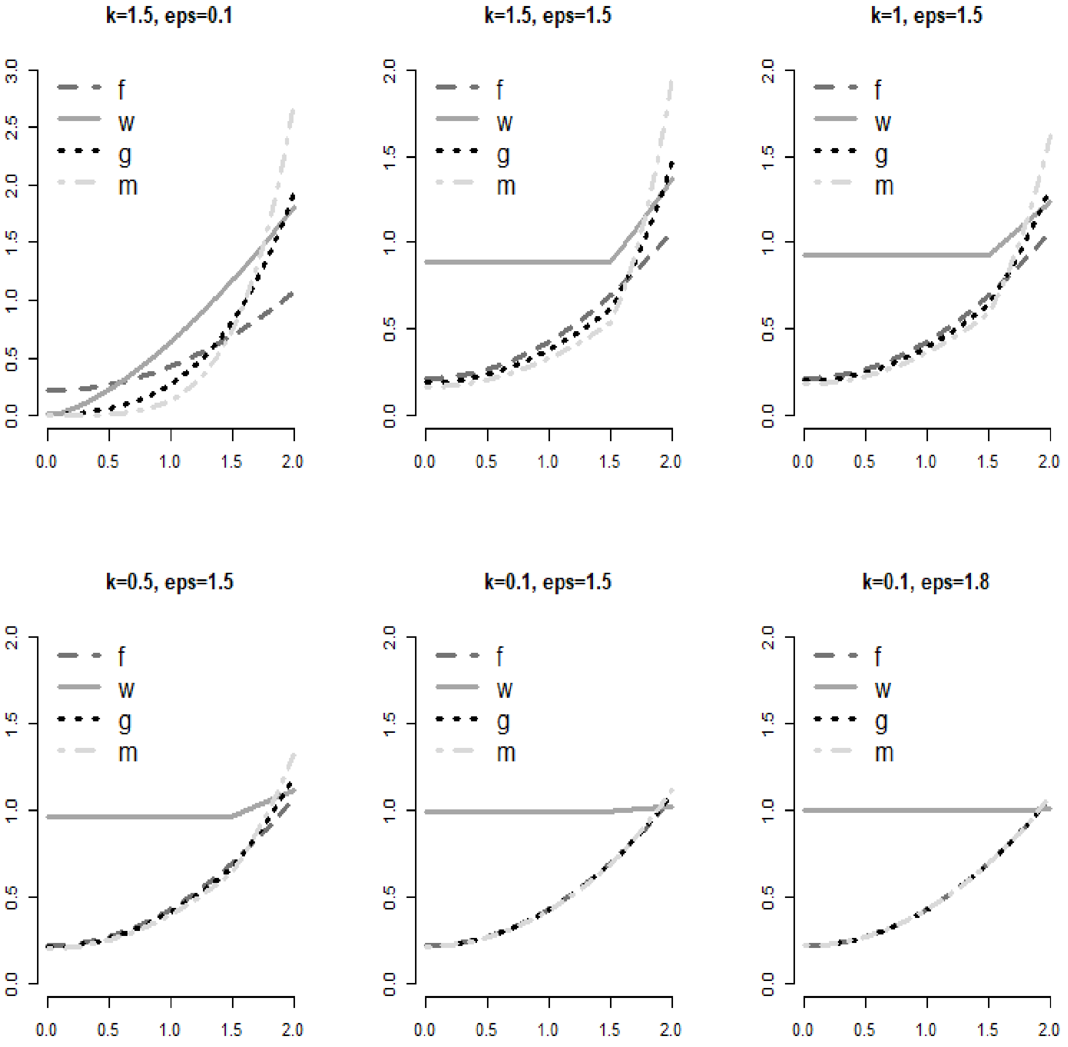

Figure 5.1: Densities g (dotted black line) and m (dashed gray line) involved in the simulated models for different values of k and ε in Setup 2 for the biasing function, w (solid gray line).

Figure 5.2 shows how, by choosing a suitable h, our proposed estimator performs very well for a very wide range of values for the smoothing parameter b.

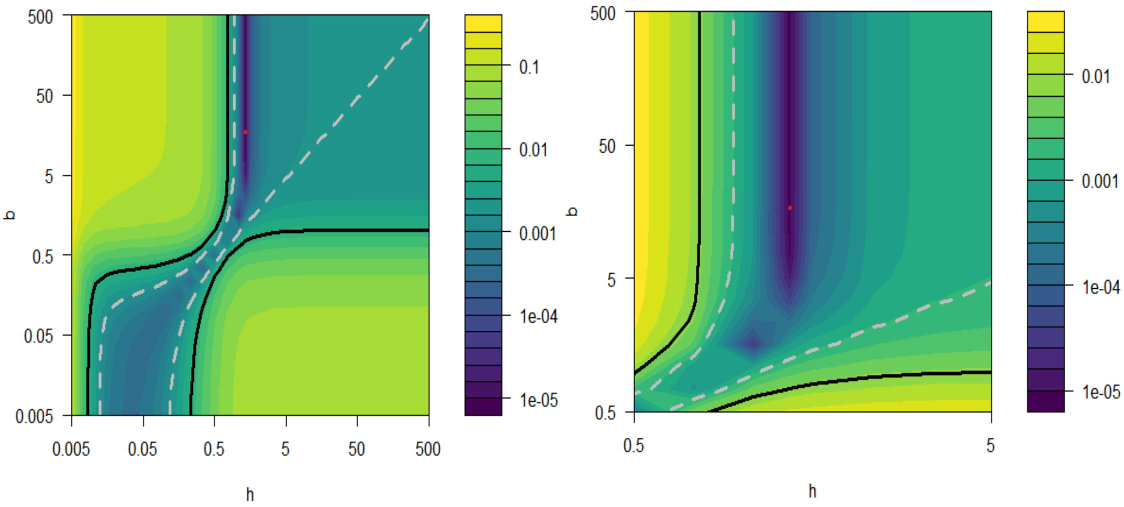

Figure 5.2: Comparison of the MSE of the proposed estimator as a function of h and b with the MSE of $\overline{Z}$ (solid black line) and the MSE of $\overline{Y}$ (dashed gray line) for the particular choice of $k = 1$ and $\varepsilon = 1.5$, with $n = 10^3$ and $N = 10^6$. The red dot represents the minimal value of the MSE.

5.3.1 Asymptotically-based simulations

To further study the properties of the proposed estimator, $\hat{\hat{\mu}}$, we carried out another simulation study. In order to explore the optimal values of the bandwidths obtained by simulation, we analyze how the optimal bandwidths behave when progressively increasing the sample sizes in two different settings, $k = 1.5$ and $\varepsilon = 0.1$ and $k = 1$ and $\varepsilon = 1.5$.

In the case of the smoothing parameter h, the results in Tables 5.2 and 5.3 show that the optimal values obtained by simulation, h_{opt}, do not tend to zero, but tend to constant values in both settings considered.

The values obtained in Tables 5.2 and 5.3 for the optimal b, b_{opt}, apart from taking enormously high values, do not seem to tend to constant values. As Figure 5.2 showed, considering the optimal value for bandwidth h, the MSE of the proposed estimator is, from a certain value, a very flat function of the smoothing parameter b. Thus the proposed estimator is rather stable for a wide range of values for bandwidth b.

Figure 5.3 shows how, from a certain value of b, the value of the MSE hardly changes. Setting the value of the optimal h and taking a bandwidth b, b'_{opt}, in which the value of the MSE, $MSE_{b'_{opt}}$, is just one millionth greater than the value of the MSE in the optimal, $MSE(\hat{\mu})$, we observe that this new b'_{opt}, with a much smaller value and giving approximately the same error, seems to tend to a constant value in both settings.

Table 5.2: Comparison of the MSE of $\overline{Z}$, $\overline{Y}$ and $\hat{\mu}^{h_{opt},b_{opt}}$ for different sample sizes and the choice of $k = 1.5$ and $\varepsilon = 0.1$ in Setup 2. h_{opt} and b_{opt} refer to the optimal values obtained by simulation. b'_{opt} refers to the smallest value of the bandwidth b in which the value of the MSE, $MSE_{b'_{opt}}$, is less than or equal to one thousandth greater than in the optimal b_{opt}, considering the same h_{opt}.

n	N	$MSE(\overline{Z})$	$MSE(\overline{Y})$	$MSE(\hat{\mu})$	h_{opt}	b_{opt}	$MSE_{b'_{opt}}$	b'_{opt}
10	100	0.163780	0.082726	$2.5 \cdot 10^{-2}$	0.58	10^{29}	$2.5 \cdot 10^{-2}$	11.67
20	400	0.162581	0.081857	$1.2 \cdot 10^{-2}$	0.54	10^{19}	$1.2 \cdot 10^{-2}$	16.56
50	$2,500$	0.159506	0.081642	$4.3 \cdot 10^{-2}$	0.51	10^{9}	$4.3 \cdot 10^{-2}$	21.83
100	$10,000$	0.159352	0.081691	$2.2 \cdot 10^{-3}$	0.50	10^{11}	$2.2 \cdot 10^{-3}$	24.38
200	$40,000$	0.158226	0.081697	$9.7 \cdot 10^{-4}$	0.49	10^{15}	$9.7 \cdot 10^{-4}$	25.29
500	$250,000$	0.158147	0.081684	$4.2 \cdot 10^{-4}$	0.49	10^{26}	$4.2 \cdot 10^{-4}$	26.24
10^3	10^6	0.157936	0.081661	$1.8 \cdot 10^{-4}$	0.49	10^{15}	$1.8 \cdot 10^{-4}$	21.04

Table 5.3: Comparison of the MSE of $\overline{Z}$, $\overline{Y}$ and $\hat{\mu}^{h_{opt},b_{opt}}$ for different sample sizes and the choice of $k = 1$ and $\varepsilon = 1.5$ in Setup 2. h_{opt} and b_{opt} refer to the optimal values obtained by simulation. b'_{opt} refers to the smallest value of the bandwidth b in which the value of the MSE, $MSE_{b'_{opt}}$, is less than or equal to one thousandth greater than in the optimal b_{opt}, considering the same h_{opt}.

n	N	$MSE(\overline{Z})$	$MSE(\overline{Y})$	$MSE(\hat{\mu})$	h_{opt}	b_{opt}	$MSE_{b'_{opt}}$	b'_{opt}
10	100	$3.5 \cdot 10^{-2}$	$4.5 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$	1.91	10^{14}	$4.0 \cdot 10^{-3}$	26.24
20	400	$2.2 \cdot 10^{-2}$	$2.4 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	1.54	10^{32}	$1.2 \cdot 10^{-3}$	32.73
50	$2,500$	$1.3 \cdot 10^{-2}$	$1.7 \cdot 10^{-3}$	$2.8 \cdot 10^{-4}$	1.44	10^{15}	$2.8 \cdot 10^{-4}$	40.83
100	$10,000$	$1.0 \cdot 10^{-2}$	$1.7 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$	1.40	10^{30}	$1.1 \cdot 10^{-4}$	40.83
200	$40,000$	$8.5 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$4.7 \cdot 10^{-5}$	1.38	10^{14}	$4.7 \cdot 10^{-5}$	46.45
500	$250,000$	$7.6 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$1.8 \cdot 10^{-5}$	1.38	10^{28}	$1.8 \cdot 10^{-5}$	79.24
10^3	10^6	$7.2 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$8.0 \cdot 10^{-6}$	1.36	10^{17}	$8.0 \cdot 10^{-6}$	29.31

This motivates studying the asymptotic behavior of the estimator under the non-standard conditions that the two bandwidths (h and b) tend to positive constants when n tends to infinity (see Theorem 5.2.2 above).

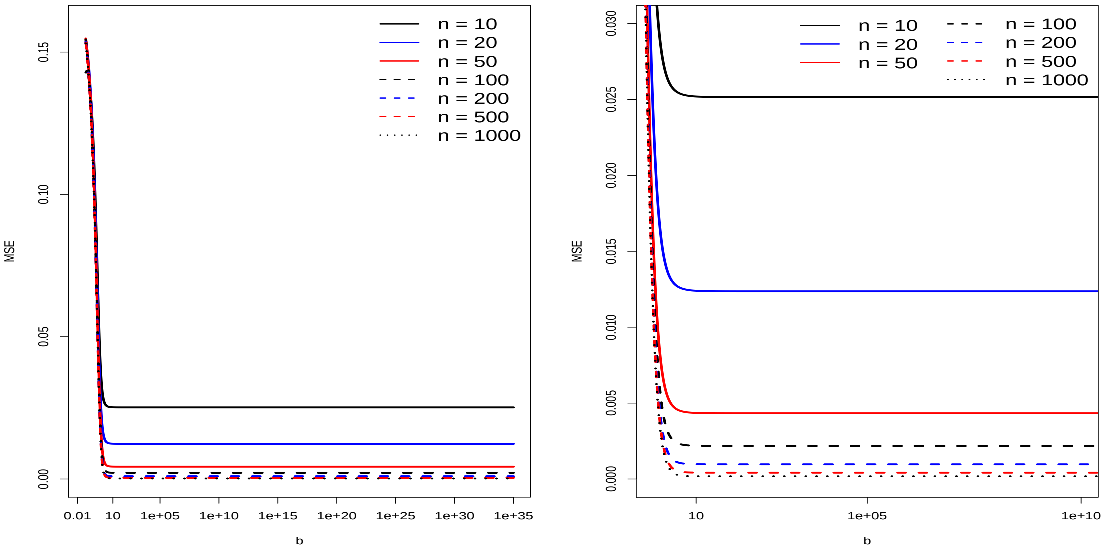

Figure 5.3: Comparison of the MSE of the proposed estimator as a function of b for the choice $K = 1.5$ and $\varepsilon = 0.1$ and different samples sizes of n and $N = n^2$, using a suitable h for each case.

5.4 Bootstrap algorithm

Since the sample $\mathbf{X}$ is not available in Setup 2, in this section we proposed a new version of the bootstrap algorithm presented for Setup 1 (Subsection 4.4).

Considering relation (2.6) we can express:

$$\frac{m(x)/g(x)}{g(x)/f(x)} = \frac{m(x) \cdot f(x)}{g(x)^2} = c,$$

for some constant c, i.e.

$$f(x) = c \cdot \frac{g(x)^2}{m(x)}.$$

Since

$$1 = \int f(x)dx = \int c \cdot \frac{g(x)^2}{m(x)}dx = c \cdot \int \frac{g(x)^2}{m(x)}dx,$$

then

$$c = \frac{1}{\displaystyle\int \frac{g(x)^2}{m(x)}dx}.$$

The bootstrap method proceeds as follows:

1. Based on the original **Z** and **Y** samples, the estimated densities $\hat{m}_{h_{pil}}$ and $\hat{g}_{b_{pil}}$, where h_{pil} and b_{pil} denote suitable pilot bandwidths, are considered as the true population densities in the bootstrap world.

2. The value of c is estimated using $\hat{m}_{h_{pil}}$ and $\hat{g}_{b_{pil}}$:
$$\hat{c} = \frac{1}{\displaystyle\int \frac{\hat{g}_{b_{pil}}(x)^2}{\hat{m}_{h_{pil}}(x)}dx},$$
which allows to obtain estimations of the density f:
$$\hat{f}_{h_{pil},b_{pil}}(x) = \frac{\dfrac{\hat{g}_{b_{pil}}(x)^2}{\hat{m}_{h_{pil}}(x)}}{\displaystyle\int \frac{\hat{g}_{b_{pil}}(x)^2}{\hat{m}_{h_{pil}}(x)}dx}$$
and the bootstrap population mean:
$$\mu^* = \int x \cdot \hat{f}_{h_{pil},b_{pil}}(x)dx.$$

3. Bootstrap resamples, $\mathbf{Z}^* = (Z_1^*, \ldots, Z_n^*)$ and $\mathbf{Y}^* = (Y_1^*, \ldots, Y_N^*)$, of sizes n and N respectively, are obtained from the estimated densities $\hat{m}_{h_{pil}}$ and $\hat{g}_{b_{pil}}$ as follows:

   (a) $Z_i^* = \psi_i^* + h_{pil} \cdot U_i$, where $\psi^* = (\psi_1^*, \ldots, \psi_n^*)$ is a simple random sample obtained from the empirical distribution computed with the values $\mathbf{Z} = (Z_1, \ldots, Z_n)$ and $U = (U_1, \ldots, U_n)$, with U_i simulated from the density K (a $N(0,1)$ when considering a Gaussian kernel), for $i = 1, \ldots, n$.

   (b) $Y_i^* = \eta_i^* + b_{pil} \cdot V_i$, where $\eta^* = (\eta_1^*, \ldots, \eta_N^*)$ is a simple random sample obtained from the empirical distribution computed with the values $\mathbf{Y} = (Y_1, \ldots, Y_N)$ and $V = (V_1, \ldots, V_N)$, with V_i simulated from the density K (a $N(0,1)$ when considering a Gaussian kernel), for $i = 1, \ldots, N$.

4. The estimator $\hat{\hat{\mu}}^{2,h,b,*}$ is computed using the resamples $\mathbf{Z}^*$ and $\mathbf{Y}^*$ and considering a very wide range of values for the smoothing parameters h and b.

5. Steps 2 and 3 are repeated a large number of times, B, in order to obtain an approximation of the bootstrap mean squared error (MSE^*) of the estimator,
$$MSE^*(h,b) = \frac{1}{B}\sum_{j=1}^{B}\left(\hat{\hat{\mu}}_j^{2,h,b,*} - \mu^*\right)^2.$$

6. The bandwidths h^* and b^* that minimize the function $MSE^*(h,b)$ are considered as bootstrap bandwidth selectors.

5.4.1 Simulations

The performance of the bootstrap algorithm is studied via simulation. We generated 100 pairs of datasets, each with sample size $n = 10^3$ in the case of the sample $\mathbf{Z}$ and sample size $N = 10^6$ for the sample $\mathbf{Y}$. For each pair of generated samples, an iterative search of the bootstrap selectors is performed. This search consists of applying the bootstrap algorithm 3 times using 100 resamples $\mathbf{Z}^*$ and 100 resamples $\mathbf{Y}^*$ each time. In each search, a grid of size 5 is used for each bandwidth (25 combinations of bandwidths), progressively reducing the area of this grid in each iteration of the method.

The choice of the pilot bandwidths has a certain influence on the behavior of our estimator. In this case we have used the bandwidths obtained by the Seather-Jones method.

Table 5.4: Comparison of the MSE of $\overline{Z}$, $\overline{Y}$ and $\hat{\hat{\mu}}^{2,h^*,b^*}$ for different values of k and ε ($n = 10^3$, $N = 10^6$, trials=100, $B = 100$). $Median(h^*)$ and $Median(b^*)$ refer to the median of the bandwidths obtained using the bootstrap algorithm.

k	ε	$MSE(\overline{Z})$	$MSE(\overline{Y})$	$MSE(\hat{\hat{\mu}}^{2,h^*,b^*})$	$Median(h^*)$	$Median(b^*)$
1.5	0.1	$1.58 \cdot 10^{-1}$	$8.17 \cdot 10^{-2}$	$2.35 \cdot 10^{-2}$	0.49	$3.90 \cdot 10^7$
1.5	1.5	$1.68 \cdot 10^{-2}$	$3.81 \cdot 10^{-3}$	$4.20 \cdot 10^{-4}$	1.19	$2.04 \cdot 10^5$
1	1.5	$7.26 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$3.74 \cdot 10^{-4}$	1.36	$7.11 \cdot 10^7$
0.5	1.5	$1.93 \cdot 10^{-3}$	$3.97 \cdot 10^{-4}$	$2.20 \cdot 10^{-4}$	2.37	$1.70 \cdot 10^4$
0.1	1.5	$3.52 \cdot 10^{-4}$	$1.61 \cdot 10^{-5}$	$9.35 \cdot 10^{-5}$	$5.08 \cdot 10^2$	29.58
0.1	1.8	$2.92 \cdot 10^{-4}$	$7.92 \cdot 10^{-7}$	$1.99 \cdot 10^{-4}$	$1.48 \cdot 10^2$	4.40

Table 5.4 contains the true values of the MSE of the classical estimators, the median of the bootstrap bandwidths h^* and b^* obtained from the algorithm presented above and the mean squared error of the proposed mean estimator. The results in this table show how, in the first four settings considered (more biased settings), the estimator obtained with the bootstrap bandwidths outperforms the two classical sample means, $\overline{Z}$ and $\overline{Y}$. In the last two cases (less biased settings), our estimator for the mean has a slightly worse behavior than $\overline{Y}$, which is reasonable given that bias in these situations is practically imperceptible and the size of sample $\mathbf{Y}$ is very large ($N = 10^6$).

In conclusion, in situations where the existence of bias is confirmed, the bootstrap algorithm will be used. This will provide a better estimation than the one provided by the samples involved. On the contrary, when we reject the presence of bias, we could directly use the estimator based on the sample $\mathbf{Y}$.

5.5 Sketch of the proofs

5.5.1 Sketch of the proof of Theorem 5.2.1

A set of lemmas needed to proof Theorem 5.2.1 is listed below. Their detailed proofs can be found in Appendix A.2. Along this subsection, Conditions A1, A11 and A12 are assumed for Lemmas 5.5.1-5.5.10.

Lemma 5.5.1. *The difference $\hat{\hat{\mu}}_v - \mu_v$ can be expressed as follows*

$$\hat{\hat{\mu}}_v - \mu_v = \frac{\widehat{A}^\bullet}{\widehat{B}^\bullet} \simeq \frac{\widehat{A}^\bullet}{c}, \tag{5.5}$$

where

$$\widehat{A}^\bullet = \frac{1}{N}\sum_{i=1}^N \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}(v(Y_i) - \mu_v) \tag{5.6}$$

and

$$\widehat{B}^\bullet = \frac{1}{N}\sum_{i=1}^N \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}, \tag{5.7}$$

considering c given by the relation

$$\frac{g(y)}{m(y)} = c \cdot \frac{f(y)}{g(y)}. \tag{5.8}$$

The term in (5.6) can be splitted into different terms, $\widehat{A}^\bullet = \widehat{A}_1^\bullet - \widehat{A}_2^\bullet + \widehat{A}_3^\bullet - \widehat{A}_4^\bullet + \widehat{A}_5^\bullet$, where

$$\begin{aligned}
\widehat{A}_1^\bullet &:= \frac{1}{N}\sum_{i=1}^N \frac{g(Y_i)}{m(Y_i)}(v(Y_i) - \mu_v),\\
\widehat{A}_2^\bullet &:= \frac{1}{N}\sum_{i=1}^N \frac{g(Y_i)(\hat{m}_h(Y_i) - m(Y_i))}{m(Y_i)^2}(v(Y_i) - \mu_v),\\
\widehat{A}_3^\bullet &:= \frac{1}{N}\sum_{i=1}^N \frac{\hat{g}_b(Y_i) - g(Y_i)}{m(Y_i)}(v(Y_i) - \mu_v),\\
\widehat{A}_4^\bullet &:= \frac{1}{N}\sum_{i=1}^N \frac{(\hat{g}_b(Y_i) - g(Y_i))(\hat{m}_h(Y_i) - m(Y_i))}{m(Y_i)^2}(v(Y_i) - \mu_v),\\
\widehat{A}_5^\bullet &:= \frac{1}{N}\sum_{i=1}^N \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\left(\frac{\hat{m}_h(Y_i) - m(Y_i)}{m(Y_i)}\right)^2 (v(Y_i) - \mu_v).
\end{aligned}$$

Since the terms $\widehat{A}_4^\bullet$ and $\widehat{A}_5^\bullet$ have some factors of quadratic nature whitin the sum (i.e. $(\hat{g}_b(Y_i) - g(Y_i))(\hat{m}_h(Y_i) - m(Y_i))$ and $(\hat{m}_h(Y_i) - m(Y_i))^2$) it is expected that one could prove negligibility of this terms. Thus we will consider

$$\widehat{A}^\bullet \simeq \widehat{A}_1^\bullet - \widehat{A}_2^\bullet + \widehat{A}_3^\bullet.$$

Lemma 5.5.2. *The expectation and variance of $\widehat{A}^\bullet$ can be approximated by*

$$\begin{aligned} E\left(\widehat{A}^\bullet\right) &\simeq E\left(\widehat{A}_1^\bullet\right) - E\left(\widehat{A}_2^\bullet\right) + E\left(\widehat{A}_3^\bullet\right), \qquad (5.9)\\ Var\left(\widehat{A}^\bullet\right) &\simeq Var\left(\widehat{A}_1^\bullet\right) + Var\left(\widehat{A}_2^\bullet\right) + Var\left(\widehat{A}_3^\bullet\right)\\ &- 2Cov\left(\widehat{A}_1^\bullet, \widehat{A}_2^\bullet\right) + 2Cov\left(\widehat{A}_1^\bullet, \widehat{A}_3^\bullet\right) - 2Cov\left(\widehat{A}_2^\bullet, \widehat{A}_3^\bullet\right). \qquad (5.10)\end{aligned}$$

The proof of Theorem 5.2.1 proceeds by analyzing the expectations and variances involved.

Lemma 5.5.3. *The expectation of $\widehat{A}^\bullet$ is*

$$\begin{aligned} E(\widehat{A}^\bullet) &\simeq D_1^\bullet \frac{1}{Nb} + D_2^\bullet b^2 + D_3^\bullet b^4 - D_2^\bullet \frac{b^2}{N} - D_3^\bullet \frac{b^4}{N}\\ &+ D_4^\bullet h^2 + D_5^\bullet h^4 + O(b^6) + O(h^6), \qquad (5.11)\end{aligned}$$

where

$$\begin{aligned} D_1^\bullet &:= K(0)c\int \gamma(y)dy,\\ D_2^\bullet &:= \frac{\mu_2(K)}{2}c\int \gamma(y)g''(y)dy,\\ D_3^\bullet &:= \frac{\mu_4(K)}{24}c\int \gamma(y)g^{(4)}dy,\\ D_4^\bullet &:= -\frac{\mu_2(K)}{2}c^2 B^\bullet\left(\Omega''\right),\\ D_5^\bullet &:= -\frac{\mu_4(K)}{24}c^2 B^\bullet(\Omega^{(4)}),\end{aligned}$$

the operator $B^\bullet$ is defined by

$$B^\bullet(\phi) := \int \phi(x)m(x)dx,$$

$$\gamma(y) := \frac{f(y)}{g(y)}(v(y) - \mu_v)$$

and

$$\Omega(y) := \frac{f(y)^2}{g(y)^2}(v(y) - \mu_v).$$

Lemma 5.5.4. *The variance of* $\widehat{A}_1^\bullet$ *is*

$$Var\left(\widehat{A}_1^\bullet\right) = \frac{D_6^\bullet}{N},$$

where

$$D_6^\bullet := c^2 \int \beta(y)dy$$

with

$$\beta(y) := \frac{f(y)^2}{g(y)}(v(y) - \mu_v)^2.$$

Lemma 5.5.5. *The variance of* $\widehat{A}_2^\bullet$ *is*

$$\begin{aligned} Var\left(\widehat{A}_2^\bullet\right) &= D_7^\bullet \frac{1}{n} + D_8^\bullet \frac{1}{Nn} + D_9^\bullet \frac{1}{Nnh} + D_{10}^\bullet \frac{h^2}{n} + D_{11}^\bullet \frac{h^4}{n} + D_{12}^\bullet \frac{h^4}{N} + D_{13}^\bullet \frac{h}{Nn} \\ &+ D_{14}^\bullet \frac{h^2}{Nn} + D_{15}^\bullet \frac{h^3}{Nn} + D_{16}^\bullet \frac{h^6}{n} + D_{17}^\bullet \frac{h^6}{N} + O\left(\frac{h^8}{n}\right) + O\left(\frac{h^4}{Nn}\right), \end{aligned} \tag{5.12}$$

where

$$\begin{aligned} D_7^\bullet &:= c^4 B^\bullet(\Omega^2), \\ D_8^\bullet &:= -D_6^\bullet - D_7^\bullet, \\ D_9^\bullet &:= \mu_0(K^2)c^3 \int \psi(y)dy, \\ D_{10}^\bullet &:= \mu_2(K)c^4 B^\bullet(\Omega \cdot \Omega''), \\ D_{11}^\bullet &:= \frac{\mu_2(K)^2}{4} c^4 \left[B^\bullet(\Omega''^2) - B^\bullet(\Omega'')^2\right] + \frac{\mu_4(K)}{12} c^4 B^\bullet(\Omega \cdot \Omega^{(4)}), \\ D_{12}^\bullet &:= \frac{\mu_2(K)^2}{4} c^4 \left[\int \xi(y) m''(y)^2 dy - B^\bullet(\Omega'')^2\right], \\ D_{13}^\bullet &:= \frac{\mu_2(K^2)}{2} c^4 \int \xi(y) m''(y) dy, \\ D_{14}^\bullet &:= -\mu_2(K)c^4 B^\bullet(\Omega \cdot \Omega'') - \mu_2(K)c^3 \int \psi(y) m''(y) dy, \\ D_{15}^\bullet &:= \frac{\mu_4(K^2)}{24} c^4 \int \xi(y) m^{(4)}(y) dy, \\ D_{16}^\bullet &:= \frac{\mu_2(K)\mu_4(K)}{24} c^4 \left[B^\bullet(\Omega'' \cdot \Omega^{(4)}) - B^\bullet(\Omega'')B^\bullet(\Omega^{(4)})\right] + \frac{\mu_6(K)}{360} c^4 B^\bullet(\Omega \cdot \Omega^{(6)}), \\ D_{17}^\bullet &:= \frac{\mu_2(K)\mu_4(K)}{24} c^4 \left[\int \xi(y) m''(y) m^{(4)}(y) dy - B^\bullet(\Omega'')B^\bullet(\Omega^{(4)})\right], \end{aligned}$$

being

$$\begin{aligned} \psi(y) &:= \frac{f(y)^3}{g(y)^3}(v(y) - \mu_v)^2, \\ \xi(y) &:= \frac{f(y)^4}{g(y)^5}(v(y) - \mu_v)^2. \end{aligned}$$

Lemma 5.5.6. *The variance of $\widehat{A}_3^\bullet$ is*

$$\begin{aligned} Var\left(\widehat{A}_3^\bullet\right) &= D_{18}^\bullet \frac{1}{N} + D_{19}^\bullet \frac{b^2}{N} + D_{20}^\bullet \frac{b^4}{N} + D_{21}^\bullet \frac{1}{N^2 b} + D_{22}^\bullet \frac{1}{N^2} + D_{23}^\bullet \frac{b}{N^2} \\ &+ \quad D_{24}^\bullet \frac{1}{N^3 b^2} + D_{25}^\bullet \frac{1}{N^3 b} + D_{26}^\bullet \frac{1}{N^3} + O\left(\frac{b^6}{N}\right) + O\left(\frac{b^2}{N^2}\right), \qquad (5.13) \end{aligned}$$

where

$$\begin{aligned} D_{18}^\bullet &:= c^2 \int \beta(y) dy = D_6^\bullet, \\ D_{19}^\bullet &:= \mu_2(K) c^2 \left[\int \alpha(y) g''(y) dy + \int \gamma''(y) f(y)(v(y) - \mu_v) dy \right], \\ D_{20}^\bullet &:= \frac{\mu_4(K)}{12} c^2 \left[\int \alpha(y) g^{(4)}(y) dy + \int \gamma^{(4)}(y) f(y)(v(y) - \mu_v) dy \right] \\ &+ \frac{\mu_2(K)^2}{4} c^2 \left[\int \delta(y) g''(y)^2 dy - 4 \left(\int \gamma(y) g''(y) dy \right)^2 \right. \\ &+ \left. \int \gamma''(y)^2 g(y) dy + 2 \int \gamma(y) \gamma''(y) g''(y) dy \right], \\ D_{21}^\bullet &:= 2c^2 \left[\mu_0(K^2) + K(0) \right] \int \alpha(y) dy, \\ D_{22}^\bullet &:= -8c^2 \int \beta(y) dy = -8 D_6^\bullet, \\ D_{23}^\bullet &:= \mu_2(K) K(0) c^2 \left[\int \delta(y) g''(y) dy + \int \gamma(y) \gamma''(y) dy \right. \\ &- \left. \left(\int \gamma(y) g''(y) dy + \int \gamma''(y) g(y) dy \right) \left(\int \gamma(y) dy \right) \right] \\ &+ \frac{\mu_2(K^2)}{2} c^2 \left[\int \delta(y) g''(y) dy + \int \gamma(y) \gamma''(y) dy \right], \\ D_{24}^\bullet &:= K(0)^2 c^2 \left[\int \delta(y) dy - \left(\int \gamma(y) dy \right)^2 \right], \\ D_{25}^\bullet &:= -2c^2 \left[\mu_0(K^2) + 2K(0) \right] \int \alpha(y) dy, \\ D_{26}^\bullet &:= 8c^2 \int \beta(y) dy = 8 D_6^\bullet, \end{aligned}$$

with

$$\begin{aligned} \alpha(y) &:= \frac{f(y)^2}{g(y)^2} (v(y) - \mu_v)^2, \\ \delta(y) &:= \frac{f(y)^2}{g(y)^3} (v(y) - \mu_v)^2. \end{aligned}$$

Lemma 5.5.7. *The covariance of $\widehat{A}_1^\bullet$ and $\widehat{A}_2^\bullet$ is*

$$Cov\left(\widehat{A}_1^\bullet, \widehat{A}_2^\bullet\right) = D_{27}^\bullet \frac{h^2}{N} + D_{28}^\bullet \frac{h^4}{N} + O\left(\frac{h^6}{N}\right), \tag{5.14}$$

where

$$\begin{aligned} D_{27}^\bullet &:= \frac{\mu_2(K)}{2} c^3 \int \psi(y) m''(y) dy, \\ D_{28}^\bullet &:= \frac{\mu_4(K)}{24} c^3 \int \psi(y) m^{(4)}(y) dy. \end{aligned}$$

Lemma 5.5.8. *The covariance of $\widehat{A}_1^\bullet$ and $\widehat{A}_3^\bullet$ is*

$$\begin{aligned} Cov\left(\widehat{A}_1^\bullet, \widehat{A}_3^\bullet\right) &= D_{29}^\bullet \frac{1}{N} + D_{30}^\bullet \frac{1}{N^2 b} + D_{31}^\bullet \frac{1}{N^2} + D_{32}^\bullet \frac{b^2}{N} \\ &+ D_{33}^\bullet \frac{b^4}{N} + O\left(\frac{b^2}{N^2}\right) + O\left(\frac{b^6}{N}\right), \end{aligned} \tag{5.15}$$

where

$$\begin{aligned} D_{29}^\bullet &:= c^2 \int \beta(y) dy = D_6^\bullet, \\ D_{30}^\bullet &:= K(0) c^2 \int \alpha(y) dy, \\ D_{31}^\bullet &:= -2c^2 \int \beta(y) dy = -2D_6^\bullet, \\ D_{32}^\bullet &:= \frac{\mu_2(K)}{2} c^2 \left[\int \alpha(y) g''(y) dy + \int \gamma''(y) f(y)(v(y) - \mu_v) dy\right], \\ D_{33}^\bullet &:= \frac{\mu_4(K)}{24} c^2 \left[\int \alpha(y) g^{(4)}(y) dy + \int \gamma^{(4)}(y) f(y)(v(y) - \mu_v) dy\right]. \end{aligned}$$

Lemma 5.5.9. *The covariance of $\widehat{A}_2^\bullet$ and $\widehat{A}_3^\bullet$ is*

$$\begin{aligned} Cov\left(\widehat{A}_2^\bullet, \widehat{A}_3^\bullet\right) &= D_{34}^\bullet \frac{h^2}{N} + D_{35}^\bullet \frac{h^2 b^2}{N} + D_{36}^\bullet \frac{h^4}{N} + D_{37}^\bullet \frac{h^4 b^2}{N} + D_{38}^\bullet \frac{h^2 b^4}{N} \\ &+ D_{39}^\bullet \frac{h^2}{N^2 b} + D_{40}^\bullet \frac{h^2}{N^2} + D_{41}^\bullet \frac{h^4}{N^2 b} + D_{42}^\bullet \frac{h^4}{N^2} + D_{43}^\bullet \frac{h^2 b^2}{N^2} + O\left(\frac{h^6}{N}\right) \\ &+ O\left(\frac{h^2 b^6}{N}\right) + O\left(\frac{h^4 b^4}{N}\right) + O\left(\frac{h^6}{N^2 b}\right) + O\left(\frac{h^2 b^4}{N^2}\right) + O\left(\frac{h^4 b^2}{N^2}\right), \end{aligned} \tag{5.16}$$

where

$$\begin{aligned} D_{34}^\bullet &:= \frac{\mu_2(K)}{2} c^3 \int \psi(y) m''(y) dy = D_{27}^\bullet, \\ D_{35}^\bullet &:= \frac{\mu_2(K)^2}{4} c^3 \left[\int \zeta(y) g''(y) m''(y) dy + \int m''(y) \gamma''(y) \Omega(y) dy \right. \\ &- \left. 2B^\bullet(\Omega'') \int \gamma(y) g''(y) dy\right], \\ D_{36}^\bullet &:= \frac{\mu_4(K)}{24} c^3 \int \psi(y) m^{(4)}(y) dy = D_{28}^\bullet, \end{aligned}$$

$$\begin{aligned}
D^\bullet_{37} &:= \frac{\mu_2(K)\mu_4(K)}{48}c^3\left[\int\zeta(y)m^{(4)}(y)g''(y)dy+\int m^{(4)}(y)\gamma''(y)\Omega(y)dy\right.\\
&- \left.2B^\bullet(\Omega^{(4)})\int\gamma(y)g''(y)dy\right],\\
D^\bullet_{38} &:= \frac{\mu_2(K)\mu_4(K)}{48}c^3\left[\int\zeta(y)(y)m''(y)g^{(4)}(y)dy+\int m''(y)\gamma^{(4)}(y)\Omega(y)dy\right.\\
&- \left.2B^\bullet(\Omega'')\int\gamma(y)g^{(4)}(y)dy\right],\\
D^\bullet_{39} &:= c^3\frac{\mu_2(K)K(0)}{2}\left[\int\zeta(y)m''(y)dy-B^\bullet(\Omega'')\int\gamma(y)dy\right],\\
D^\bullet_{40} &:= -2D^\bullet_{34},\\
D^\bullet_{41} &:= \frac{\mu_4(K)K(0)}{24}c^3\left[\int\zeta(y)m^{(4)}(y)dy-B^\bullet(\Omega^{(4)})\int\gamma(y)dy\right],\\
D^\bullet_{42} &:= -2D^\bullet_{36},\\
D^\bullet_{43} &:= -D^\bullet_{35},
\end{aligned}$$

with

$$\zeta(y) := \frac{f(y)^3}{g(y)^4}(v(y)-\mu_v)^2.$$

Lemma 5.5.10. *The variance of $\widehat{A}^\bullet$ is*

$$\begin{aligned}
Var\left(\widehat{A}^\bullet\right) &\simeq D^\bullet_7\frac{1}{n}+D^\bullet_8\frac{1}{Nn}+4D^\bullet_6\frac{1}{N}-12D^\bullet_6\frac{1}{N^2}+D^\bullet_{26}\frac{1}{N^3}+D^\bullet_9\frac{1}{Nnh}+D^\bullet_{44}\frac{1}{N^2b}\\
&+ D^\bullet_{24}\frac{1}{N^3b^2}+D^\bullet_{25}\frac{1}{N^3b}-2D^\bullet_{39}\frac{h^2}{N^2b}-2D^\bullet_{41}\frac{h^4}{N^2b}+D^\bullet_{10}\frac{h^2}{n}+D^\bullet_{11}\frac{h^4}{n}-4D^\bullet_{27}\frac{h^2}{N}\\
&+ 2D^\bullet_{19}\frac{b^2}{N}+D^\bullet_{46}\frac{h^4}{N}+D^\bullet_{45}\frac{b^4}{N}-2D^\bullet_{35}\frac{h^2b^2}{N}-2D^\bullet_{37}\frac{h^4b^2}{N}-2D^\bullet_{38}\frac{h^2b^4}{N}+D^\bullet_{13}\frac{h}{Nn}\\
&+ D^\bullet_{14}\frac{h^2}{Nn}+D^\bullet_{15}\frac{h^3}{Nn}+D^\bullet_{23}\frac{b}{N^2}+4D^\bullet_{27}\frac{h^2}{N^2}+4D^\bullet_{28}\frac{h^4}{N^2}+2D^\bullet_{35}\frac{h^2b^2}{N^2}+O\left(\frac{h^6}{n}\right)\\
&+ O\left(\frac{b^6}{N}\right)+O\left(\frac{h^4b^4}{N}\right)+O\left(\frac{h^4}{Nn}\right)+O\left(\frac{b^2}{N^2}\right)+O\left(\frac{h^6}{N^2b}\right),
\end{aligned}$$

where

$$\begin{aligned}
D^\bullet_{44} &:= D^\bullet_{21}+2D^\bullet_{30}=2\mu_0(K^2)c^2\int\alpha(y)dy+4K(0)c^2\int\alpha(y)dy,\\
D^\bullet_{45} &:= D^\bullet_{20}+2D^\bullet_{33}=\frac{\mu_4(K)}{6}c^2\left[\int\alpha(y)g^{(4)}(y)dy+\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)\right]\\
&+ \frac{\mu_2(K)^2}{4}c^2\left[\int\delta(y)g''(y)^2dy-4\left(\int\gamma(y)g''(y)dy\right)^2\right.\\
&+ \left.\int\gamma''(y)^2g(y)dy+2\int\gamma(y)\gamma''(y)g''(y)dy\right].
\end{aligned}$$

$$
\begin{aligned}
D_{46}^{\bullet} &:= D_{12}^{\bullet} - 4D_{28}^{\bullet} = \frac{\mu_2(K)^2}{4}c^4\left[\int \xi(y)m''(y)^2dy - B^{\bullet}(\Omega'')^2\right] \\
&- \frac{\mu_4(K)}{6}c^3\int \psi(y)m^{(4)}(y)dy.
\end{aligned}
$$

The proof of Theorem 5.2.1 is a consequence of Lemmas from 5.5.3 to 5.5.10, considering $C_1^{\bullet} := D_2^{\bullet}/c$, $C_2^{\bullet} := D_1^{\bullet}/c$, $C_3 := D_4^{\bullet}/c$, $C_4^{\bullet} := D_7^{\bullet}/c^2$, $C_5^{\bullet} := D_8^{\bullet}/c^2$, $C_6 := -12D_6^{\bullet}/c^2$, $C_7^{\bullet} := D_9^{\bullet}/c^2$, $C_8^{\bullet} := D_{44}^{\bullet}/c^2$, $C_9 := D_{10}^{\bullet}/c^2$, $C_{10}^{\bullet} := -2D_{39}^{\bullet}/c^2$, $C_{11}^{\bullet} := D_{46}^{\bullet}/c^2$, $C_{12}^{\bullet} := -2D_{35}^{\bullet}/c^2$, $C_{13}^{\bullet} := 4_{16}^{\bullet}/c^2$, $C_{14}^{\bullet} := -4D_{27}^{\bullet}/c^2$ and $C_{15}^{\bullet} := 2D_{19}^{\bullet}/c^2$.

5.5.2 Sketch of the proof of Theorem 5.2.2

The proof of Theorem 5.2.2 follows parallel lines to that of Theorem 5.2.1. The proofs of the following lemmas are available in Appendix A.2. Along this subsection, Conditions A1 and A13-A15 are assumed for Lemmas 5.5.11-5.5.19.

Lemma 5.5.11. *The expectation and variance of $\widehat{A}^{*\bullet}$, being*

$$\widehat{A}^{*\bullet} := \widehat{A}_1^{*\bullet} - \widehat{A}_2^{*\bullet} + \widehat{A}_3^{*\bullet},$$

are

$$
\begin{aligned}
E\left(\widehat{A}^{*\bullet}\right) &= E\left(\widehat{A}_1^{*\bullet}\right) - E\left(\widehat{A}_2^{*\bullet}\right) + E\left(\widehat{A}_3^{*\bullet}\right), && (5.17)\\
Var\left(\widehat{A}^{*\bullet}\right) &= Var\left(\widehat{A}_1^{*\bullet}\right) + Var\left(\widehat{A}_2^{*\bullet}\right) + Var\left(\widehat{A}_3^{*\bullet}\right) \\
&- 2Cov\left(\widehat{A}_1^{*\bullet}, \widehat{A}_2^{*\bullet}\right) + 2Cov\left(\widehat{A}_1^{*\bullet}, \widehat{A}_3^{*\bullet}\right) - 2Cov\left(\widehat{A}_2^{*\bullet}, \widehat{A}_3^{*\bullet}\right), && (5.18)
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{A}_1^{*\bullet} &:= \frac{1}{N}\sum_{i=1}^{N}\frac{(K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v), \\
\widehat{A}_2^{*\bullet} &:= \frac{1}{N}\sum_{i=1}^{N}\frac{(K_b * g)(Y_i)(\hat{m}_h(Y_i) - (K_h * m)(Y_i))}{(K_h * m)(Y_i)^2}(v(Y_i) - \mu_v), \\
\widehat{A}_3^{*\bullet} &:= \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v).
\end{aligned}
$$

Lemma 5.5.12. *The expectation of $\widehat{A}^{*\bullet}$ is*

$$E(\widehat{A}^{*\bullet}) = D_1^{*\bullet} + D_2^{*\bullet}\frac{1}{N}. \qquad (5.19)$$

where

$$\begin{aligned} D_1^{*\bullet} &:= \int \gamma^{*\bullet}(y)g(y)dy, \\ D_2^{*\bullet} &:= -D_1^{*\bullet} + \frac{K(0)}{b}\int \frac{(v(y)-\mu_v)}{(K_h * m)(y)}g(y)dy, \end{aligned}$$

with

$$\gamma^{*\bullet}(y) := \frac{(K_b * g)(y)}{(K_h * m)(y)}(v(y)-\mu_v).$$

Lemma 5.5.13. *The variance of $\widehat{A}_1^{*\bullet}$ is*

$$Var\left(\widehat{A}_1^{*\bullet}\right) = \frac{D_3^{*\bullet}}{N}. \tag{5.20}$$

with

$$D_3^{*\bullet} := \int \alpha^{*\bullet}(y)g(y)dy - D_1^{*\bullet 2},$$

where

$$\alpha^{*\bullet}(y) := \frac{(K_b * g)(y)^2}{(K_h * m)(y)^2}(v(y)-\mu_v)^2.$$

Lemma 5.5.14. *The variance of $\widehat{A}_2^{*\bullet}$ is*

$$Var\left(\widehat{A}_2^{*\bullet}\right) = D_4^{*\bullet}\frac{1}{n} + D_5^{*\bullet}\frac{1}{Nn}. \tag{5.21}$$

where

$$\begin{aligned} D_4^{*\bullet} &:= \int \left(\int \varphi^{*\bullet}(y)K_h(y-z)g(y)dy - D_1^{*\bullet}\right)^2 m(z)dz, \\ D_5^{*\bullet} &:= \int \varphi^{*\bullet}(y)^2((K_h)^2 * m)(y)g(y)dy - \int \alpha^{*\bullet}(y)g(y)dy - D_4^{*\bullet}, \end{aligned}$$

where

$$\varphi^{*\bullet}(y) := \frac{(K_b * g)(y)}{(K_h * m)(y)^2}(v(y)-\mu_v).$$

Lemma 5.5.15. *The variance of $\widehat{A}_3^{*\bullet}$ is*

$$Var\left(\widehat{A}_3^{*\bullet}\right) = D_6^{*\bullet}\frac{1}{N} + D_7^{*\bullet}\frac{1}{N^2} + D_8^{*\bullet}\frac{1}{N^3}, \tag{5.22}$$

where

$$
\begin{aligned}
D_6^{*\bullet} &:= \int \left[\int \frac{v(y)-\mu_v}{(K_h*m)(y)} K_b(y-z)g(y)dy - \int \gamma^{*\bullet}(y)g(y)dy\right]^2 g(z)dz \\
&= \int \left[\int \frac{v(y)-\mu_v}{(K_h*m)(y)} K_b(y-z)g(y)dy - D_1^{*\bullet}\right]^2 g(z)dz,
\end{aligned}
$$

$$
\begin{aligned}
D_7^{*\bullet} &:= \frac{2K(0)}{b}\left[\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz\right. \\
&- \left.\int \gamma^{*\bullet}(y)g(y)dy \int \frac{v(y)-\mu_v}{(K_h*m)(y)} g(y)dy\right] \\
&- 4\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz + 3\left(\int \gamma^{*\bullet}(y)g(y)dy\right)^2 \\
&- \int \alpha^{*\bullet}(y)g(y)dy + \int \left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 ((K_b)^2*g)(y)g(y)dy \\
&+ \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)^2 g(y)g(z)dydz \\
&- 3\int \left[\int \frac{v(y)-\mu_v}{(K_h*m)(y)} K_b(y-z)g(y)dy - \int \gamma^{*\bullet}(y)g(y)dy\right]^2 g(z)dz \\
&= \frac{2K(0)}{b}\left[\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz\right. \\
&- \left.\int \gamma^{*\bullet}(y)g(y)dy \int \frac{v(y)-\mu_v}{(K_h*m)(y)} g(y)dy\right] \\
&- 4\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz + 3D_1^{*\bullet 2} \\
&- \int \alpha^{*\bullet}(y)g(y)dy + \int \left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 ((K_b)^2*g)(y)g(y)dy \\
&+ \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)^2 g(y)g(z)dydz - 3D_6^{*\bullet},
\end{aligned}
$$

$$
\begin{aligned}
D_8^{*\bullet} &:= \frac{K(0)^2}{b^2}\left[\int \left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 g(y)dy - \left(\int \frac{v(y)-\mu_v}{(K_h*m)(y)} g(y)dy\right)^2\right] \\
&+ \frac{2K(0)}{b}\left[-\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz\right. \\
&+ \left. 2\int \gamma^{*\bullet}(y)g(y)dy \int \frac{v(y)-\mu_v}{(K_h*m)(y)} g(y)dy - \int \rho^{*\bullet}(y)g(y)dy\right] \\
&+ 4\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz - 4\left(\int \gamma^{*\bullet}(y)g(y)dy\right)^2 \\
&+ 2\int \alpha^{*\bullet}(y)g(y)dy - \int \left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 ((K_b)^2*g)(y)g(y)dy
\end{aligned}
$$

$$\begin{aligned}
&- \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)^2 g(y)g(z)dydz \\
&+ 2\int\left[\int \frac{v(y)-\mu_v}{(K_h*m)(y)} K_b(y-z)g(y)dy - \int \gamma^{*\bullet}(y)g(y)dy\right]^2 g(z)dz \\
&= \frac{K(0)^2}{b^2}\left[\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 g(y)dy - \left(\int \frac{v(y)-\mu_v}{(K_h*m)(y)} g(y)dy\right)^2\right] \\
&+ \frac{2K(0)}{b}\left[-\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz\right. \\
&+ \left. 2\int \gamma^{*\bullet}(y)g(y)dy \int \frac{v(y)-\mu_v}{(K_h*m)(y)} g(y)dy - \int \rho^{*\bullet}(y)g(y)dy\right] \\
&+ 4\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz - 4D_1^{*\bullet 2} \\
&+ 2\int \alpha^{*\bullet}(y)g(y)dy - \int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 ((K_b)^2*g)(y)g(y)dy \\
&- \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)^2 g(y)g(z)dydz + 2D_6^{*\bullet},
\end{aligned}$$

with

$$\rho^{*\bullet}(y) := \frac{(K_b*g)(y)}{(K_h*m)(y)^2}(v(y)-\mu_v)^2.$$

Lemma 5.5.16. *The covariance of $\widehat{A}_1^{*\bullet}$ and $\widehat{A}_2^{*\bullet}$ is*

$$Cov\left(\widehat{A}_1^{*\bullet}, \widehat{A}_2^{*\bullet}\right) = 0.$$

Lemma 5.5.17. *The covariance of $\widehat{A}_1^{*\bullet}$ and $\widehat{A}_3^{*\bullet}$ is*

$$Cov\left(\widehat{A}_1^{*\bullet}, \widehat{A}_3^{*\bullet}\right) = D_9^{*\bullet}\frac{1}{N} + D_{10}^{*\bullet}\frac{1}{N^2}, \tag{5.23}$$

where

$$\begin{aligned}
D_9^{*\bullet} &:= \int\int \gamma^{*\bullet}(z)\frac{v(y)-\mu_v}{(K_h*m)(y)} K_b(y-z)g(y)g(z)dydz - D_1^{*\bullet 2}, \\
D_{10}^{*\bullet} &:= \frac{K(0)}{b}\left[\int \rho^{*\bullet}(y)g(y)dy - \left(\int \frac{v(y)-\mu_v}{(K_h*m)(y)} g(y)dy\right) D_1^{*\bullet}\right] - D_3^{*\bullet} - D_9^{*\bullet}.
\end{aligned}$$

Lemma 5.5.18. *The covariance of $\widehat{A}_2^{*\bullet}$ and $\widehat{A}_3^{*\bullet}$ is*

$$Cov\left(\widehat{A}_2^{*\bullet}, \widehat{A}_3^{*\bullet}\right) = 0.$$

Lemma 5.5.19. *The variance of $\widehat{A}^{\bullet}$ is*

$$Var\left(\widehat{A}^{*\bullet}\right) = D_4^{*\bullet}\frac{1}{n} + D_{11}^{*\bullet}\frac{1}{N} + D_5^{*\bullet}\frac{1}{Nn} + D_{12}^{*\bullet}\frac{1}{N^2} + D_8^{*\bullet}\frac{1}{N^3},$$

where

$$\begin{aligned} D_{11}^{*\bullet} &:= D_3^{*\bullet} + D_6^{*\bullet} + 2D_9^{*\bullet}, \\ D_{12}^{*\bullet} &:= D_7^{*\bullet} + 2D_{10}^{*\bullet}. \end{aligned}$$

The proof of Theorem 5.2.2 is a consequence of Lemmas from 5.5.12 to 5.5.19, considering $C_1^{*\bullet} := (D_1^{*\bullet}/c)^2$, $C_2^{*\bullet} := D_4^{*\bullet}/c^2$, $C_3^{*\bullet} := 2D_1^{*\bullet}D_2^{*\bullet}/c^2 + D_{11}^{*\bullet}/c^2$, $C_4^{*\bullet} := D_5^{*\bullet}/c^2$, $C_5^{*\bullet} := (D_2^{*\bullet}/c)^2 + D_{12}^{*\bullet}/c^2$, $C_6^{*\bullet} := D_8^{*\bullet}/c^2$.

Chapter 6

Real data applications

The methods proposed in Chapters 4 and 5 are applied to several real data sets. In Section 6.1, it is considered the US airlines data set with information of the arrival and departure details of commercial flights. The air quality data set about the emissions of different air pollutants registered in the city of A Coruña is studied in Section 6.2. The content of this section was published in Borrajo & Cao (2020). In Section 6.3 the customer churn data set from the Telco Company Vodafone ES is considered. Finally, in Section 6.4 the proposed methods are applied to the study of two COVID-19 data sets with information on asymptomatic, identified and hospitalized cases.

6.1 Airline on-time performance

The airline on-time performance (AOTP) data set is available at `http://stat-computing.org/dataexpo/2009/the-data.html`. It consists of nearly 180 million records about flight arrival and departure details for all commercial flights within the US, from October 1987 to May 2018.

In this context, we would like to estimate the mean arrival delay (in minutes) of US flights for the whole year 2017. We use (4.4) considering as $\mathbf{Y}$ the whole data set for the year 2016 ($N = 5,617,658$) and assuming that only the arrival delay time for the flights of January 11th, 2017, $\mathbf{X}$, ($n = 14,568$) is available. Since the first days of January are atypical due to the holiday period and since weekends and Mondays do not always accurately reflect the behavior of a normal labor day, we decided to collect the data from the first available Wednesday (January 11th, 2017) in order to obtain something close to a SRS of the true 2017 population.

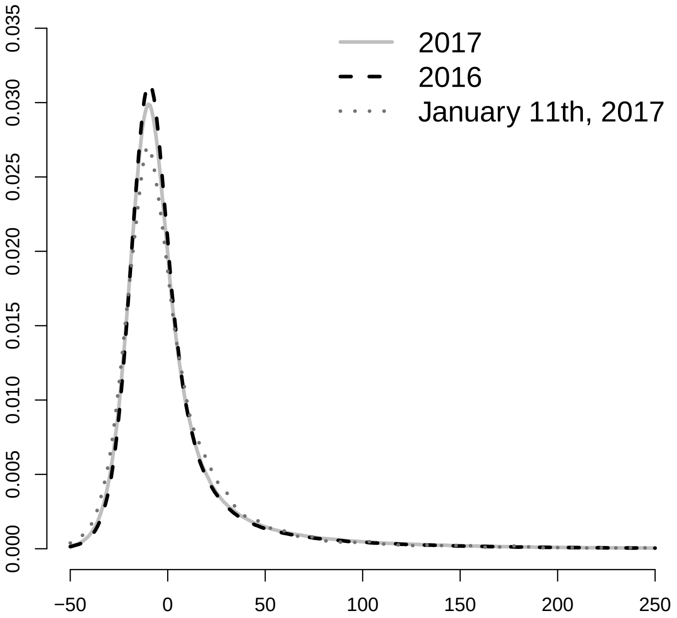

Figure 6.1: Densities involved in the case study with AOTP data. Densities of arrival delays in 2016 (dashed black line), 2017 (solid gray line) and January, 11 2017 (dotted gray line).

To illustrate the difference between the density functions of the arrival delay of US flights in 2016 and 2017, two kernel density estimations have been plotted in Figure 6.1. These density estimates are based on nearly 6 million data each. Although the two estimated annual densities are very similar, they exhibit some subtle differences, for instance the level of the density at the mode. Figure 6.1 also contains the kernel density estimation based on the arrival delays of January 11th, 2017.

Table 6.1: p-values obtained using the two-sample Kolmogorov–Smirnov (KS) test for equality of distributions and using the two-sample Student's t-test for equality of means.

Variable	KS test	$\overline{X}$	$\overline{Y}$	t-test
Arrival delay	$<8.5 \times 10^{-15}$	4.742243	3.519290	0.005194

To test for sampling bias, we can use the two-sample Kolmogorov–Smirnov test for equality of distributions and the Student's t-test for equality of means. The

p-values obtained in Table 6.1 allow to reject the null hypothesis in both cases, in favour of the presence of bias and the non equality of means.

In this context, we computed the values of $\overline{X}$ and $\overline{Y}$ and we applied the bootstrap bandwidth selection method presented in Subsection 4.4 in order to obtain the values of h and b that provide a good performance of $\hat{\mu}$. As a benchmark, we considered the sample mean of all the arrival delays in 2017, which would not be known until the end of that year. Although this is a sample mean, it is based on more than 6 million data, so we will consider that is approximately equal to the true $\mu = \mu_{2017}$.

Table 6.2 shows the estimated value, $\hat{\mu}^{2,h^*,b^*}$, with the bootstrap bandwidth selectors h^* and b^*. It is clear from this table that $\hat{\mu}^{2,h^*,b^*}$ may perform extremely well (nearly perfect) for a suitable bandwidth choice in comparison with the classical estimators.

Table 6.2: Comparison of the full 2017 sample mean and the standard deviation with the mean and variance of January, 11st 2017, the B3D sample mean and the standard deviation (2016) and the proposed estimators $\hat{\mu}$ and $\hat{\sigma}$ for the bootstrap bandwidths (h^*, b^*).

Mean estimation			Standard deviation estimation		
μ_{2017}	$\overline{X}$	$\overline{Y}$	σ_{2017}	$S_{n,X}$	$S_{N,Y}$
4.326357	4.742243	3.519290	45.8648	48.46081	41.87338
(h^*, b^*)	$\hat{\mu}^{2,h^*,b^*}$			$\hat{\sigma}^{h^*,b^*}$	
(9.12,12.02)	4.326778			49.59647	

As mentioned in Section 4.2, $E(X^2)$ can be estimated using the proposed general procedure with the choice $v(x) = x^2$. As a consequence, the variance and, therefore, the standard deviation can also be estimated. The estimated value for the standard deviation, $\hat{\sigma}^{h^*,b^*}$, is also shown in Table 6.2 for the same bandwidth values considered in the mean estimation. In this case, the performance of our estimator is slightly worse than that of the SRS sample but better than that of the big-but-biased sample when comparing with the true $\sigma = \sigma_{2017}$. This happens because we have not used a bootstrap method adapted for this parameter. Analyzing the behavior of this estimator in a wide range of values, we observe that for a suitable choice of the smoothing parameters ($h = 3.02$ and $b = 19.05$), our estimator would behave significantly well ($\hat{\sigma}^{h,b} = 45.85975$), so it is logical to think that using an adapted version of the bootstrap algorithm for the standard deviation, our estimator would improve its behavior.

6.2 Air quality in smart cities

Making a city smart has emerged as a strategy to mitigate the challenges of urban population growth and fast urbanization, provinding better quality of life to its citizens (Chourabi et al., 2012). The important role of Big Data Analytics and Information and Communication Technology in the development of smart cities initiatives is unquestionable (Hashem et al., 2016). Some of the economic, environmental and social benefits and opportunities of using Big Data in smart city applications are detailed in Al Nuaimi et al. (2015). There are many applications of Big Data in differents domains of smart cities, such as city planning, environment, sustainability, traffic management, transportation, security and education (Osman, 2019; Al Nuaimi et al., 2015).

6.2.1 Motivation

Despite the many advantages of applying Big Data Analytics to urban data, some authors have also identified some of its challenges, such as the importance of managing truthful and quality data (Al Nuaimi et al., 2015; Lim et al., 2018). This problem is highly related to the idea that with enough data, numbers speak for themselves; idea on which this thesis is based.

The present application focuses on the domain of public health in smart cities; in particular, on urban air quality, since air pollution is one of the big concerns for smart cities. Information about real-time air quality is of great importance to protect humans from damage by air pollution (Zheng et al., 2013).

There are many methods and works using Big Data Analytics to predict the air quality in smart cities. Martínez-España et al. (2018) compare several machine learning methods in order to choose the most suitable for predicting the ozone level in the Region of Murcia (Spain). In Ameer et al. (2019), four regression methods based on machine learning techniques are proposed to predict air pollution and compare their accuracy in terms of error rate and processing time, using multiple data sets. A novel deep learning model based on Long Short Term Memory networks is presented in Kök et al. (2017) to make predictions about air quality in smart cities. In Ramos et al. (2018), the authors use the existing sensor networks in smart cities to create and promote alternative pollution-free routes across cities depending on the level of pollution in each zone and apply the study carried out to Madrid (Spain).

However, none of these works deal with the problem of sampling bias in Big Data Analytics for smart cities and air quality. This application deals with the particular problem of urban air pollution in that context, using the nonparametric estimation method proposed in Chapter 4.

Subsection 6.2.2 introduces the problem of air pollution by focusing on two specific pollutants. A real data set study is carried out in Subsection 6.2.3. It is a data set with information of different variables of interest about air quality in the city of A Coruña, Galicia, NW Spain. The mean and the cumulative distribution function of the level of ozone (O_3) and nitrogen dioxide (NO_2) when the temperature is greater than or equal to 30 °C is estimated based on 15 years of big-but-biased data.

6.2.2 Urban Air Pollution

According to the World Health Organization (WHO), air pollution kills an estimated number of seven million people worldwide every year, increasing deaths from stroke, chronic obstructive pulmonary disease, lung cancer, heart disease and acute respiratory infections (World Health Organization, 2020).

Air quality control and management have been one of the priorities of the environmental policy of the City Council of A Coruña for several years. After conducting an emission analysis, four automatic air pollution control stations were installed in different points of the city, aimed at protecting human health.

The Air Quality Index (AQI) is a global indicator of the air quality of an area at a certain time of the day, based on data provided by air quality monitoring stations. The AQI is calculated from information related to different atmospheric pollutants: sulfur dioxide (SO_2), nitrogen oxides (NO_2 y NO_x), carbon monoxide (CO), tropospheric ozone (O_3), benzene (C_6H_6) and airborne particulate matter, smaller than 10 micrometers in diameter (PM10) and smaller than 2.5 micrometers in diameter (PM2.5). The AQI value changes every hour depending on the values obtained by the real-time surveillance stations. In case the air quality is poor, those responsible for the surveillance network receive an alert by email, initiating the corresponding action protocol (Ayuntamiento de A Coruña, 2020). This application focuses on the levels of two of these pollutants: ozone (O_3) and nitrogen dioxide (NO_2).

Ozone at ground level (tropospheric ozone) is formed by the reaction of pollutants such as nitrogen oxides and volatile organic compounds emitted by vehicles and industry. Solar radiation plays an important role in these reactions, since the reactions are photochemical in nature and require high temperatures to be effective. As a result, the highest levels of ozone pollution occur on sunny and hot days. Excessive ozone in the air can have a marked effect on human health. It can cause breathing problems, trigger asthma, reduce lung function and cause lung diseases (World Health Organization, 2020; Ayuntamiento de A Coruña, 2020).

Nitrogen dioxide is one of the most dangerous pollutants due to its toxic and irritating nature, which causes significant inflammation of the airways. In addition, it decomposes through light to form atomic oxygen, which is very reactive, and converts molecular oxygen into ozone. The major emissions of NO_2 are of anthropogenic origin, through combustion processes as heating, power generation and engines in vehicles and ships. Nitrogen oxides mainly affect the respiratory system and can cause bronchitis and pneumonia as well as a lower resistance to respiratory tract infections (World Health Organization, 2020; Ayuntamiento de A Coruña, 2020).

Since high temperatures are related to high values of these pollutants, the problem of estimating their levels when the temperature is greater than or equal to 30°C is considered. To carry out this study, the bias correction method proposed in Chapter 4, whose good performance has already been tested, is used.

6.2.3 Results

The air quality data set used is available in Ayuntamiento de A Coruña (2020). It consists of nearly 126 thousands hourly records about the temperature, measured in centigrades (°C), and the levels, in μg/m^{3}N, of O_3 and NO_2 in the urban air of A Coruña during the last 15 years. These data have been collected from the Santa Margarita station, one of the four automatic air pollution control stations in the city.

We are interested in estimating the mean level of ozone and nitrogen dioxide when the temperature is greater than or equal to 30 °C, since it is believed that high temperatures may be associated with an increase in the levels of these harmful pollutants in the air. For this purpose, we use (4.4), considering as $\mathbf{Y}$ the whole data set for the last 15 years ($N = 125{,}949$ in the case of ozone and $N = 126{,}056$ for nitrogen dioxide) and as $\mathbf{X}$ the data set with the level of ozone and nitrogen dioxide when the temperature is greater than or equal to 30 °C in the last 15 years ($n = 275$ and $n = 267$, respectively). The difference between the sample sizes according to the variable considered is due to the missing data.

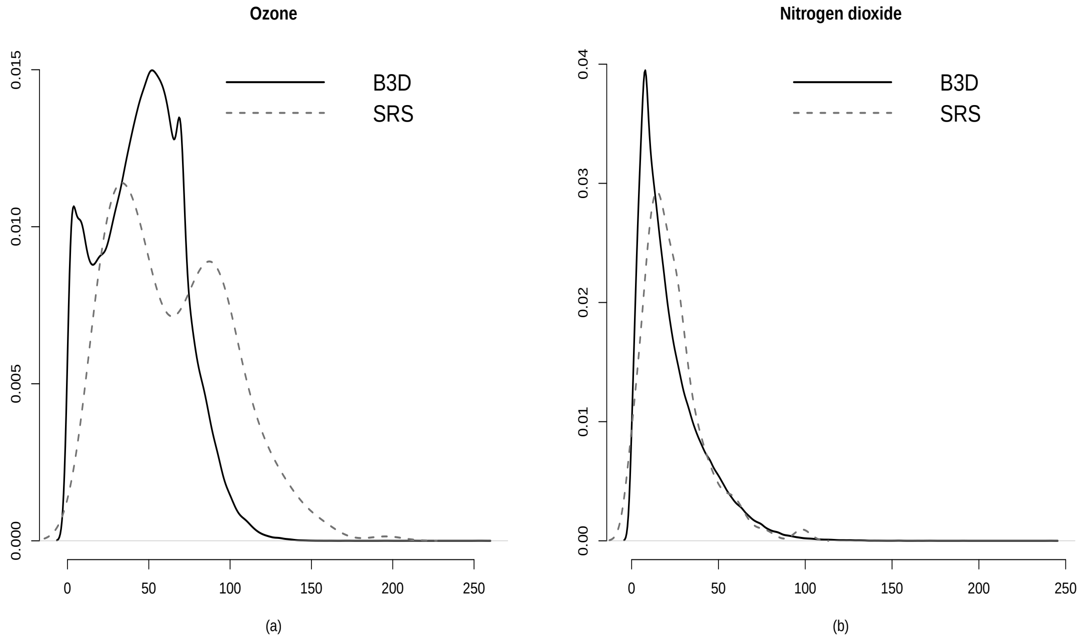

Figure 6.2: Estimated probability densities involved in the case study with air quality data. (**a**) Density of the ozone level in the last 15 years (solid black line) and its analogue for temperatures greater than or equal to 30 °C (dashed gray line). (**b**) Density of the nitrogen dioxide level in the last 15 years (solid black line) and its analogue for temperatures greater than or equal to 30 °C (dashed gray line).

Figure 6.2 shows the density functions of the levels of ozone and nitrogen dioxide with temperatures greater than or equal to 30 °C (dashed gray lines) when compared to the general levels of the last 15 years (solid black lines). The two densities are very similar for nitrogen dioxide, while they differ very much for ozone. The great difference depending on the temperature in the case of the ozone level was expected, since this connection has already been studied by several authors. Experiments performed in Cardelino & Chameides (1990) show that high temperatures increase the ozone level, while the effect on nitrogen oxides is uncertain. In Jhun et al. (2014), not only is the relation between both variables analyzed, but also the effect it has on ozone-related mortality, concluding that high temperatures increase ozone level, which leads to a rise in the mortality rate. Furthermore, Meleux et al. (2007) warn about how the increase in the ozone level will negatively affect human health, agriculture and natural ecosystems due to climate change.

In fact, we can use the two-sample Kolmogorov–Smirnov test (Kolmogorov, 1933; Smirnov, 1939) for equality of distributions to test for sampling bias. The p-values obtained (see Table 6.3) allow to reject the null hypothesis in both cases, in favour of the presence of bias, but with a higher level of confidence in the case of ozone. Table 6.3 also shows the p-values obtained using the Student's t-test, which allow to reject the equality of the means in the case of ozone with the usual significance levels. However, the hypothesis of equal means for nitrogen dioxide is accepted using the t-test.

Table 6.3: p-values obtained using the two-sample Kolmogorov–Smirnov (KS) test for equality of distributions and using the two-sample Student's t-test for equality of means.

Variable	KS test	$\overline{X}$	$\overline{Y}$	t-test
Ozone	$<2.2 \times 10^{-16}$	64.44	45.35	$<2.2 \times 10^{-16}$
Nitrogen dioxide	0.001064	23.88	22.28	0.1411

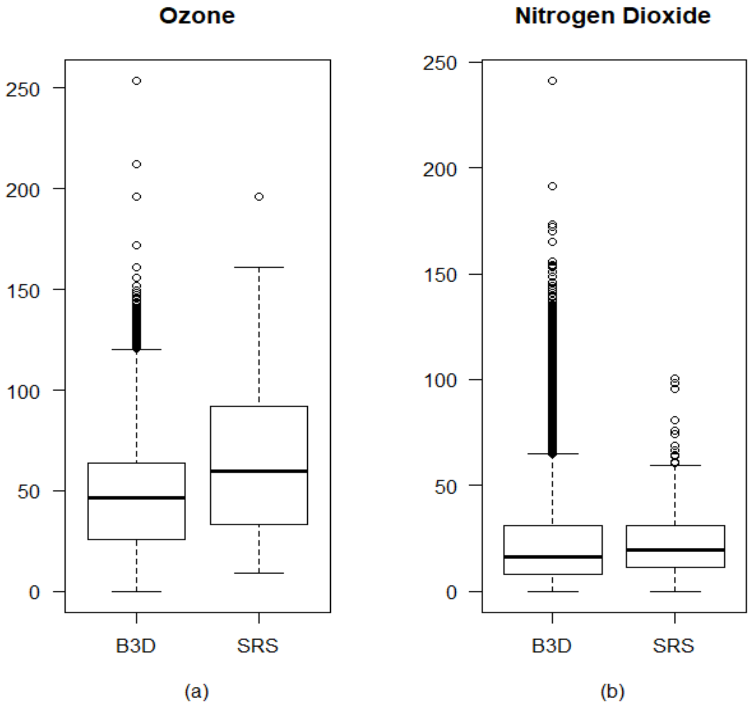

Figure 6.3: Boxplot of the four samples involved in the case study with air quality data. (**a**) Boxplot of the ozone level for the Big-But-Biased Data (B3D) sample (left) and the simple random sample (SRS) (right). (**b**) Boxplot of the nitrogen dioxide level for the B3D sample (left) and the SRS (right).

Figure 6.3 shows the presence of heavy right tails, as already observed in Figure 6.2. However, the proposed methods will not be affected by these observations, since they are not measurement errors, but unusually high values of the levels of ozone and nitrogen dioxide in those particular hourly records.

In this context, we computed the values of $\overline{X}$ and $\overline{Y}$ in each case. As the sample mean for the two pollutants when the temperature is greater than or equal to 30°C is not available, the real μ is unknown. For this reason, in order to know which values of h and b provide a good performance of our estimator, we use the bootstrap bandwidth selection method presented in Subsection 4.4.

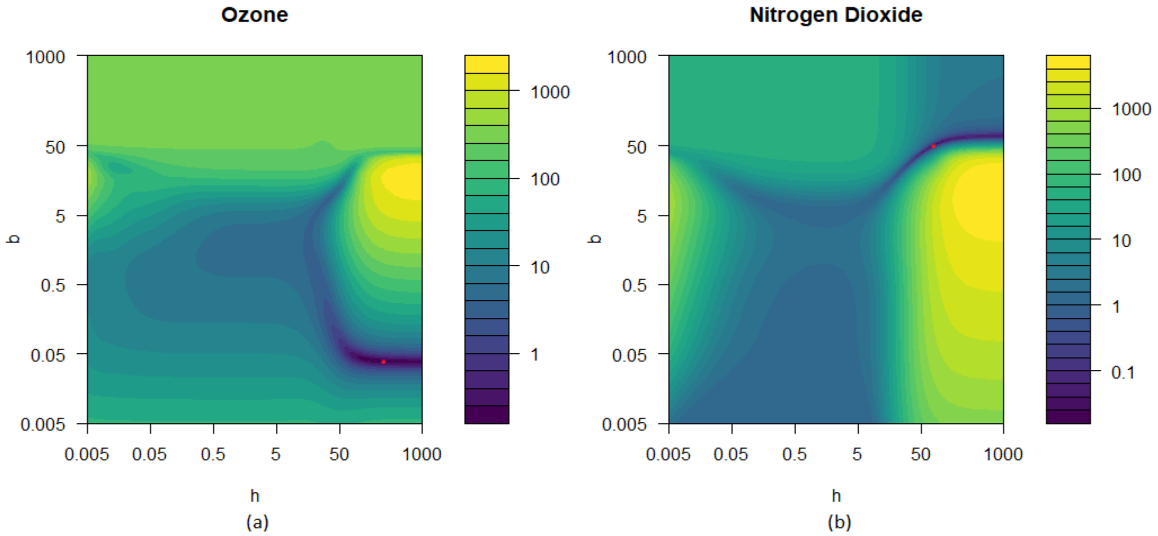

Figure 6.4: Estimation of the mean squared errors of the proposed estimator as a function of h and b, obtained by the bootstrap. (**a**) Mean squared error (MSE^*) of the estimator for the mean level of ozone. (**b**) MSE^* of the estimator for the mean level of nitrogen dioxide. The red dot represents the minimal value of the MSE^*.

Figure 6.4 shows the bootstrap mean squared errors of the proposed estimator for bandwidth selection in the case of ozone and nitrogen dioxide, respectively. This figure provides some estimate of the optimal values for the two bandwidths, those that minimize the MSE^*. These values of h and b will be the ones used in the proposed method to estimate the mean level of both pollutants when the temperature is greater than or equal to 30 °C. This figure also shows how relevant it is to properly select the smoothing parameter b; otherwise, the MSE^* would increase significantly. Once b has been chosen, it is also important to find a suitable h, although the range in which the estimator works correctly is wider for this bandwidth.

Table 6.4 shows the estimated values, $\hat{\mu}^{2,h,b}$, for the bootstrap bandwidth selectors h^* and b^*. Considering the study performed in Borrajo & Cao (2021), the resulting bandwidths in the case of nitrogen dioxide are not surprising, since in situations of little bias it is expected to obtain high values for these parameters. More surprising is the case of ozone (more bias), in which we would expect to obtain small values for both parameters, which does not happen in the case of h^*.

Table 6.4: Comparison of the full 15 years sample mean of the level of ozone and nitrogen dioxide with the mean of the analogous sample when the temperature is greater than or equal to 30 °C and the proposed estimator $\hat{\mu}$ for the values h^* and b^* obtained in the bootstrap implementation.

Variable	$\overline{X}$	$\overline{Y}$	h^*	b^*	$\hat{\mu}^{2,h^*,b^*}$
Ozone	64.44	45.35	199.05	0.0397	67.94
Nitrogen dioxide	23.88	22.28	79.24	50	23.90

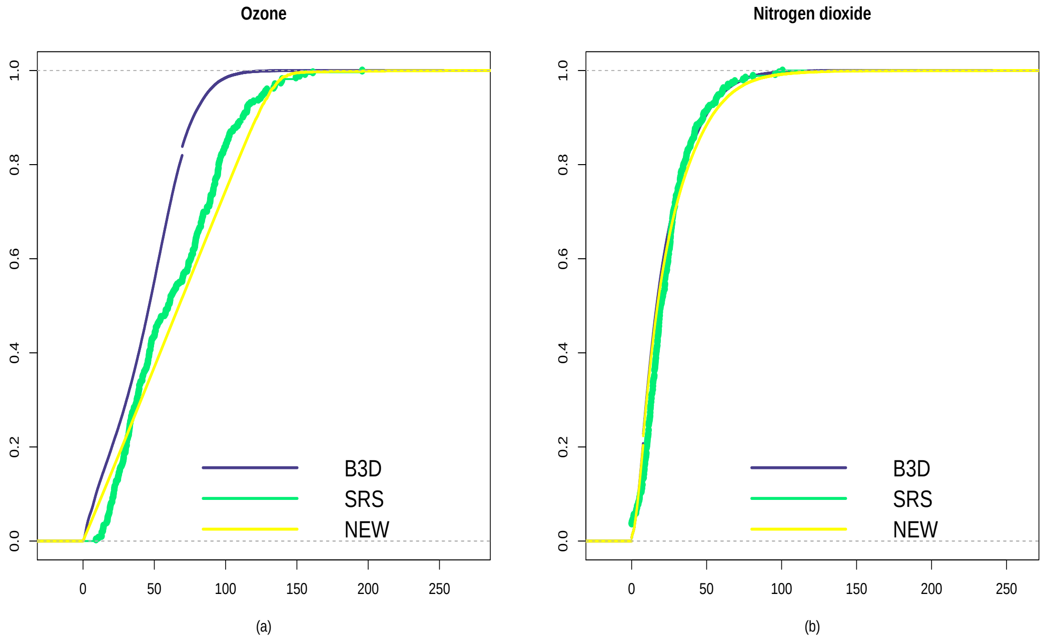

Figure 6.5: Estimated cumulative distribution functions involved in the case study with air quality data. (**a**) Empirical distribution function of the ozone level in the last 15 years (blue line), the analogue for temperatures greater than or equal to 30 °C (green line) and the estimated distribution function using the proposed estimator (yellow line). (**b**) Empirical distribution function of the nitrogen dioxide level in the last 15 years (blue line), the analogue for temperatures greater than or equal to 30 °C (green line) and the estimated distribution function using the proposed estimator (yellow line).

As already mentioned, the proposed method allows to solve other problems, such as, for example, the estimation of the cumulative distribution function. Although this requires a specific bandwidth selection, for simplicity we will use those obtained by the bootstrap algorithm for mean estimation. Figure 6.5 shows the estimated distribution function using our proposed estimator. This figure exhibits important differences in the case of ozone, which was expected, since the more bias, the more the proposed estimator can benefit and beat the classical estimators based on the two samples.

We can conclude that, in the case of estimating the mean and the distribution function of the level of nitrogen dioxide, it is irrelevant to use the classical estimators based on the two samples or our proposed estimator, since the results are very similar. However, in the case of ozone, things change. Although the proposed estimator gives a similar value for the mean and an estimated distribution function close to the one obtained in the SRS case, the difference is big enough to take it into consideration. This is a relevant issue due to the already mentioned problems caused by high values of this pollutant. This is not surprising in view of Figure 6.2.

6.3 Vodafone

A real data set from the Telco Company Vodafone ES is considered. It consists of nearly 2.5 million records and 176 variables concerning contracted services, application consumption, participation in campaigns and billing of their clients, among many other. We will focus on customer retention campaigns. According to the data protection law, clients identifiers have been previously anonymized.

We are interested in estimating the mean and the cumulative distribution function of an *index* constructed for Vodafone customers. This *index* is a new variable created from 14 significant variables that best reflect the costumer's tendency to leave the company. In this context, bias appears due to decisions taken by humans in the past and has been learned by their predictive models. For this purpose, we use (4.4), where $\mathbf{Y}$ is the data set with information of the target group (TG) for retention campaigns ($N = 194,010$), which is a biased sample of the population since it has been obtained from that biased models. The sample $\mathbf{X}$ is the data set of the universal control group (UCG) ($n = 1,466$) in these campaigns. This group corresponds to a simple random sample formed by the 1.5% of the company's customers.

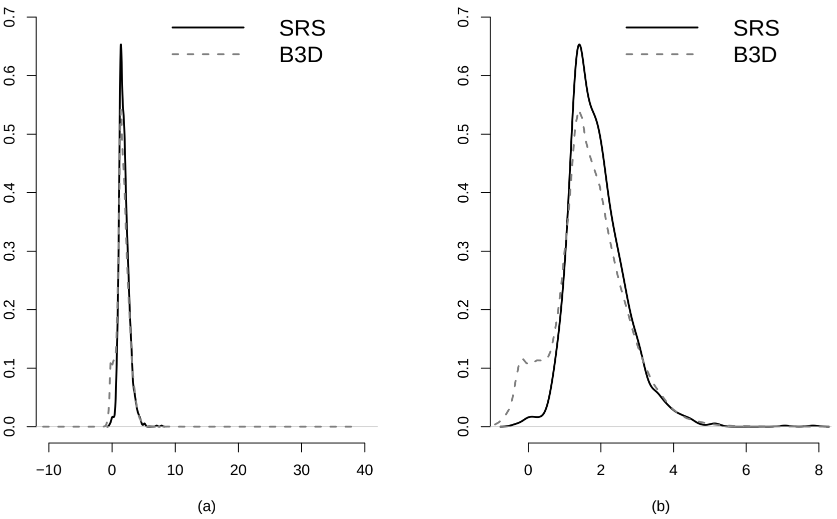

Figure 6.6: Estimated densities involved in the case study with Vodafone data. **(a)** Estimated densities of the *index* in the target group (dashed gray line) and in the universal control group (solid black line). **(b)** Zoom of the left panel.

To illustrate the difference between the density functions of the *index* in the TG and the UCG, kernel density estimations have been plotted in Figure 6.6 (a). Although, at a first glance, the two estimated densities seem to be similar, the zoom made in Figure 6.6 (b) exhibits some significant differences, such as the level of density at the mode and the left tail.

In fact, we can use the two-sample Kolmogorov-Smirnov test (Kolmogorov, 1933; Smirnov, 1939) to test for sampling bias. The p-value obtained (see Table 6.5) allows to reject the null hypothesis, so we conclude the presence of bias. Table 6.5 also shows the p-value obtained using the Student's t-test, which allows to reject that the means of the two populations are equal with the usual significance levels.

Table 6.5: p-values obtained for the two-sample Kolmogorov–Smirnov (KS) test and for the two-sample Student's t-test.

Variable	*KS test*	$\overline{X}$	$\overline{Y}$	*t-test*
index	$<2.2 \times 10^{-16}$	1.901294	1.706327	$<2.2 \times 10^{-16}$

As mentioned in Section 4.2, the cumulative distribution function, $F(t)$, can also be estimated considering $v(x) = \mathbf{1}_{\{x \leq t\}}$ in the proposed general procedure.

An adapted version for the Kolmogorov-Smirnov distance of the bootstrap algorithm proposed in Chapter 4 for the mean has been implemented. It is an algorithm for automatic bandwidth selection which allows to find values of h and b that provide a good performance of the cdf estimator. Despite the fact that in Chapter 4 a specific bootstrap algorithm for the mean was proposed, in order to obtain coherent estimates of the cdf and the mean and to avoid a high computational cost, we used the optimal bandwidths from this new algorithm to estimate both.

The bootstrap algorithm for automatic bandwidth selection in the cdf estimation problem consists of:

1. Based on the original $\mathbf{X}$ and $\mathbf{Y}$ samples, the estimated densities $\hat{f}_{h_{pil}}$ and $\hat{g}_{b_{pil}}$ (h_{pil} and b_{pil} denote the pilot bandwidths obtained from the rule-of-thumb method) are considered as the true population densities in the bootstrap world.

2. Bootstrap resamples, $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ and $\mathbf{Y}^* = (Y_1^*, \ldots, Y_N^*)$, of sizes n and N respectively, are obtained from the estimated densities $\hat{f}_{h_{pil}}$ and $\hat{g}_{b_{pil}}$ as follows:

   (a) $X_i^* = \psi_i^* + h_{pil} \cdot U_i$, where $\psi^* = (\psi_1^*, \ldots, \psi_n^*)$ is a simple random sample obtained from the empirical distribution computed with the values $\mathbf{X} = (X_1, \ldots, X_n)$ and $U = (U_1, \ldots, U_n)$, with U_i simulated from the density K (a $N(0,1)$ when considering a Gaussian kernel), for $i = 1, \ldots, n$.

   (b) $Y_i^* = \eta_i^* + b_{pil} \cdot V_i$, where $\eta^* = (\eta_1^*, \ldots, \eta_N^*)$ is a simple random sample obtained from the empirical distribution computed with the values $\mathbf{Y} = (Y_1, \ldots, Y_N)$ and $V = (V_1, \ldots, V_N)$, with V_i simulated from the density K (a $N(0,1)$ when considering a Gaussian kernel), for $i = 1, \ldots, N$.

3. The estimator $\hat{F}^{h,b,*}$ is computed using the resamples $\mathbf{X}^*$ and $\mathbf{Y}^*$ and considering a wide range of values for the smoothing parameters h and b.

4. Steps 2 and 3 are repeated a large number of times, B, in order to obtain a Monte Carlo approximation of the bootstrap Kolmogorov-Smirnov distance (d_{KS}^*) of the estimator,

$$d_{KS}^*(h,b) = \frac{1}{B} \sum_{j=1}^{B} \left(\sup_{x \in \mathbb{R}} |\hat{F}_j^{h,b,*}(x) - \hat{F}_{h_{pil}}(x)| \right).$$

5. The bandwidths h^* and b^* that minimize the function $d_{KS}^*(h,b)$ are considered as bootstrap bandwidth selectors.

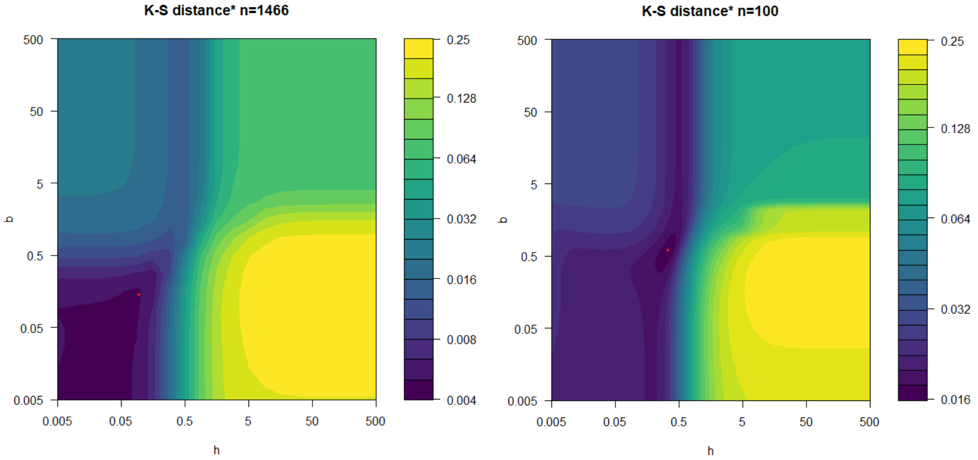

Figure 6.7: Logarithm of the bootstrap version of the mean Kolmogorov-Smirnov distance for the cdf estimator for the *index* as a function of h and b. **(a)** Considering as $\mathbf{X}$ the data set of the UCG (n =1,466). **(b)** Considering as $\mathbf{X}$ a subsample of size $n = 100$ of the UCG. The red dot represents the minimal value of the d^*_{KS}.

Figure 6.7 (a) shows the logarithm of the bootstrap version of the mean Kolmogorov-Smirnov distance for this estimator. This figure provides some estimate of the optimal values for the two bandwidths, those that minimize the bootstrap version of this distance.

The values of $\overline{X}$ and $\overline{Y}$ have been computed (see Table 6.6) and the cumulative distribution functions of the samples involved have been plotted (see Figure 6.8 (a)). Table 6.6 shows the estimated value $\hat{\mu}^{2,h,b}$ (first row) and Figure 6.8 (a) depicts the estimated distribution function using the bootstrap bandwidth selectors h^* and b^*. These results exhibit a few differences with the estimator based on the SRS. This was expected, since the "small" sample size ($n = 1,466$) is quite large in this data set.

Table 6.6: Comparison of the sample mean of the *index* in the universal control group and in a subsample of size $n = 100$ of this group ($\overline{X}$) with the mean of the target group of the retention campaign ($\overline{Y}$) and the proposed estimator, $\hat{\mu}$, using the values h^* and b^* obtained in the bootstrap implementation.

Variable	n	$\overline{X}$	$\overline{Y}$	h^*	b^*	$\hat{\mu}^{2,h^*,b^*}$
index	1,466	1.9013	1.7063	0.1104	0.1662	1.8947
index	100	2.0251	1.7063	0.3339	0.5691	1.9563

Just to show the effect of n in the practical behaviour of the estimator, we considered a simple random subsample of just 100 clients in the universal control group. The second row in Table 6.6 and Figure 6.8 (b) shows how, in that case, our proposed estimators for the mean and the cdf differ more from the classic estimators based on the new SRS for this more moderate sample size ($n = 100$). In fact, the new estimations for $n = 100$ are close to the ones obtained with the original $\mathbf{X}$ ($n = 1,466$), whose good performance has already been shown.

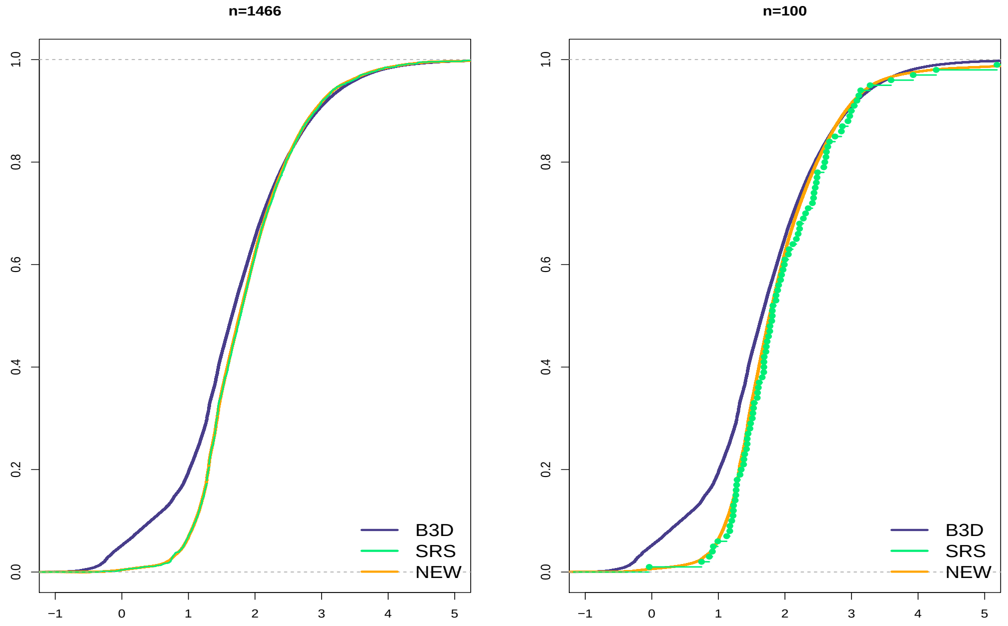

Figure 6.8: Estimated cumulative distribution functions involved in the case study with Vodafone data. **(a)** Comparison of the empirical distribution functions of the *index* in the TG (blue line), the analogue in the UCG (green line) and the estimated distribution function using the proposed estimator (orange line) with $h^* = 0.1104$ and $b^* = 0.1662$. **(b)** Comparison of the empirical distribution functions of the *index* in the TG (blue line), the analogue in a subsample of size $n = 100$ of the UCG (green line) and the estimated distribution function using the proposed estimator (orange line) with $h^* = 0.3339$ and $b^* = 0.5691$.

6.4 COVID-19

In this section, an analysis about the age of infection with COVID-19 in Spain is performed. Two databases from the Centro de Coordinación de Alertas y Emergencias Sanitarias (CCAES) are considered. One of them contains the results of the National Study of sero-Epidemiology of the infection with SARS-CoV-2 in Spain

(ENECOVID), carried out on a sample of 68,296 people obtained through stratified two-stages sampling. This study was conducted in 3 rounds from April 27th to June 22nd, 2020, showing an approximated national prevalence of 5%. An additional fourth round, not considered here, corresponds to November 2020. The other data set used contains the data of confirmed cases of COVID-19, which are obtained from the information that the Autonomous Communities notify to the National Epidemiological Surveillance Network (RENAVE), within the framework of the National Surveillance Strategy (Strategy for early detection, surveillance and control of COVID-19) and through the web platform of the Spanish Surveillance System (SiViES). This second database contains information of more than 3 million records about cases of infection identified until March 4th, 2021. In addition to the variable age, there are other relevant variables, such as the one that indicates the need for hospitalization. According to the data protection law, pacients identifiers have been previously anonymized.

We address the problem of mean age estimation for COVID-19 patients in Setup 1 and Setup 2. For this purpose, we use (4.4) and (5.4), considering as $\mathbf{X}$ the sample of the ages of the participants in the ENECOVID survey whose analysis showed the presence of antibodies during the weeks of the study (n =4,555), as $\mathbf{Y}$ the ages of the identified cases from RENAVE until May 11th, 2020 (N = 228,879) and as $\mathbf{Z}$ a subsample of $\mathbf{Y}$ corresponding to the ages of the patients who were hospitalized until that time (n =106,637). The reason for restricting the second data set is that the criteria for conducting diagnostic tests changed on May 11th, 2020.

On the one hand, the design of the ENECOVID study, available at `https://portalcne.isciii.es/enecovid19/`, guarantees that sample $\mathbf{X}$ is very close to a simple random sample of the population in which we are interested. On the other hand, sample $\mathbf{Y}$ is clearly biased due to the already known high presence of asymptomatic cases among the population that are not identified by the health system.

It is reasonable to think that the bias mechanism is similar in both setups, since the main characteristic to identify someone infected with SARS-CoV-2 is an increase in the severity of their symptoms, as well as to hospitalize an already identified patient. To illustrate the difference between the density functions of the age of those infected with SARS-CoV-2 in the samples considered, three kernel density estimations have been plotted in Figure 6.9.

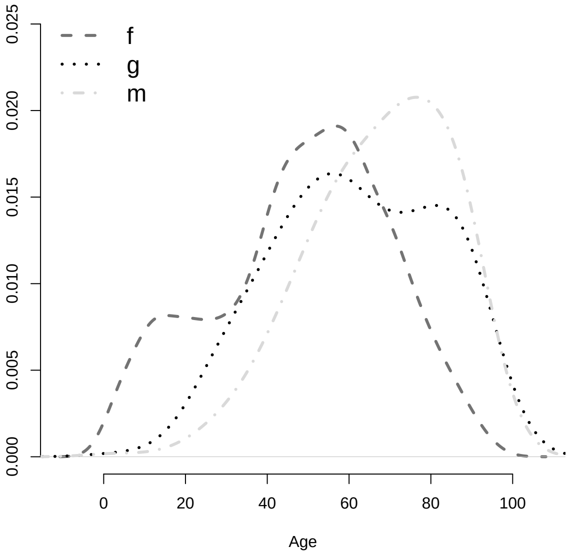

Figure 6.9: Densities involved in the case study with COVID-19 data. Densities of the age of the positive cases among the participants in the ENECOVID survey (dashed dark gray line), the positive cases identified by RENAVE (dotted black line) and those who required hospitalization (dashed light gray line) until May 11th, 2020.

To test the presence of sampling bias and the equality of means we use the two-sample Kolmogorov-Smirnov test, throught the *ks.test* function, and the Student's t-test, respectively. The p-values obtained (see Table 6.7) allow to reject the null hypothesis in both cases, so we conclude the presence of bias.

Table 6.7: p-values obtained for the two-sample Kolmogorov–Smirnov (KS) test and for the two-sample Student's t-test in both setups.

Setup	*ks.test*	$\overline{X}$	$\overline{Y}$	$\overline{Z}$	*t-test*
1	$<2.2 \times 10^{-16}$	48.57	62.00	-	$<2.2 \times 10^{-16}$
2	$<2.2 \times 10^{-16}$	-	62.00	67.61	$<2.2 \times 10^{-16}$

Table 6.8: Comparison of some bias indices in both setups.

Setup	i_1	i_2	i_3	i_6
1	0.6457611	0.2152082	0.04953267	0.7357414
2	0.2988901	0.1488982	0.03567542	0.2208698

Table 6.8 shows the estimated values of some of the bias indices presented in

Chapter 3. All the indices considered show that bias is greater in Setup 1 than in Setup 2.

6.4.1 Setup 1

In this context, we computed the values of $\overline{X}$ and $\overline{Y}$. Since the real mean of the variable age in the population of COVID-19 pacients, μ, is unknown, we use the bootstrap bandwidth selection method presented in Subsection 4.4 in order to select the values of h and b for the estimator (4.4).

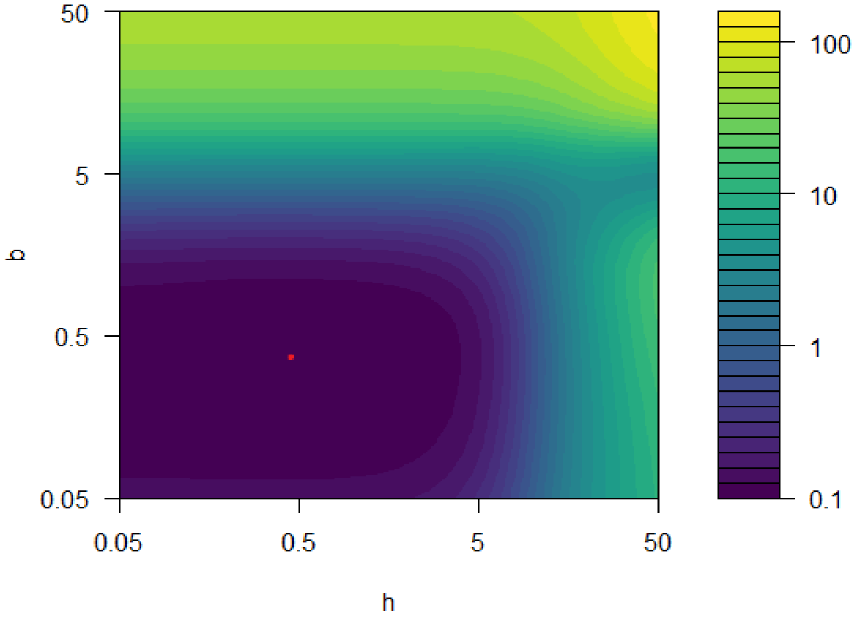

Figure 6.10: Bootstrap estimation of the mean squared error (MSE^*) of the proposed estimator for the mean age of the infected with SARS-CoV-2 as a function of h and b in Setup 1. The red dot represents the minimal value of the MSE^*.

Figure 6.10 shows the bootstrap mean squared error of the proposed estimator for bandwidth selection in Setup 1. This figure provides the bandwidths that minimize the MSE^* and which will be used to estimate the mean age for COVID-19 patients. As it can be observed in this figure, the estimator works correctly for a large number of combinations of h and b (region in purple).

Table 6.9 shows how the mean age of the COVID-19 cases identified by RENAVE, 62 years, is greater than that of the ENECOVID study participants who had passed the infection, 48.57 years. This makes perfect sense since during the first epidemic wave, the diagnosis of severe cases that required hospitalization was prioritized, a situation observed in older age groups. On the contrary, the ENECOVID survey estimates the age of people who were infected with SARS-CoV-2 in that period, regardless of whether they were symptomatic, asymptomatic or had contact with health systems, therefore their mean age was lower than that of the cases notified to RENAVE. This table also shows the values of the bootstrap bandwidth selectors and the estimated value $\hat{\mu}^{2,h^*,b^*}$. Regarding the analysis performed in Chapter 4, the obtained values for the bandwidths and the mean estimator are not surprising, since in situations in which bias is quite significant (see Table 6.8), we would expect to obtain small values for both parameters and an estimation of μ very close to $\overline{X}$.

Table 6.9: Comparison of the mean age of the ENECOVID respondents who passed the disease ($\overline{X}$) with the mean age of the cases identified by RENAVE ($\overline{Y}$) and the proposed estimator $\hat{\mu}$ for the values h^* and b^* obtained in the bootstrap implementation for Setup 1.

Variable	$\overline{X}$	$\overline{Y}$	h^*	b^*	$\hat{\mu}^{2,h^*,b^*}$
Age	48.57	62.00	0.43	0.40	48.62

6.4.2 Setup 2

In Setup 2, we computed the values of $\overline{Z}$ and $\overline{Y}$ and we use the bootstrap method proposed in Section 5.4 to select the bandwidths h and b that provide a good performance of the estimator (5.4).

Since the sample $\mathbf{X}$ is not available in this setup, we proceed as shown in the bootstrap algorithm in Section 5.4. Figure 6.11 shows the estimated density function $\hat{f}_{h_{pil},b_{pil}}$ for a suitable choice of the pilot bandwidths, $h_{pil} = 6.22$ and $b_{pil} = 6.94$.

The differences observed between the estimated density $\hat{f}_h$ with $h = 3.567$ obtained from sample $\mathbf{X}$ in Setup 1 (Figure 6.9) and the estimated density $\hat{f}_{h_{pil},b_{pil}}$ using samples $\mathbf{Y}$ and $\mathbf{Z}$ in Setup 2 (Figure 6.11) are due to the fact that the bias mechanism, although similar, is not exactly equal. For the same reason, the estimated mean from the density $\hat{f}_{h_{pil},b_{pil}}$, $\mu^* = 54.07$, slightly differs from $\overline{X} = 48.57$.

Figure 6.12 shows the bootstrap mean squared error of the proposed estimator for bandwidth selection in Setup 2. This figure provides the bandwidths that minimize the MSE^* and which will be used to estimate the mean age of COVID-19 pacients in this setup.

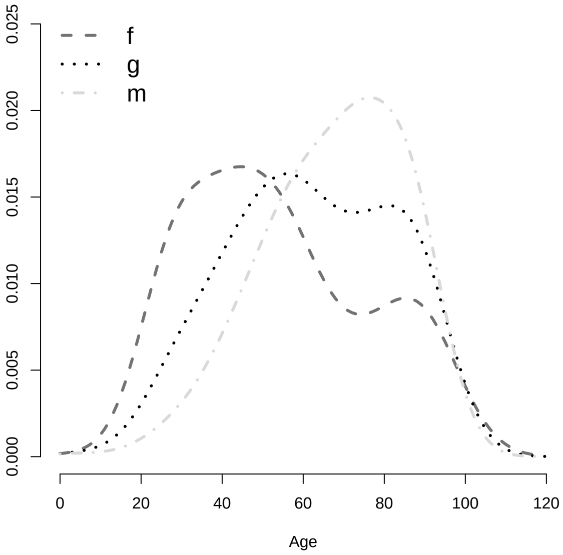

Figure 6.11: Estimated densities of the age of the positive cases identified by RENAVE (dotted black line) with $b_{pil} = 6.94$ and those who required hospitalization (dashed light gray line) with $h_{pil} = 6.22$ until May 11th, 2020, and resulting estimated density $\hat{f}$ (dashed dark gray line).

Table 6.10 includes the values of the classical means, 62 year in the case of the biased sample and 67.61 years in the twice biased sample. In this case, the difference between the sample means is much smaller compared to that obtained in Setup 1. This is due to the fact that in the first wave, many of the identified cases had severe symptoms, requiring hospitalization many of them (46.6 %) and many others could not be hospitalized due to the epidemic situation at that time. This table also shows the values of the bootstrap bandwidth selectors and the estimated value $\hat{\hat{\mu}}^{2,h^*,b^*}$. In Setup 2, in which sample $\mathbf{X}$ is not observed, we would obtain an estimation of $\hat{\hat{\mu}} = 53.84$ years. Assuming that the true mean, μ, is close to the value $\overline{X} = 48.57$, our estimation differs substantially from it but significantly improves those offered by the classical estimators: $\overline{Y} = 62$ and $\overline{Z} = 67.61$ years.

Table 6.10: Comparison of the the mean age of the cases identified by RENAVE ($\overline{Y}$) with the mean age of those that required hospitalization ($\overline{Z}$) and the proposed estimator $\hat{\hat{\mu}}$ for the values h^* and b^* obtained in the bootstrap implementation for Setup 2.

Variable	$\overline{Y}$	$\overline{Z}$	h^*	b^*	$\hat{\hat{\mu}}^{2,h^*,b^*}$
Age	62.00	67.61	9.18	13.57	53.84

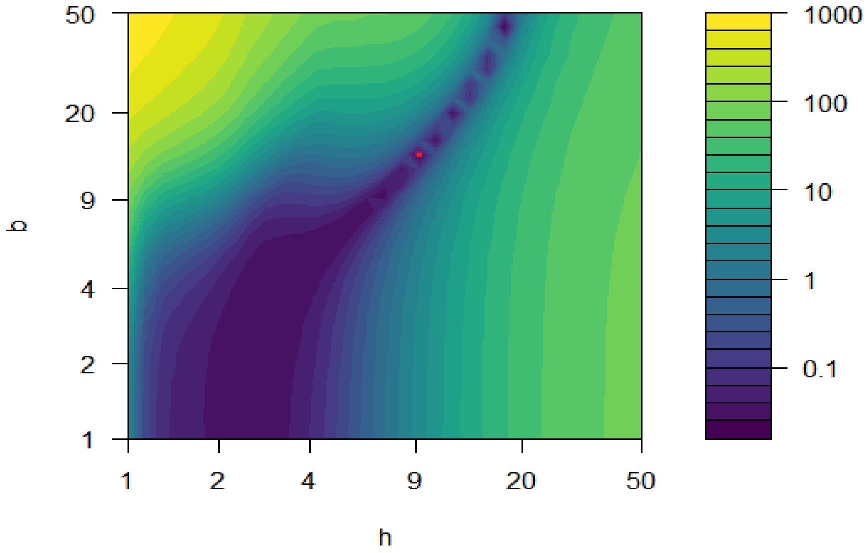

Figure 6.12: Bootstrap estimation of the mean squared error (MSE^*) of the proposed estimator for the mean of the age of COVID-19 pacients as a function of h and b in Setup 2. The red dot represents the minimal value of the MSE^*.

Chapter 7

Conclusion and further research

In Section 7.1 the main conclusions of the work developed in this thesis are included. Section 7.2 contains some interesting challenges that remain as open problems. They are considered to be dealt in the future.

7.1 Conclusion

In the era of big data, sampling bias is more present than before in statistical data analysis. Testing for sampling bias is an extremely important issue in a big data context. Several existing methods have been used to test for no sampling bias (i.e. $f = g$ in Setup 1 and $g = m$ in Setup 2). Of course, the fact that $N/n \to \infty$ makes a difference with the classical asymptotic theory.

Assuming that our large-sized sample is coming from a distribution which is different from the one we are interested in, two setups are considered to correct for this sampling bias. The first one assumes that another simple random sample (possibly of small or moderate size) from the real population is available. The second setup assumes that a doubly biased sample of small size is available. Both setups have been studied in detail. Density-based estimators for the mean of a transformation of the underlying population are proposed in both setups. The estimator in Setup 1 reduces to classical estimators (sample means of the transformation of either the large-sized biased sample and the small-sized sample from the real population in Setup 1) when using equal and extreme bandwidths. Similar comments apply to Setup 2. Asymptotic properties and Monte Carlo simulations show the very good performance of the proposed estimators and its rather stable behavior as a function of the two smoothing parameters needed. Moreover, both asymptotic theory as well as simulations show that optimal bandwidths satisfy a non-standard limit condition, namely they tend to positive constants. This is a very striking condition since implies inconsistency of the nonparametric density estimators. Anyhow, the region for the bandwidths where the new estimator performs better than the sample means is rather wide. This makes the problem of bandwidth selection not so critical. Bootstrap algorithms for bandwidth selection seems to work well in practice. Finally, several applications to real datasets show the good performance of the methods proposed throughout this thesis.

7.2 Future work

The research carried out so far opens further interesting topics in the field:

- Half sampling devices could be useful when defining $\hat{\mu}$ to force independence between the subsamples used for density estimation and those used for evaluation of kernel estimators. Some loss of power of the methods is anticipated, but the proofs for the asymptotic results would probably be simpler.

- Application of the methods proposed to the estimation of the variance-covariance matrix and the correlation matrix. This would allow applying the results to carry out principal component analysis (PCA) and linear discriminant analysis (LDA).

- Adaptation of the bootstrap algorithm to obtain the bandwidth selectors for the estimation of the general parameter μ_v.

- Extension of the methods proposed to the study of other type of variables, such as the extension to categorical settings using, for instance, the estimators proposed in Li & Racine (2003) or the extension to multidimensional $\mathbf{X}$ and $\mathbf{Y}$; including covariate dependence in the biasing weight.

- Adaptation of the proposed methods to the study of dependent samples of $\mathbf{X}$ and $\mathbf{Y}$, for example, considering $\mathbf{X}$ a subsample of $\mathbf{Y}$ obtained by an acceptance/rejection method.

- Developing an R package to apply the proposed statistical techniques to real data sets.

Appendix A

Appendix

A.1 Proofs of the results in Chapter 4

A.1.1 Proof of Lemma 4.2.1

Let us first remember the expressions of the estimators involved in Lemma 4.2.1:

$$\hat{\mu}_v^{1,h,b} = \frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)}{\hat{w}_{h,b}\left(Y_i\right)} = \frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)\hat{f}_h\left(Y_i\right)}{\hat{g}_b\left(Y_i\right)}, \tag{4.3}$$

$$\hat{\mu}_v^{2,h,b} = \frac{\frac{1}{N}\sum_{i=1}^{N} \frac{v(Y_i)}{\hat{w}_{h,b}\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{\hat{w}_{h,b}\left(Y_i\right)}} = \frac{\frac{1}{N}\sum_{i=1}^{N} v(Y_i)\frac{\hat{f}_h\left(Y_i\right)}{\hat{g}_b\left(Y_i\right)}}{\frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h\left(Y_i\right)}{\hat{g}_b\left(Y_i\right)}}. \tag{4.4}$$

Lemma 4.2.1. *Consider two fixed samples, $\boldsymbol{X}$ and $\boldsymbol{Y}$, with sizes n and N, and equal smoothing parameters, $h = b$. Then the extreme undersmoothing and oversmoothing versions of the estimators in (4.3) and (4.4) is reduced as follows. If Condition A1 holds, then*

$$\lim_{h\to\infty} \hat{\mu}^{1,h,h} = \overline{Y}, \ \lim_{h\to\infty} \hat{\mu}^{2,h,h} = \overline{Y}.$$

Under Condition A2, and assuming there are no ties between any Y_i and any X_j, we have

$$\lim_{h\to 0^+} \hat{\mu}^{1,h,h} = 0.$$

Assuming there are no ties in the union of the X and Y samples, Conditions A2, A3 and A4 imply

$$\lim_{h\to 0^+} \hat{\mu}^{2,h,h} = \arg\min_{y\in\{Y_1,\ldots Y_N\}} \min_{j\in\{1,\ldots,n\}} \left|y - X_j\right|.$$

Finally, either assuming Condition A3 or A5 and defining $b_N = (\log N)/N$ *we have*

$$\hat{\mu}^{1,b_N,b_N} \simeq \overline{X},\ \hat{\mu}^{2,b_N,b_N} \simeq \overline{X}.$$

Proof. Starting from (4.3), straightforward calculations lead to

$$\hat{\mu}^{1,h,h} = \frac{1}{n}\sum_{i=1}^{N} Y_i \sum_{j=1}^{n} \frac{K\left(\frac{Y_i - X_j}{h}\right)}{\sum_{\ell=1}^{N} K\left(\frac{Y_i - Y_\ell}{h}\right)}. \tag{A.1}$$

Now, using A1, all the terms of the form $K\left(\frac{a}{h}\right)$ have limit $K(0)$ when $h \to \infty$. Consequently,

$$\lim_{h\to\infty} \hat{\mu}^{1,h,h} = \frac{1}{n}\sum_{i=1}^{N} Y_i \sum_{j=1}^{n} \frac{K(0)}{\sum_{\ell=1}^{N} K(0)} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \overline{Y}.$$

Parallel arguments can be used to prove that the denominator in (4.4) has limit 1 when $h = b \to \infty$. Thus $\lim_{h\to\infty} \hat{\mu}^{2,h,h} = \overline{Y}$.

To prove the second part of the lemma, recall that, using A2, all the terms of the form $K(a/h)$, with $a \neq 0$, in (A.1) have limit 0 when $h \to 0^+$. As a consequence, $\lim_{h\to 0^+} \hat{\mu}^{1,h,h} = 0$. Similarly, the limit of the denominator of (4.4) when $h = b \to 0^+$ is also zero. So to study the limit behaviour of $\hat{\mu}^{2,h,h}$ we need to obtain the dominant terms in the numerator and denominator in (4.4) when $h = b$. Its is very easy to prove that the dominant term of (A.1) $h \to 0^+$ is the same as the dominant term of

$$\frac{1}{nK(0)}\sum_{i=1}^{N} Y_i \sum_{j=1}^{n} K\left(\frac{Y_i - X_j}{h}\right),$$

whose limit behaviour, in view of A2-A4, is the one of

$$\frac{K\left(\frac{b_{\min}}{h}\right) Y_{i_0}}{nK(0)},$$

where $b_{\min} = |Y_{i_0} - X_{j_0}| = \min B$ for the set B in A4. Parallel arguments show that the limit behaviour of the denominator of (4.4) when $h = b \to 0^+$ is the one of

$$\frac{K\left(\frac{b_{\min}}{h}\right)}{nK(0)}.$$

As a consequence,

$$\lim_{h\to 0^+} \hat{\mu}^{2,h,h} = Y_{i_0} = \arg \min_{y\in\{Y_1,\dots Y_N\}} \min_{j\in\{1,\dots,n\}} |y - X_j|.$$

For the last part of the lemma, using (A.1) the estimator $\hat{\mu}^{1,b_N,b_N}$ can be approximated as follows

$$\hat{\mu}^{1,b_N,b_N} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{N}\frac{K\left(\frac{Y_i-X_j}{b_N}\right)}{\sum_{\ell=1}^{N}K\left(\frac{Y_i-Y_\ell}{b_N}\right)}Y_i \simeq \tilde{\mu}^{1,b_N,b_N} \tag{A.2}$$

where

$$\tilde{\mu}^{1,b_N,b_N} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{N}\frac{K\left(\frac{Y_i-X_j}{b_N}\right)(Y_i-X_j)}{\sum_{\ell=1}^{N}K\left(\frac{X_j-Y_\ell}{b_N}\right)} + \overline{X}. \tag{A.3}$$

Now, Condition A5 implies that

$$\left|\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{N}\frac{K\left(\frac{Y_i-X_j}{b_N}\right)(Y_i-X_j)}{\sum_{\ell=1}^{N}K\left(\frac{X_j-Y_\ell}{b_N}\right)}\right| \leq b_N\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{N}\frac{K\left(\frac{Y_i-X_j}{b_N}\right)\left|\frac{Y_i-X_j}{b_N}\right|}{\sum_{\ell=1}^{N}K\left(\frac{X_j-Y_\ell}{b_N}\right)} \leq b_N. \tag{A.4}$$

Using this expression in (A.3) gives $\left|\tilde{\mu}^{1,b_N,b_N} - \overline{X}\right| \leq b_N$, which, together with (A.2), implies $\hat{\mu}^{1,b_N,b_N} \simeq \tilde{\mu}^{1,b_N,b_N} \simeq \overline{X}$. In a similar way, the denominator of $\hat{\mu}^{2,b_N,b_N}$ in (4.4) can be proven to be approximately 1. Consequently $\hat{\mu}^{2,b_N,b_N} \simeq \overline{X}$.

When assuming Condition A3, the left-hand side of expression (A.4) can be handled in the following way

$$\left|\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{N}\frac{K\left(\frac{Y_i-X_j}{b_N}\right)(Y_i-X_j)}{\sum_{\ell=1}^{N}K\left(\frac{X_j-Y_\ell}{b_N}\right)}\right| \simeq \frac{1}{n}\sum_{j=1}^{n}\min\left\{\left|Y_i-X_j\right|; i=1,\ldots,N\right\} \simeq 0,$$

if N is much larger than n. Consequently $\hat{\mu}^{1,b_N,b_N} \simeq \overline{X}$ and $\hat{\mu}^{2,b_N,b_N} \simeq \overline{X}$ also when assuming Condition A3 and the lemma is proven. □

A.1.2 General weighted estimator in Subsection 4.2.3

Lemma A.1.1. *The mean squared error of the estimator*

$$\hat{\mu}_v^{2,\tau} = \frac{\frac{1}{N}\sum_{i=1}^{N}\tau\left(Y_i\right)v(Y_i)}{\frac{1}{N}\sum_{i=1}^{N}\tau\left(Y_i\right)} \tag{A.5}$$

is

$$\begin{aligned} MSE(\hat{\mu}_v^{2,\tau}) \simeq & \left[\int \tau(y)(v(y)-\mu_v)g(y)dy\right]^2 \\ + & \frac{1}{N}\int\left[\tau(y)(v(y)-\mu_v) - \int\tau(x)(v(x)-\mu_v)g(x)dx\right]^2 g(y)dy. \end{aligned} \tag{A.6}$$

Proof. The estimator in (A.5) can be expressed as the ratio:

$$\hat{\mu}_v^{2,\tau} = \frac{\hat{\mu}_v^\tau}{\hat{\mu}_0^\tau}$$

where

$$\hat{\mu}_v^\tau := \frac{1}{N}\sum_{i=1}^{N}\tau(Y_i)v(Y_i) \quad \text{and} \quad \hat{\mu}_0^\tau := \frac{1}{N}\sum_{i=1}^{N}\tau(Y_i).$$

So,

$$\begin{aligned}
\hat{\mu}_v^{2,\tau} - \mu_v &= \frac{\hat{\mu}_v^\tau}{\hat{\mu}_0^\tau} - \mu_v = \frac{\hat{\mu}_v^\tau}{\hat{\mu}_0^\tau}\left(\hat{\mu}_0^\tau + 1 - \hat{\mu}_0^\tau\right) - \mu_v = \hat{\mu}_v^\tau - \mu_v + \frac{\hat{\mu}_v^\tau}{\hat{\mu}_0^\tau}\left(1 - \hat{\mu}_0^\tau\right) \\
&= \hat{\mu}_v^\tau - \mu_v + \frac{\hat{\mu}_v^\tau}{\hat{\mu}_0^\tau}\left(\hat{\mu}_0^\tau + 1 - \hat{\mu}_0^\tau\right)\left(1 - \hat{\mu}_0^\tau\right) \\
&= \hat{\mu}_v^\tau - \mu_v + \hat{\mu}_v^\tau\left(1 - \hat{\mu}_0^\tau\right) + \frac{\hat{\mu}_v^\tau}{\hat{\mu}_0^\tau}\left(1 - \hat{\mu}_0^\tau\right)^2 \\
&= \hat{\mu}_v^\tau - \mu_v + \mu_v\left(1 - \hat{\mu}_0^\tau\right) + \left(\hat{\mu}_v^\tau - \mu_v\right)\left(1 - \hat{\mu}_0^\tau\right) + \hat{\mu}_v^\tau\left(1 - \hat{\mu}_0^\tau\right)^2 \\
&= A + B + C + D, \qquad \text{(A.7)}
\end{aligned}$$

with $A = \hat{\mu}_v^\tau - \mu_v$, $B = \mu_v\left(1 - \hat{\mu}_0^\tau\right)$, $C = \left(\hat{\mu}_v^\tau - \mu_v\right)\left(1 - \hat{\mu}_0^\tau\right)$ and $D = \hat{\mu}_v^{2,\tau}\left(1 - \hat{\mu}_0^\tau\right)^2$, where C and D are negligible.

We now analyze the terms A and B:

$$\begin{aligned}
E(A) &= E(\hat{\mu}_v^\tau) - \mu_v = E(\tau(Y_1)v(Y_1)) - \mu_v, \qquad \text{(A.8)} \\
Var(A) &= Var\left(\hat{\mu}_v^\tau\right) = \frac{Var\left(\tau(Y_1)v(Y_1)\right)}{N}, \qquad \text{(A.9)} \\
E(B) &= \mu_v E\left(1 - \hat{\mu}_0^\tau\right) = \mu_v\left[1 - E\left(\hat{\mu}_0^\tau\right)\right] = \mu_v\left[1 - E\left(\tau(Y_1)\right)\right], \qquad \text{(A.10)} \\
Var(B) &= \mu_v^2 Var\left(1 - \hat{\mu}_0^\tau\right) = \mu_v^2 Var\left(\hat{\mu}_0^\tau\right) = \mu_v^2\frac{Var\left(\tau(Y_1)\right)}{N}, \qquad \text{(A.11)} \\
Cov(A,B) &= Cov\left(\hat{\mu}_v^\tau - \mu_v, \mu_v\left(1 - \hat{\mu}_0^\tau\right)\right) \\
&= \mu Cov\left(\hat{\mu}_v^\tau - \mu_v, 1 - \hat{\mu}_0^\tau\right) = -\mu_v Cov\left(\hat{\mu}_v^\tau, \hat{\mu}_0^\tau\right) \\
&= -\mu_v Cov\left(\frac{1}{N}\sum_{i=1}^{N}\tau(Y_i)v(Y_i), \frac{1}{N}\sum_{j=1}^{N}\tau(Y_j)\right) \\
&= -\mu_v\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}Cov\left(\tau(Y_i)v(Y_i), \tau(Y_j)\right) \\
&= -\frac{\mu}{N^2}\sum_{i=1}^{N}Cov\left(\tau(Y_i)v(Y_i), \tau(Y_i)\right) \\
&= -\frac{\mu}{N}Cov\left(\tau(Y_1)v(Y_1), \tau(Y_1)\right). \qquad \text{(A.12)}
\end{aligned}$$

Using (A.7) and then (A.8)-(A.12) we have:

$$\begin{aligned}
E\left[\left(\hat{\mu}_v^{2,\tau}-\mu_v\right)^2\right] &= \left[E(\hat{\mu}^{2,\tau})_v-\mu_v\right]^2+Var(\hat{\mu}_v^{2,\tau}) \\
&\simeq [E(A)+E(B)]^2+Var(A)+Var(B)+2Cov(A,B) \\
&= \left[E(\tau(Y_1)v(Y_1))-\mu_v+\mu_v\left[1-E\left(\tau(Y_1)\right)\right]\right]^2 \\
&+ \frac{1}{N}\left[Var\left(\tau(Y_1)v(Y_1)\right)+\mu_v^2 Var\left(\tau(Y_1)\right)-2\mu_v Cov\left(\tau(Y_1)v(Y_1),\tau(Y_1)\right)\right] \\
&= \left[E(\tau(Y_1)v(Y_1))-\mu_v+\mu_v-\mu_v E\left(\tau(Y_1)\right)\right]^2 \\
&+ \frac{1}{N}\left[Var\left(\tau(Y_1)v(Y_1)\right)+Var\left(\mu_v\tau(Y_1)\right)-2Cov\left(\tau(Y_1)v(Y_1),\mu_v\tau(Y_1)\right)\right] \\
&= \left[E(\tau(Y_1)v(Y_1))-\mu_v E\left(\tau(Y_1)\right)\right]^2+\frac{1}{N}Var\left(\tau(Y_1)v(Y_1)-\mu_v\tau(Y_1)\right) \\
&= \left[E\left(\tau(Y_1)(v(Y_1)-\mu_v)\right)\right]^2+\frac{1}{N}Var\left(\tau(Y_1)(v(Y_1)-\mu_v)\right). \qquad \text{(A.13)}
\end{aligned}$$

But

$$E\left(\tau(Y_1)(v(Y_1)-\mu_v)\right)=\int\tau(y)(v(y)-\mu_v)g(y)dy \qquad \text{(A.14)}$$

and

$$Var\left(\tau(Y_1)(v(Y_1)-\mu_v)\right)=\int\left[\tau(y)(v(y)-\mu_v)-\int\tau(x)(v(x)-\mu_v)g(x)dx\right]^2 g(y)dy. \qquad \text{(A.15)}$$

As a consequence of (A.13), (A.14) and (A.15) we obtain (A.6). □

A.1.3 Proof of Theorem 4.2.1

Let us first state an auxiliary lemma:

Lemma 4.5.1. *The difference $\hat{\mu}_v-\mu_v$ can be expressed as follows*

$$\hat{\mu}_v-\mu_v=\widehat{A}+\widehat{A}\left(1-\widehat{B}\right)+\frac{\widehat{A}\left(1-\widehat{B}\right)^2}{\widehat{B}}\simeq\widehat{A}, \qquad \text{(4.10)}$$

where

$$\widehat{A}=\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}(v(Y_i)-\mu_v) \qquad \text{(4.11)}$$

and

$$\widehat{B}=\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}. \qquad \text{(4.12)}$$

Proof. Considering the estimator $\hat{\mu}$ defined in (4.4), the difference $\hat{\mu}_v - \mu_v$ can be expressed as follows:

$$\hat{\mu}_v - \mu_v = \frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}(v(Y_i) - \mu_v)}{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}} = \frac{\widehat{A}}{\widehat{B}},$$

being $\widehat{A}$ and $\widehat{B}$ the terms defined in (4.11) and (4.12), respectively.

Intuitively $\widehat{B}$ is a consistent estimator of

$$B = E\left(\frac{f(Y)}{g(Y)}\right) = \int \frac{f(y)}{g(y)} g(y)dy = \int f(y)dy = 1, \tag{A.16}$$

so the most important term to study is $\widehat{A}$, but for a final result, properties that jointly consider $\widehat{A}$ and $\widehat{B}$ will be needed. For instance,

$$\begin{aligned}\hat{\mu}_v - \mu_v &= \frac{\widehat{A}}{\widehat{B}} = \frac{\widehat{A}}{\widehat{B}}\left(\widehat{B} + \left(1 - \widehat{B}\right)\right) = \widehat{A} + \frac{\widehat{A}\left(1 - \widehat{B}\right)}{\widehat{B}} \\ &= \widehat{A} + \frac{\widehat{A}\left(1 - \widehat{B}\right)}{\widehat{B}}\left(\widehat{B} + \left(1 - \widehat{B}\right)\right) = \widehat{A} + \widehat{A}\left(1 - \widehat{B}\right) + \frac{\widehat{A}\left(1 - \widehat{B}\right)^2}{\widehat{B}}\end{aligned}$$

and this expression can be iterated as much as needed in terms of negligibility of the term $\dfrac{\widehat{A}\left(1 - \widehat{B}\right)^r}{\widehat{B}}$ for a suitable $r \in \mathbb{N}$.

As a consequence of (A.16), we obtain (4.10).

□

The term $\widehat{A}$ can be split in different terms:

$$\widehat{A} = \widehat{A}_1 + \widehat{A}_2 - \widehat{A}_3 - \widehat{A}_4 + \widehat{A}_5,$$

where

$$\begin{aligned}\widehat{A}_1 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)}{g(Y_i)}(v(Y_i) - \mu_v), \\ \widehat{A}_2 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i) - f(Y_i)}{g(Y_i)}(v(Y_i) - \mu_v), \\ \widehat{A}_3 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)(\hat{g}_b(Y_i) - g(Y_i))}{g(Y_i)^2}(v(Y_i) - \mu_v),\end{aligned}$$

$$\begin{aligned}
\widehat{A}_4 &:= \frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{f}_h(Y_i)-f(Y_i))(\hat{g}_b(Y_i)-g(Y_i))}{g(Y_i)^2}(v(Y_i)-\mu_v),\\
\widehat{A}_5 &:= \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i)-g(Y_i)}{g(Y_i)}\right)^2(v(Y_i)-\mu_v).
\end{aligned}$$

Since the terms $\widehat{A}_4$ and $\widehat{A}_5$ have some factors of quadratic nature inside the sum (i.e. $(\hat{f}_h(Y_i)-f(Y_i))(\hat{g}_b(Y_i)-g(Y_i))$ and $(\hat{g}_b(Y_i)-g(Y_i))^2$) they are negligible with respect to other terms. Consequently,

$$\widehat{A} \simeq \widehat{A}_1 + \widehat{A}_2 - \widehat{A}_3.$$

Lemma 4.5.2. *The expectation and variance of $\widehat{A}$ can be approximated by*

$$\begin{aligned}
E\left(\widehat{A}\right) &\simeq E\left(\widehat{A}_1\right)+E\left(\widehat{A}_2\right)-E\left(\widehat{A}_3\right), &(4.13)\\
Var\left(\widehat{A}\right) &\simeq Var\left(\widehat{A}_1\right)+Var\left(\widehat{A}_2\right)+Var\left(\widehat{A}_3\right)\\
&+\ 2Cov\left(\widehat{A}_1,\widehat{A}_2\right)-2Cov\left(\widehat{A}_1,\widehat{A}_3\right)-2Cov\left(\widehat{A}_2,\widehat{A}_3\right). &(4.14)
\end{aligned}$$

Proof. We may write some expression for the ratio $\hat{f}_h(Y_i)/\hat{g}_b(Y_i)$ involved in the definition of $\widehat{A}$ in (4.11):

$$\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} = \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i)}{g(Y_i)}+1-\frac{\hat{g}_b(Y_i)}{g(Y_i)}\right),$$

which gives:

$$\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} = \frac{\hat{f}_h(Y_i)}{g(Y_i)}+\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{g(Y_i)-\hat{g}_b(Y_i)}{g(Y_i)}\right).$$

As a consequence,

$$\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}-\frac{f(Y_i)}{g(Y_i)} = \frac{\hat{f}_h(Y_i)-f(Y_i)}{g(Y_i)}+\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{g(Y_i)-\hat{g}_b(Y_i)}{g(Y_i)}\right).$$

Applying similar techniques once more to the term $\hat{f}_h(Y_i)/\hat{g}_b(Y_i)$ in the right-hand side of the previous expression gives:

$$\begin{aligned}
\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}-\frac{f(Y_i)}{g(Y_i)} &= \frac{\hat{f}_h(Y_i)-f(Y_i)}{g(Y_i)}+\frac{f(Y_i)}{g(Y_i)}\left(\frac{g(Y_i)-\hat{g}_b(Y_i)}{g(Y_i)}\right)\\
&+ \left(\frac{\hat{f}_h(Y_i)-f(Y_i)}{g(Y_i)}\right)\left(\frac{g(Y_i)-\hat{g}_b(Y_i)}{g(Y_i)}\right)+\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{g(Y_i)-\hat{g}_b(Y_i)}{g(Y_i)}\right)^2,
\end{aligned}$$

which can also be expressed as:

$$\begin{aligned}\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} - \frac{f(Y_i)}{g(Y_i)} &= \frac{\hat{f}_h(Y_i) - f(Y_i)}{g(Y_i)} - \frac{f(Y_i)}{g(Y_i)}\left(\frac{\hat{g}_b(Y_i) - g(Y_i)}{g(Y_i)}\right) \\ &- \frac{(\hat{f}_h(Y_i) - f(Y_i))(\hat{g}_b(Y_i) - g(Y_i))}{g(Y_i)^2} + \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i) - g(Y_i)}{g(Y_i)}\right)^2.\end{aligned}$$

The term $\widehat{A}$ defined in (4.11) can be splitted in different terms:

$$\widehat{A} = \widehat{A}_1 + \widehat{A}_2 - \widehat{A}_3 - \widehat{A}_4 + \widehat{A}_5,$$

where

$$\begin{aligned}\widehat{A}_1 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)}{g(Y_i)}(v(Y_i) - \mu_v), \qquad (A.17)\\ \widehat{A}_2 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i) - f(Y_i)}{g(Y_i)}(v(Y_i) - \mu_v), \\ \widehat{A}_3 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{f(Y_i)(\hat{g}_b(Y_i) - g(Y_i))}{g(Y_i)^2}(v(Y_i) - \mu_v), \\ \widehat{A}_4 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{(\hat{f}_h(Y_i) - f(Y_i))(\hat{g}_b(Y_i) - g(Y_i))}{g(Y_i)^2}(v(Y_i) - \mu_v), \\ \widehat{A}_5 &:= \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i) - g(Y_i)}{g(Y_i)}\right)^2 (v(Y_i) - \mu_v).\end{aligned}$$

Since the terms $\widehat{A}_4$ and $\widehat{A}_5$ have some factors of quadratic nature inside the sum (i.e. $(\hat{f}_h(Y_i) - f(Y_i))(\hat{g}_b(Y_i) - g(Y_i))$ and $(\hat{g}_b(Y_i) - g(Y_i))^2$) it is expected that one could prove negligibility of this terms.

Thus we will consider

$$\widehat{A} \simeq \widehat{A}_1 + \widehat{A}_2 - \widehat{A}_3.$$

Since we want to obtain the mean and variance of $\widehat{A}$, we proceed as shown in (4.13) and (4.14). □

The proof of Theorem 4.2.1 proceeds by analyzing the expectations and variances involved.

Lemma 4.5.3. *The expectation of $\widehat{A}$ is*

$$\begin{aligned}E(\widehat{A}) &\simeq D_1\frac{1}{Nb} + D_2 b^2 + D_3 b^4 - D_2\frac{b^2}{N} - D_3\frac{b^4}{N} \\ &+ D_4 h^2 + D_5 h^4 + O(b^6) + O(h^6), \qquad (4.15)\end{aligned}$$

where

$$\begin{aligned}
D_1 &:= -K(0)\int \gamma(y)dy,\\
D_2 &:= -\frac{\mu_2(K)}{2}\int \gamma(y)g''(y)dy,\\
D_3 &:= -\frac{\mu_4(K)}{24}\int \gamma(y)g^{(4)}(y)dy,\\
D_4 &:= \frac{\mu_2(K)}{2}\int v''(y)f(y)dy,\\
D_5 &:= \frac{\mu_4(K)}{24}\int v^{(4)}(y)f(y)dy,
\end{aligned}$$

with

$$\gamma(y) := \frac{f(y)}{g(y)}(v(y)-\mu_v).$$

Proof. We now consider the terms in the right-hand side of (4.13).

$$\begin{aligned}
E\left(\widehat{A}_1\right) &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{f(Y_i)}{g(Y_i)}(v(Y_i)-\mu_v)\right] = \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right]\\
&= E\left[\frac{f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right] = \int \frac{f(y)}{g(y)}(v(y)-\mu_v)g(y)dy\\
&= \int v(y)f(y)dy - \mu_v\int f(y)dy = \mu_v-\mu_v = 0.
\end{aligned}$$

Since the random variables

$$\eta_i := \Psi\left(X_1,\ldots,X_n,Y_i\right) := \frac{\hat{f}_h(Y_i)-f(Y_i)}{g(Y_i)}(v(Y_i)-\mu_v), \quad i=1,2,\ldots,N$$

are identically distributed (but not independent) then

$$\begin{aligned}
E\left(\widehat{A}_2\right) &= \frac{1}{N}\sum_{i=1}^{N} E\left(\eta_i\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(\eta_1\right) = E\left(\eta_1\right) = E\left[E\left(\eta_1|Y_1\right)\right]\\
&= E\left[\frac{E(\hat{f}_h(Y_1)|Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right]\\
&= E\left[\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right]\\
&= \int \frac{(K_h*f)(y)-f(y)}{g(y)}(v(y)-\mu_v)g(y)dy\\
&= \int \left((K_h*f)(y)-f(y)\right)(v(y)-\mu_v)dy\\
&= \int\left(\int K_h(y-z)f(z)dz\right)(v(y)-\mu_v)dy - \int f(y)(v(y)-\mu_v)dy
\end{aligned}$$

$$\begin{aligned}
&= \int f(z)\left(\int K(t)(v(z+ht)-\mu_v)dt\right)dz \\
&= \int f(z)\left(\int K(t)v(z+ht)dt\right)dz-\mu_v \\
&= \int v(z)f(z)dz+\frac{h^2}{2}\mu_2(K)\int v''(z)f(z)dz \\
&+ \frac{h^4}{24}\mu_4(K)\int v^{(4)}(z)f(z)dz+O(h^6)-\mu_v \\
&= \frac{h^2}{2}\mu_2(K)\int v''(z)f(z)dz+\frac{h^4}{24}\mu_4(K)\int v^{(4)}(z)f(z)dz+O(h^6). \qquad \text{(A.18)}
\end{aligned}$$

Finally,

$$\begin{aligned}
E(\widehat{A}_3) &= \frac{1}{N}\sum_{i=1}^{N}E\left[\frac{f(Y_i)(\hat{g}_b(Y_i)-g(Y_i))}{g(Y_i)^2}(v(Y_i)-\mu_v)\right] \\
&= E\left[\frac{f(Y_1)(\hat{g}_b(Y_1)-g(Y_1))}{g(Y_1)^2}(v(Y_1)-\mu_v)\right] \\
&= E\left[E\left(\frac{f(Y_1)(\hat{g}_b(Y_1)-g(Y_1))}{g(Y_1)^2}(v(Y_1)-\mu_v)|Y_1\right)\right] \\
&= E\left[\frac{f(Y_1)(E\left[\hat{g}_b(Y_1)|Y_1\right]-g(Y_1))}{g(Y_1)^2}(v(Y_1)-\mu_v)\right].
\end{aligned}$$

But since

$$\begin{aligned}
\hat{g}_b(Y_1) &= \frac{1}{N}\sum_{i=1}^{N}K_b(Y_1-Y_i)=\frac{1}{N}\left(K_b(0)+\sum_{i=2}^{N}K_b(Y_1-Y_i)\right) \\
&= \frac{K(0)}{Nb}+\frac{N-1}{N}\frac{1}{N-1}\sum_{i=2}^{N}K_b(Y_1-Y_i)=\frac{K(0)}{Nb}+\frac{N-1}{N}\hat{g}_b^{(-1)}(Y_1),
\end{aligned}$$

we get

$$E\left[\hat{g}_b(Y_1)|Y_1\right]=\frac{K(0)}{Nb}+\frac{N-1}{N}E\left[\hat{g}_b^{(-1)}(Y_1)|Y_1\right]=\frac{K(0)}{Nb}+\frac{N-1}{N}(K_b*g)(Y_1).$$

Using this expression above we have:

$$E(\widehat{A}_3)=E\left[\frac{f(Y_1)\left[\frac{K(0)}{Nb}+\frac{N-1}{N}(K_b*g)(Y_1)-g(Y_1)\right]}{g(Y_1)^2}(v(Y_1)-\mu_v)\right]$$

$$= \int\frac{f(y)\left(\frac{K(0)}{Nb}+\frac{N-1}{N}(K_b*g)(y)-g(y)\right)}{g(y)^2}(v(y)-\mu_v)g(y)dy$$

$$
\begin{aligned}
&= \frac{K(0)}{Nb}\int \frac{f(y)}{g(y)}(v(y)-\mu_v)dy + \int f(y)(v(y)-\mu_v)dy \\
&+ \frac{N-1}{N}\int \frac{f(y)\left[g(y)+\frac{\mu_2(K)}{2}b^2 g''(y)+\frac{\mu_4(K)}{4!}b^4 g^{(4)}(y)+O(b^6)\right]}{g(y)}(v(y)-\mu_v)dy \\
&= \frac{K(0)}{Nb}\int \gamma(y)dy + \frac{\mu_2(K)}{2}\frac{(N-1)b^2}{N}\int \frac{f(y)g''(y)}{g(y)}(v(y)-\mu_v)dy \\
&+ \frac{\mu_4(K)}{24}\frac{(N-1)b^4}{N}\int \frac{f(y)g^{(4)}(y)}{g(y)}(v(y)-\mu_v)dy + O(b^6). \qquad \text{(A.19)}
\end{aligned}
$$

From (A.18) and (A.19), since $E(\widehat{A}) \simeq E(\widehat{A}_2) - E(\widehat{A}_3)$, we get (4.15).

□

We now consider the terms in the right-hand side of (4.14):

Lemma 4.5.4. *The variance of $\widehat{A}_1$ is*

$$Var\left(\widehat{A}_1\right) = \frac{D_6}{N},$$

where

$$D_6 := \int \beta(y)dy,$$

with

$$\beta(y) := \frac{f(y)^2}{g(y)}(v(y)-\mu_v)^2.$$

Proof.

$$
\begin{aligned}
Var\left(\widehat{A}_1\right) &= \frac{1}{N^2}\sum_{i=1}^{N} Var\left[\frac{f(Y_i)}{g(Y_i)}(v(Y_i)-\mu_v)\right] = \frac{1}{N}Var\left[\frac{f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right] \\
&= \frac{1}{N}\left\{E\left[\frac{f(Y_1)^2}{g(Y_1)^2}(v(Y_1)-\mu_v)^2\right] - \left(E\left[\frac{f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right]\right)^2\right\} \\
&= \frac{1}{N}\int \beta(y)dy.
\end{aligned}
$$

□

Lemma 4.5.5. *The variance of $\widehat{A}_2$ is*

$$
\begin{aligned}
Var\left(\widehat{A}_2\right) &= D_7\frac{1}{n} + D_8\frac{1}{Nn} + D_9\frac{1}{Nnh} + D_{10}\frac{h^2}{n} + D_{11}\frac{h^4}{n} + D_{12}\frac{h^4}{N} + D_{13}\frac{h}{Nn} \\
&+ D_{14}\frac{h^2}{Nn} + D_{15}\frac{h^3}{Nn} + D_{16}\frac{h^6}{n} + D_{17}\frac{h^6}{N} + O\left(\frac{h^8}{n}\right) + O\left(\frac{h^4}{Nn}\right), \qquad (4.16)
\end{aligned}
$$

where

$$\begin{aligned}
D_7 &:= B(v^2) - \mu_v^2, \\
D_8 &:= -D_6 - B(v^2) + \mu_v^2, \\
D_9 &:= \mu_0(K^2) \int \theta(y) dy, \\
D_{10} &:= \mu_2(K) \left[B(v \cdot v'') - \mu_v B(v'') \right], \\
D_{11} &:= \frac{\mu_2(K)^2}{4} \left[B(v''^2) - B(v'')^2 \right] + \frac{\mu_4(K)}{12} \left[B(v \cdot v^{(4)}) - \mu_v B(v^{(4)}) \right], \\
D_{12} &:= \frac{\mu_2(K)^2}{4} \left[\int \frac{f''(y)^2}{g(y)} (v(y) - \mu_v)^2 dy - B(v'')^2 \right], \\
D_{13} &:= \frac{\mu_2(K^2)}{2} \int \frac{f''(y)}{g(y)} (v(y) - \mu_v)^2 dy, \\
D_{14} &:= -\mu_2(K) \left[\int \theta(y) f''(y) dy + B(v \cdot v'') - \mu_v B(v'') \right], \\
D_{15} &:= \frac{\mu_4(K^2)}{24} \int \frac{f^{(4)}(y)}{g(y)} (v(y) - \mu_v)^2 dy, \\
D_{16} &:= \frac{\mu_2(K)\mu_4(K)}{24} \left[B(v'' \cdot v^{(4)}) - B(v'')B(v^{(4)}) \right] \\
&+ \frac{\mu_6(K)}{360} \left[B(v \cdot v^{(6)}) - \mu_v B(v^{(6)}) \right], \\
D_{17} &:= \frac{\mu_2(K)\mu_4(K)}{24} \left[\int \frac{f''(y) f^{(4)}(y)}{g(y)} (v(y) - \mu_v)^2 dy - B(v'')B(v^{(4)}) \right],
\end{aligned}$$

where the operator B is defined by

$$B(\phi) := \int \phi(x) f(x) dx$$

and

$$\theta(y) := \frac{f(y)}{g(y)} (v(y) - \mu_v)^2.$$

Proof. In order to compute the variance of $\widehat{A}_2$, let us rewrite the terms η_i as follows:

$$\eta_i = \frac{1}{n} \sum_{j=1}^{n} \frac{K_h(Y_i - X_j) - f(Y_i)}{g(Y_i)} (v(Y_i) - \mu_v) = \frac{1}{n} \sum_{j=1}^{n} \eta_{ij}$$

with

$$\eta_{ij} := \frac{K_h(Y_i - X_j) - f(Y_i)}{g(Y_i)} (v(Y_i) - \mu_v), \quad i = 1, \ldots, N; j = 1, \ldots, n. \tag{A.20}$$

Using (A.20), the variance of $\widehat{A}_2$ can be written as

$$Var(\widehat{A}_2) = \frac{1}{Nn} \sum_{i=1}^{N} \sum_{j=1}^{n} \eta_{ij} = \frac{1}{N^2 n^2} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{N} \sum_{l=1}^{n} Cov(\eta_{ij}, \eta_{kl}). \tag{A.21}$$

Collecting all the equal terms in (A.21) gives:

$$Var(\widehat{A}_2) = \frac{n-1}{Nn} Cov(\eta_{11}, \eta_{12}) + \frac{N-1}{Nn} Cov(\eta_{11}, \eta_{21}) + \frac{1}{Nn} Var(\eta_{11}). \qquad \text{(A.22)}$$

We now work these covariance terms out:

$$Cov(\eta_{11}, \eta_{12}) = Cov(E\left(\eta_{11}|Y_1\right), E\left(\eta_{12}|Y_1\right)) + E\left[Cov(\eta_{11}, \eta_{12}|Y_1)\right].$$

But $Cov(\eta_{11}, \eta_{12}|Y_1) = 0$, since

$$\eta_{11} = \frac{K_h(Y_1 - X_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)$$

and

$$\eta_{12} = \frac{K_h(Y_1 - X_2) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)$$

are conditionally independent on Y_1 (because X_1 and X_2 are independent).

On the other hand,

$$\begin{aligned} E\left(\eta_{11}|Y_1\right) &= E\left(\eta_{12}|Y_1\right) = \frac{E\left[K_h(Y_1 - X_1)|Y_1\right] - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v) \\ &= \frac{E\left[K_h(X_1 - Y_1)|Y_1\right] - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v) \\ &= \frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v). \qquad \text{(A.23)} \end{aligned}$$

Now

$$\begin{aligned} Cov(E\left(\eta_{11}|Y_1\right), E\left(\eta_{12}|Y_1\right)) &= Var(E\left(\eta_{11}|Y_1\right)) \\ &= E\left[\left(\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right)^2\right] \\ &- \left[E\left(\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right)\right]^2, \end{aligned}$$

with

$$\begin{aligned} & E\left[\left(\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right)^2\right] \\ &= \int\left(\frac{(K_h * f)(y) - f(y)}{g(y)}(v(y) - \mu_v)\right)^2 g(y)dy \\ &= \int \frac{((K_h * f)(y) - f(y))^2}{g(y)}(v(y) - \mu_v)^2 dy \\ &= \int \frac{\left(\frac{\mu_2(K)}{2}h^2 f''(y) + \frac{\mu_4(K)}{4!}h^4 f^{(4)}(y) + O(h^6)\right)^2}{g(y)}(v(y) - \mu_v)^2 dy \end{aligned}$$

$$\begin{aligned} &= \frac{\mu_2(K)^2}{4}h^4\int\frac{f''(y)^2}{g(y)}(v(y)-\mu_v)^2dy \\ &+ \frac{\mu_2(K)\mu_4(K)}{24}h^6\int\frac{f''(y)f^{(4)}(y)}{g(y)}(v(y)-\mu_v)^2dy+O(h^8)\end{aligned}$$

and

$$\begin{aligned} &\left[E\left(\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right)\right]^2 \\ &= \left[\int\frac{(K_h*f)(y)-f(y)}{g(y)}(v(y)-\mu_v)g(y)dy\right]^2 \\ &= \left[\int((K_h*f)(y)-f(y))\,(v(y)-\mu_v)dy\right]^2 \\ &= \left[\int(K_h*f)(y)(v(y)-\mu_v)dy-\int f(y)(v(y)-\mu_v)dy\right]^2 \\ &= \left[\int\left(\int K_h(y-z)f(z)dz\right)(v(y)-\mu_v)dy\right]^2 \\ &= \left[\int f(z)\left(\int K(t)(v(z+ht)-\mu_v)dt\right)dz\right]^2 \\ &= \left[\int f(z)\left(\int K(t)v(z+ht)dt\right)dz-\mu_v\right]^2 \\ &= \left[\frac{h^2}{2}\mu_2(K)\int v''(z)f(z)dz+\frac{h^4}{24}\mu_4(K)\int v^{(4)}(z)f(z)dz+O(h^6)\right]^2 \\ &= \frac{\mu_2(K)^2}{4}h^4\left(\int v''(z)f(z)dz\right)^2 \\ &+ \frac{\mu_2(K)\mu_4(K)}{24}h^6\int v''(z)f(z)dz\int v^{(4)}(z)f(z)dz+O(h^8)\end{aligned}$$

Consequently:

$$\begin{aligned} Cov(\eta_{11},\eta_{12}) &= Cov(E(\eta_{11}|Y_1),E(\eta_{12}|Y_1))=Var(E(\eta_{11}|Y_1)) \\ &= \frac{\mu_2(K)^2}{4}h^4\left(\int\frac{f''(y)^2}{g(y)}(v(y)-\mu_v)^2dy-\left(\int v''(z)f(z)dz\right)^2\right) \\ &+ \frac{\mu_2(K)\mu_4(K)}{24}h^6\left(\int\frac{f''(y)f^{(4)}(y)}{g(y)}(v(y)-\mu_v)^2dy\right. \\ &- \left.\int v''(z)f(z)dz\int v^{(4)}(z)f(z)dz\right)+O(h^8). \qquad (A.24)\end{aligned}$$

Now we deal with the term $Cov(\eta_{11},\eta_{21})$ in (A.22):

$$\begin{aligned} Cov(\eta_{11},\eta_{21}) &= Cov(E(\eta_{11}|X_1),E(\eta_{21}|X_1))+E[Cov(\eta_{11},\eta_{21}|X_1)] \\ &= Cov(E(\eta_{11}|X_1),E(\eta_{21}|X_1)),\end{aligned}$$

since $Cov(\eta_{11}, \eta_{21}|X_1) = 0$ because

$$\eta_{11} = \frac{K_h(Y_1 - X_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)$$

and

$$\eta_{21} = \frac{K_h(Y_2 - X_1) - f(Y_2)}{g(Y_2)}(v(Y_2) - \mu_v)$$

are conditionally independent given X_1 (since Y_1 and Y_2 are independent). But

$$\begin{aligned} E\left(\eta_{11}|X_1\right) &= E\left(\eta_{21}|X_1\right) = \int \frac{K_h(y - X_1) - f(y)}{g(y)}(v(y) - \mu_v)g(y)dy \\ &= \int K_h(y - X_1)(v(y) - \mu_v)dy - \int f(y)(v(y) - \mu_v)dy \\ &= \int K(t)(v(X_1 + ht) - \mu_v)dt. \end{aligned}$$

So,

$$\begin{aligned} Cov\left(E\left(\eta_{11}|X_1\right), E\left(\eta_{21}|X_1\right)\right) &= Var\left(\int K(t)(v(X_1 + ht) - \mu_v)dt\right) \\ &= Var\left(\int K(t)v(X_1 + ht)dt\right). \end{aligned}$$

On the one hand,

$$\begin{aligned} & E\left[\int K(t)v(X_1 + ht)dt\right] = \int \left(\int K(t)v(x + ht)dt\right) f(x)dx \\ = & \mu_v + \frac{h^2}{2}\mu_2(K)\int v''(x)f(x)dx + \frac{h^4}{24}\mu_4(K)\int v^{(4)}(x)f(x)dx \\ + & \frac{h^6}{720}\mu_6(K)\int v^{(6)}(x)f(x)dx + O(h^8) \end{aligned}$$

and consequently

$$\begin{aligned} & \left(E\left[\int K(t)v(X_1 + ht)dt\right]\right)^2 = \mu_v^2 + h^2\mu_2(K)\mu_v\int v''(x)f(x)dx \\ + & \frac{h^4}{4}\mu_2(K)^2\left(\int v''(x)f(x)dx\right)^2 + \frac{h^4}{12}\mu_4(K)\mu_v\int v^{(4)}(x)f(x)dx \\ + & \frac{h^6}{360}\mu_6(K)\mu_v\int v^{(6)}(x)f(x)dx \\ + & \frac{h^6}{24}\mu_2(K)\mu_4(K)\int v''(x)f(x)dx\int v^{(4)}(x)f(x)dx + O(h^8). \qquad \text{(A.25)} \end{aligned}$$

On the other hand,

$$
\begin{aligned}
E\left[\left(\int K(t)v(X_1+ht)dt\right)^2\right] &= E\left[\int\int K(s)v(X_1+hs)K(t)v(X_1+ht)dsdt\right] \\
= & \int\left(\int\int K(s)v(X_1+hs)K(t)v(X_1+ht)dsdt\right)f(x)dx \\
= & \int v(x)^2 f(x)dx + h^2\mu_2(K)\int v(x)v''(x)f(x)dx \\
+ & \frac{h^4}{12}\mu_4(K)\int v(x)v^{(4)}(x)f(x)dx + \frac{h^4}{4}\mu_2(K)^2\int v''(x)^2 f(x)dx \\
+ & \frac{h^6}{360}\mu_6(K)\int v(x)v^{(6)}(x)f(x)dx \\
+ & \frac{h^6}{24}\mu_2(K)\mu_4(K)\int v''(x)v^{(4)}(x)f(x)dx + O(h^8). \qquad \text{(A.26)}
\end{aligned}
$$

From (A.25) and (A.26), we obtain

$$
\begin{aligned}
Cov(\eta_{11},\eta_{21}) = Var&\left(\int K(t)v(X_1+ht)dt\right) = B(v^2)-\mu_v^2 \\
+ & \ h^2\mu_2(K)\left[B(v\cdot v'')-\mu_v B(v'')\right] + \frac{h^4}{4}\mu_2(K)^2\left[B(v''^2)-B(v'')^2\right] \\
+ & \ \frac{h^4}{12}\mu_4(K)\left[B(v\cdot v^{(4)})-\mu_v B(v^{(4)})\right] \\
+ & \ \frac{h^6}{24}\mu_2(K)\mu_4(K)\left[B(v''\cdot v^{(4)})-B(v'')B(v^{(4)})\right] \\
+ & \ \frac{h^6}{360}\mu_6(K)\left[B(v\cdot v^{(6)})-\mu_v B(v^{(6)})\right] + O(h^8), \qquad \text{(A.27)}
\end{aligned}
$$

where the operator B is defined by

$$B(\phi) := \int \phi(x)f(x)dx$$

and in particular $B(v)=\mu_v$.

We now examine the term $Var(\eta_{11})$ in (A.22):

$$
\begin{aligned}
Var(\eta_{11}) &= Var\left[E\left(\eta_{11}|X_1\right)\right] + E\left[Var(\eta_{11}|X_1)\right] \\
&= Var\left(\int K(t)v(X_1+ht)dt\right) + E\left[Var(\eta_{11}|X_1)\right],
\end{aligned}
$$

since $E\left(\eta_{11}|X_1\right) = \int K(t)(v(X_1+ht)-\mu_v)dt$.

Now

$$
\begin{aligned}
Var(\eta_{11}|X_1) &= E\left(\eta_{11}^2|X_1\right) - \left[E\left(\eta_{11}|X_1\right)\right]^2 \\
&= E\left(\eta_{11}^2|X_1\right) - \left(\int K(t)v(X_1+ht)dt-\mu_v\right)^2
\end{aligned}
$$

and since

$$
\begin{aligned}
E\left[\left[E\left(\eta_{11}|X_1\right)\right]^2\right] &= E\left[\left(\int K(t)v(X_1+ht)dt-\mu_v\right)^2\right]\\
&= E\left[\left(\int K(t)v(X_1+ht)dt-E\left(\int K(t)v(X_1+ht)dt\right)\right.\right.\\
&+ \left.\left.E\left(\int K(t)v(X_1+ht)dt\right)-\mu_v\right)^2\right]\\
&= Var\left(\int K(t)v(X_1+ht)dt\right)+\left(E\left(\int K(t)v(X_1+ht)dt\right)-\mu_v\right)^2
\end{aligned}
$$

we obtain

$$
\begin{aligned}
Var(\eta_{11}) &= Var\left(\int K(t)v(X_1+ht)dt\right)+E\left[Var(\eta_{11}|X_1)\right]\\
&= Var\left(\int K(t)v(X_1+ht)dt\right)+E\left[E\left(\eta_{11}^2|X_1\right)\right]\\
&- Var\left(\int K(t)v(X_1+ht)dt\right)-\left(E\left(\int K(t)v(X_1+ht)dt\right)-\mu_v\right)^2\\
&= E\left(\eta_{11}^2\right)-\left(E\left(\int K(t)v(X_1+ht)dt\right)-\mu_v\right)^2\\
&= E\left[E\left(\eta_{11}^2|Y_1\right)\right]-\frac{h^4}{4}\mu_2(K)^2B(v'')^2+O(h^6)\\
&= E\left[E\left(\left(\frac{K_h(Y_1-X_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right)^2|Y_1\right)\right]\\
&- \frac{h^4}{4}\mu_2(K)^2B(v'')^2+O(h^6)\\
&= E\left[\frac{((K_h)^2*f)(Y_1)-2f(Y_1)(K_h*f)(Y_1)+f(Y_1)^2}{g(Y_1)^2}(v(Y_1)-\mu_v)^2\right]\\
&- \frac{h^4}{4}\mu_2(K)^2B(v'')^2+O(h^6)\\
&= \int\frac{((K_h)^2*f)(y)-2f(y)(K_h*f)(y)+f(y)^2}{g(y)^2}(v(y)-\mu_v)^2g(y)dy\\
&- \frac{h^4}{4}\mu_2(K)^2B(v'')^2+O(h^6)\\
&= \int\frac{((K_h)^2*f)(y)-2f(y)(K_h*f)(y)+f(y)^2}{g(y)}(v(y)-\mu_v)^2dy\\
&- \frac{h^4}{4}\mu_2(K)^2B(v'')^2+O(h^6). \qquad \text{(A.28)}
\end{aligned}
$$

But,

$$\begin{aligned}((K_h)^2 * f)(y) &= \int (K_h)^2(y-z)f(z)dz \\ &= \int \left[K_h(y-z)\right]^2 f(z)dz = \frac{1}{h}\int K(t)^2 f(y-ht)dt \\ &= \frac{1}{h}\left[\mu_0(K^2)f(y) + \frac{\mu_2(K^2)}{2}h^2 f''(y) + \frac{\mu_4(K^2)}{4!}h^4 f^{(4)}(y) + O(h^6)\right] \\ &= \frac{\mu_0(K^2)}{h}f(y) + \frac{\mu_2(K^2)}{2}hf''(y) + \frac{\mu_4(K^2)}{24}h^3 f^{(4)}(y) + O(h^5).\end{aligned}$$

Now (A.28) becomes:

$$\begin{aligned}Var(\eta_{11}) &= \int g(y)^{-1}\left[\frac{\mu_0(K^2)}{h}f(y) + \frac{\mu_2(K^2)}{2}hf''(y) + \frac{\mu_4(K^2)}{24}h^3 f^{(4)}(y) + O(h^5)\right. \\ &- 2f(y)\left(f(y) + \frac{\mu_2(K)}{2}h^2 f''(y) + \frac{\mu_4(K)}{24}h^4 f^{(4)}(y) + O(h^6)\right) \\ &+ \left. f(y)^2\right](v(y)-\mu_v)^2 dy - \frac{h^4}{4}\mu_2(K)^2 B(v'')^2 + O(h^6) \\ &= \frac{\mu_0(K^2)}{h}\int \theta(y)dy + \int \beta(y)dy \\ &+ \frac{\mu_2(K^2)}{2}h\int \frac{f''(y)}{g(y)}(v(y)-\mu_v)^2 dy - \mu_2(K)h^2\int \theta(y)f''(y)dy \\ &+ \frac{\mu_4(K^2)}{24}h^3\int \frac{f^{(4)}(y)}{g(y)}(v(y)-\mu_v)^2 dy - \frac{\mu_4(K)}{12}h^4\int \theta(y)f^{(4)}dy \\ &- \frac{h^4}{4}\mu_2(K)^2 B(v'')^2 + O(h^5). \qquad (A.29)\end{aligned}$$

Now, using (A.24), (A.27) and (A.29) in (A.22) gives:

$$\begin{aligned}Var(\widehat{A}_2) &= \frac{n-1}{Nn}\left[\frac{\mu_2(K)^2}{4}h^4\left(\int \frac{f''(y)^2}{g(y)}(v(y)-\mu_v)^2 dy - \left(\int v''(z)f(z)dz\right)^2\right)\right. \\ &+ \frac{\mu_2(K)\mu_4(K)}{24}h^6\left(\int \frac{f''(y)f^{(4)}(y)}{g(y)}(v(y)-\mu_v)^2 dy\right. \\ &- \left.\left.\int v''(z)f(z)dz\int v^{(4)}(z)f(z)dz\right)\right] \\ &+ \frac{N-1}{Nn}\left[B(v^2) - \mu_v^2 + h^2\mu_2(K)\left[B(v\cdot v'') - \mu_v B(v'')\right]\right. \\ &+ \frac{h^4}{4}\mu_2(K)^2\left[B(v''^2) - B(v'')^2\right] \\ &+ \left.\frac{h^4}{12}\mu_4(K)\left[B(v\cdot v^{(4)}) - \mu_v B(v^{(4)})\right] + O(h^6)\right]\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{Nn}\left[\frac{\mu_0(K^2)}{h}\int \theta(y)dy - \int \beta(y)dy\right. \\
&+ \frac{\mu_2(K^2)}{2}h\int \frac{f''(y)}{g(y)}(v(y)-\mu_v)^2dy - \mu_2(K)h^2\int \theta(y)f''(y)dy \\
&+ \frac{\mu_4(K^2)}{24}h^3\int \frac{f^{(4)}(y)}{g(y)}(v(y)-\mu_v)^2dy - \frac{\mu_4(K)}{12}h^4\int \theta(y)f^{(4)}dy \\
&- \left.\frac{h^4}{4}\mu_2(K)^2B(v'')^2 + O(h^5)\right]. \qquad \text{(A.30)}
\end{aligned}
$$

Shortening the expression (A.30), we obtain (4.16). □

Lemma 4.5.6. *The variance of* $\widehat{A}_3$ *is*

$$
\begin{aligned}
Var\left(\widehat{A}_3\right) &= D_{18}\frac{1}{N} + D_{19}\frac{b^2}{N} + D_{20}\frac{b^4}{N} + D_{21}\frac{1}{N^2b} + D_{22}\frac{1}{N^2} + D_{23}\frac{b}{N^2} \\
&+ D_{24}\frac{1}{N^3b^2} + D_{25}\frac{1}{N^3b} + D_{26}\frac{1}{N^3} + O\left(\frac{b^6}{N}\right) + O\left(\frac{b^2}{N^2}\right), \qquad (4.17)
\end{aligned}
$$

where

$$
\begin{aligned}
D_{18} &:= \int \beta(y)dy, \\
D_{19} &:= \mu_2(K)\left[\int \alpha(y)g''(y)dy + \int \gamma''(y)f(y)(v(y)-\mu_v)dy\right], \\
D_{20} &:= \frac{\mu_4(K)}{12}\left[\int \alpha(y)g^{(4)}(y)dy + \int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy\right] \\
&+ \frac{\mu_2(K)^2}{4}\left[\int \delta(y)g''(y)^2dy - 4\left(\int \gamma(y)g''(y)dy\right)^2\right. \\
&+ \left.\int \gamma''(y)^2g(y)dz + 2\int \gamma(y)\gamma''(y)g''(y)dy\right], \\
D_{21} &:= 2\left[\mu_0(K^2) + K(0)\right]\int \alpha(y)dy, \\
D_{22} &:= -8\int \beta(y)dy = -8D_{18}, \\
D_{23} &:= \mu_2(K)K(0)\left[\int \delta(y)g''(y)dy + \int \gamma(y)\gamma''(y)dy\right. \\
&- \left.\left(\int \gamma(y)g''(y)dy + \int \gamma''(y)g(y)dy\right)\left(\int \gamma(y)dy\right)\right] \\
&+ \frac{\mu_2(K^2)}{2}\left[\int \delta(y)g''(y)dy + \int \gamma(y)\gamma''(y)dy\right],
\end{aligned}
$$

$$\begin{aligned}
D_{24} &:= K(0)^2\left[\int \delta(y)dy - \left(\int \gamma(y)dy\right)^2\right],\\
D_{25} &:= -2\left[\mu_0(K^2) + 2K(0)\right]\int \alpha(y)dy,\\
D_{26} &:= 8\int \beta(y)dy = 8D_{18},
\end{aligned}$$

with

$$\begin{aligned}
\alpha(y) &:= \frac{f(y)^2}{g(y)^2}(v(y)-\mu_v)^2,\\
\delta(y) &:= \frac{f(y)^2}{g(y)^3}(v(y)-\mu_v)^2.
\end{aligned}$$

Proof. To deal with the variance of $\widehat{A}_3$ we first consider

$$\begin{aligned}
\tau_i &:= \frac{f(Y_i)(\hat{g}_b(Y_i)-g(Y_i))}{g(Y_i)^2}(v(Y_i)-\mu_v)\\
&= \frac{1}{N}\sum_{j=1}^{N}\frac{f(Y_i)(K_b(Y_i-Y_j)-g(Y_i))}{g(Y_i)^2}(v(Y_i)-\mu_v) = \frac{1}{N}\sum_{j=1}^{N}\tau_{ij}.
\end{aligned}$$

As a consequence, $\widehat{A}_3$ can be written as

$$\widehat{A}_3 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij}.$$

To compute the variance of $\widehat{A}_3$ we consider

$$\begin{aligned}
Var(\widehat{A}_3) &= Cov(\widehat{A}_3,\widehat{A}_3) = Cov\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij}, \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\tau_{kl}\right)\\
&= \frac{1}{N^4}\sum_{i,j,k,l=1}^{N} Cov(\tau_{ij},\tau_{kl}).
\end{aligned}$$

Thus, the variance of $\widehat{A}_3$ can be written as:

$$\begin{aligned}
Var(\widehat{A}_3) = \frac{(N-1)(N-2)}{N^3}&Cov(\tau_{12},\tau_{13}) + \frac{2(N-1)(N-2)}{N^3}Cov(\tau_{12},\tau_{31})\\
+ \quad \frac{(N-1)(N-2)}{N^3}&Cov(\tau_{12},\tau_{32}) + \frac{N-1}{N^3}Var(\tau_{12}) + \frac{N-1}{N^3}Cov(\tau_{12},\tau_{21})\\
+ \quad \frac{2(N-1)}{N^3}&Cov(\tau_{12},\tau_{11}) + \frac{2(N-1)}{N^3}Cov(\tau_{12},\tau_{22}) + \frac{1}{N^3}Var(\tau_{11}). \qquad (A.31)
\end{aligned}$$

Lemmas A.1.2, A.1.3, A.1.4, A.1.5, A.1.6, A.1.7, A.1.8 and A.1.9 can be used in (A.31) to conclude with the asymptotic expression (4.17). □

Let's deal with every one of these terms; but first, in order to save space, let us write τ_{ij} in a more compact form:

$$\tau_{ij} = \varphi(Y_i)(K_b(Y_i - Y_j) - g(Y_i)), \quad i,j = 1, \ldots, N$$

with

$$\varphi(y) := \frac{f(y)}{g(y)^2}(v(y) - \mu_v).$$

Lemma A.1.2. *The covariance of* τ_{12} *and* τ_{13} *is*

$$\begin{aligned} Cov(\tau_{12}, \tau_{13}) &= \frac{\mu_2(K)^2}{4} b^4 \left[\int \delta(y) g''(y)^2 dy - \left(\int \gamma(y) g''(y) dy \right)^2 \right] \\ &+ \frac{\mu_2(K)\mu_4(K)}{24} b^6 \left[\int \delta(y) g''(y) g^{(4)}(y) dy \right. \\ &- \left. \left(\int \gamma(y) g''(y) dy \right) \left(\int \gamma(y) g^{(4)}(y) dy \right) \right] + O(b^8). \qquad \text{(A.32)} \end{aligned}$$

Proof.

$$\begin{aligned} Cov(\tau_{12}, \tau_{13}) &= E\left[Cov(\tau_{12}, \tau_{13}|Y_1)\right] + Cov\left(E\left(\tau_{12}|Y_1\right), E\left(\tau_{13}|Y_1\right)\right) \\ &= Var\left(E\left(\tau_{12}|Y_1\right)\right) = Var\left(\varphi(Y_1)\left[(K_b * g)(Y_1) - g(Y_1)\right]\right) \end{aligned}$$

because

$$\tau_{12} = \varphi(Y_1)(K_b(Y_1 - Y_2) - g(Y_1))$$

and

$$\tau_{13} = \varphi(Y_1)(K_b(Y_1 - Y_3) - g(Y_1))$$

are conditionally independent given Y_1 (so $Cov(\tau_{12}, \tau_{13}|Y_1) = 0$) and

$$\begin{aligned} E\left(\tau_{12}|Y_1\right) &= E\left(\tau_{13}|Y_1\right) = \varphi(Y_1)\left[E\left[K_b(Y_1 - Y_2)|Y_1\right] - g(Y_1)\right] \\ &= \varphi(Y_1)\left[(K_b * g)(Y_1) - g(Y_1)\right]. \qquad \text{(A.33)} \end{aligned}$$

Now

$$\begin{aligned} Cov(\tau_{12}, \tau_{13}) &= E\left(\varphi(Y_1)^2 \left[(K_b * g)(Y_1) - g(Y_1)\right]^2\right) \\ &- \left[E\left(\varphi(Y_1)\left[(K_b * g)(Y_1) - g(Y_1)\right]\right)\right]^2, \end{aligned}$$

where

$$\begin{aligned} E\left(\varphi(Y_1)^2 \left[(K_b * g)(Y_1) - g(Y_1)\right]^2\right) &= \int \varphi(y)^2 \left((K_b * g)(y) - g(y)\right)^2 g(y) dy \\ = \int \varphi(y)^2 &\left[\frac{\mu_2(K)}{2} b^2 g''(y) + \frac{\mu_4(K)}{24} b^4 g^{(4)}(y) + O(b^6) \right]^2 g(y) dy \\ = \frac{\mu_2(K)^2}{4} b^4 &\int \delta(y) g''(y)^2 dy + \frac{\mu_2(K)\mu_4(K)}{24} b^6 \int \delta(y) g''(y) g^{(4)}(y) dy + O(b^8) \end{aligned}$$

and

$$\begin{aligned}
&E\left(\varphi(Y_1)\left[(K_b * g)(Y_1) - g(Y_1)\right]\right)\\
&\quad = \int \varphi(y)\left[\frac{\mu_2(K)}{2}b^2 g''(y) + \frac{\mu_4(K)}{24}b^4 g^{(4)}(y) + O(b^6)\right] g(y)dy\\
&\quad = \frac{\mu_2(K)}{2}b^2 \int \gamma(y)g''(y)dy + \frac{\mu_4(K)}{24}b^4 \int \gamma(y)g^{(4)}(y)dy + O(b^6). \qquad \text{(A.34)}
\end{aligned}$$

Using these two expressions we obtain (A.32). □

Lemma A.1.3. *The covariance of* τ_{12} *and* τ_{31} *is*

$$\begin{aligned}
Cov(\tau_{12}, \tau_{31}) &= \frac{\mu_2(K)}{2}b^2 \int \alpha(z)g''(z)dz + \frac{\mu_4(K)}{24}b^4 \int \alpha(z)g^{(4)}(z)dz\\
&+ \frac{\mu_2(K)^2}{4}b^4 \int \gamma(z)\gamma''(z)g''(z)dz - \frac{\mu_2(K)^2}{4}b^4 \left(\int \gamma(y)g''(y)dy\right)^2 + O(b^6)\\
&= \frac{\mu_2(K)}{2}b^2 \int \alpha(y)g''(y)dy + b^4\left[\frac{\mu_4(K)}{24}\int \alpha(y)g^{(4)}(y)dy\right.\\
&+ \left.\frac{\mu_2(K)^2}{4}\left(\int \gamma(y)\gamma''(y)g''(y)dy - \left(\int \gamma(y)g''(y)dy\right)^2\right)\right] + O(b^6). \qquad \text{(A.35)}
\end{aligned}$$

Proof.

$$\begin{aligned}
Cov(\tau_{12}, \tau_{31}) &= E\left[Cov(\tau_{12}, \tau_{31}|Y_1)\right] + Cov\left(E\left(\tau_{12}|Y_1\right), E\left(\tau_{31}|Y_1\right)\right)\\
&= Cov\left(E\left(\tau_{12}|Y_1\right), E\left(\tau_{31}|Y_1\right)\right).
\end{aligned}$$

We know that

$$E\left(\tau_{12}|Y_1\right) = \varphi(Y_1)\left[(K_b * g)(Y_1) - g(Y_1)\right]$$

and we can also deal with

$$\begin{aligned}
E\left(\tau_{31}|Y_1\right) &= E\left[\varphi(Y_3)(K_b(Y_3 - Y_1) - g(Y_3))\right]\\
&= \int \varphi(y)\left[K_b(y - Y_1) - g(y)\right] g(y)dy. \qquad \text{(A.36)}
\end{aligned}$$

Now

$$\begin{aligned}
Cov\left(E\left(\tau_{12}|Y_1\right), E\left(\tau_{31}|Y_1\right)\right) &= E\left[E\left(\tau_{12}|Y_1\right)E\left(\tau_{31}|Y_1\right)\right]\\
&- E\left[E\left(\tau_{12}|Y_1\right)\right]E\left[E\left(\tau_{31}|Y_1\right)\right]\\
&= E\left[E\left(\tau_{12}|Y_1\right)E\left(\tau_{31}|Y_1\right)\right] - E(\tau_{12})E(\tau_{31})\\
&= E\left[E\left(\tau_{12}|Y_1\right)E\left(\tau_{31}|Y_1\right)\right] - \left[E(\tau_{12})\right]^2.
\end{aligned}$$

But since $E\left(\tau_{12}|Y_1\right) = \varphi(Y_1)\left[(K_b * g)(Y_1) - g(Y_1)\right]$, equation (A.34) lead to:

$$E(\tau_{12}) = \frac{\mu_2(K)}{2}b^2\int\gamma(y)g''(y)dy + \frac{\mu_4(K)}{24}b^4\int\gamma(y)g^{(4)}(y)dy + O(b^6). \tag{A.37}$$

On the other hand, using the expression (A.36), we have:

$$\begin{aligned}
&E\left[E\left(\tau_{12}|Y_1\right)E\left(\tau_{31}|Y_1\right)\right]\\
&= E\left[\varphi(Y_1)\left[(K_b * g)(Y_1) - g(Y_1)\right]\int\varphi(y)\left[K_b(y-Y_1) - g(y)\right]g(y)dy\right]\\
&= \int\varphi(z)\left[(K_b * g)(z) - g(z)\right]\left(\int\varphi(y)\left[K_b(y-z) - g(y)\right]g(y)dy\right)g(z)dz\\
&= \int\varphi(z)\left[(K_b * g)(z) - g(z)\right]\left[\int\varphi(y)K_b(y-z)g(y)dy - \int\varphi(y)g(y)^2dy\right]g(z)dz\\
&= \int\varphi(z)\left[(K_b * g)(z) - g(z)\right]\left(\int\varphi(y)K_b(y-z)g(y)dy\right)g(z)dz\\
&- \left(\int\varphi(y)g(y)^2dy\right)\left(\int\varphi(z)\left[(K_b * g)(z) - g(z)\right]g(z)dz\right).
\end{aligned}$$

But

$$\int\varphi(y)g(y)^2dy = \int\frac{f(y)}{g(y)^2}(v(y)-\mu_v)g(y)^2dy = \int f(y)(v(y)-\mu_v)dy = 0,$$

and

$$\begin{aligned}
\int\varphi(y)K_b(y-z)g(y)dy &= \int\varphi(z+bt)K(t)g(z+bt)dt\\
&= \gamma(z) + \frac{\mu_2(K)}{2}b^2\gamma''(z) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z) + O(b^6),
\end{aligned} \tag{A.38}$$

and since

$$(K_b * g)(z) - g(z) = \frac{\mu_2(K)}{2}b^2g''(z) + \frac{\mu_4(K)}{24}b^4g^{(4)}(z) + O(b^6),$$

we conclude that

$$\begin{aligned}
E\left[E\left(\tau_{12}|Y_1\right)E\left(\tau_{31}|Y_1\right)\right] &= \int\varphi(z)\left[\frac{\mu_2(K)}{2}b^2g''(z) + \frac{\mu_4(K)}{24}b^4g^{(4)}(z) + O(b^6)\right]\\
&\cdot\left[\gamma(z) + \frac{\mu_2(K)}{2}b^2\gamma''(z) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z) + O(b^6)\right]g(z)dz\\
&= \frac{\mu_2(K)}{2}b^2\int\varphi(z)\gamma(z)g''(z)g(z)dz + \frac{\mu_4(K)}{24}b^4\int\varphi(z)\gamma(z)g^{(4)}(z)g(z)dz\\
&+ \frac{\mu_2(K)^2}{4}b^4\int\varphi(z)\gamma''(z)g''(z)g(z)dz + O(b^6)\\
&= \frac{\mu_2(K)}{2}b^2\int\alpha(z)g''(z)dz + \frac{\mu_4(K)}{24}b^4\int\alpha(z)g^{(4)}(z)dz\\
&+ \frac{\mu_2(K)^2}{4}b^4\int\gamma(z)\gamma''(z)g''(z)dz + O(b^6).
\end{aligned} \tag{A.39}$$

Using (A.37) and (A.39) we get (A.35). □

Lemma A.1.4. *The covariance of* τ_{12} *and* τ_{32} *is*

$$\begin{aligned} Cov(\tau_{12},\tau_{32}) &= \int \beta(y)dy + \mu_2(K)b^2 \int \gamma''(y)f(y)(v(y)-\mu_v)dy \\ &+ \frac{\mu_2(K)^2}{4}b^4 \int \gamma''(y)^2 g(y)dy + \frac{\mu_4(K)}{12}b^4 \int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy \\ &- \frac{\mu_2(K)^2}{4}b^4 \left(\int \gamma(y)g''(y)dy\right)^2 + O(b^6). \end{aligned} \tag{A.40}$$

Proof.

$$\begin{aligned} Cov(\tau_{12},\tau_{32}) &= E\left[Cov(\tau_{12},\tau_{32}|Y_2)\right] + Cov\left(E\left(\tau_{12}|Y_2\right), E\left(\tau_{32}|Y_2\right)\right) \\ &= Cov\left(E\left(\tau_{12}|Y_2\right), E\left(\tau_{32}|Y_2\right)\right) = E\left[E\left(\tau_{12}|Y_2\right)E\left(\tau_{32}|Y_2\right)\right] \\ &- E\left[E\left(\tau_{12}|Y_2\right)\right]E\left[E\left(\tau_{32}|Y_2\right)\right] = E\left[E\left(\tau_{12}|Y_2\right)E\left(\tau_{32}|Y_2\right)\right] - E(\tau_{12})E(\tau_{32}) \\ &= E\left[E\left(\tau_{12}|Y_2\right)E\left(\tau_{32}|Y_2\right)\right] - \left[E(\tau_{12})\right]^2 = E\left[\left[E\left(\tau_{12}|Y_2\right)\right]^2\right] - \left[E(\tau_{12})\right]^2. \end{aligned}$$

But

$$E\left(\tau_{12}|Y_2\right) = E\left[\varphi(Y_1)(K_b(Y_1-Y_2)-g(Y_1))\right] = \int \varphi(y)\left[K_b(y-Y_2)-g(y)\right]g(y)dy$$

and

$$\begin{aligned} E\left[\left[E\left(\tau_{12}|Y_2\right)\right]^2\right] &= E\left[\left(\int \varphi(y)\left[K_b(y-Y_2)-g(y)\right]g(y)dy\right)^2\right] \\ &= E\left[\int\int \varphi(y_1)\left(K_b(y_1-Y_2)-g(y_1)\right)g(y_1)\varphi(y_2)\left(K_b(y_2-Y_2)g(y_2)\right)g(y_2)dy_1dy_2\right] \\ &= \int \left(\int \varphi(y_1)\left(K_b(y_1-z)-g(y_1)\right)g(y_1)dy_1\right) \\ &\cdot \left(\int \varphi(y_2)\left(K_b(y_2-z)-g(y_2)\right)g(y_2)dy_2\right)g(z)dz \\ &= \int \left[\int \varphi(y)\left(K_b(y-z)-g(y)\right)g(y)dy\right]^2 g(z)dz, \end{aligned}$$

where

$$\begin{aligned} \int \varphi(y)\left(K_b(y-z)-g(y)\right)g(y)dy &= \int \varphi(y)K_b(y-z)g(y)dy - \int \varphi(y)g(y)^2dy \\ &= \int \varphi(z+bt)K_b(t)g(z+bt)dt - \int f(y)(v(y)-\mu_v)dy \\ &= \gamma(z) + \frac{\mu_2(K)}{2}b^2\gamma''(z) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z) + O(b^6). \end{aligned}$$

So,

$$\begin{aligned}
E\left[\left[E\left(\tau_{12}|Y_2\right)\right]^2\right] &= \int \gamma(z)^2 g(z)dz + \mu_2(K)b^2 \int \gamma(z)\gamma''(z)g(z)dz \\
+ & \frac{\mu_2(K)^2}{4}b^4 \int \gamma''(z)^2 g(z)dz + \frac{\mu_4(K)}{12}b^4 \int \gamma(z)\gamma^{(4)}(z)g(z)dz + O(b^6) \\
= & \int \beta(z)dz + \mu_2(K)b^2 \int \gamma''(z)f(z)(v(z)-\mu_v)dz \\
+ & \frac{\mu_2(K)^2}{4}b^4 \int \gamma''(z)^2 g(z)dz + \frac{\mu_4(K)}{12}b^4 \int \gamma^{(4)}(z)f(z)(v(z)-\mu_v)dz + O(b^6).
\end{aligned}$$

Using now equation (A.37) for $E(\tau_{12})$ gives (A.40). □

Lemma A.1.5. *The variance of τ_{12} is*

$$\begin{aligned}
Var(\tau_{12}) &= \frac{\mu_0(K^2)}{b}\int \alpha(y)dy - \int \beta(y)dy + \frac{\mu_2(K^2)}{2}b\int \delta(y)g''(y)dy \\
&- \mu_2(K)b^2 \int \alpha(y)g''(y)dy + O(b^3). \qquad \text{(A.41)}
\end{aligned}$$

Proof. Let us consider

$$Var(\tau_{12}) = E(\tau_{12}^2) - \left[E(\tau_{12})\right]^2,$$

where

$$\begin{aligned}
E(\tau_{12}^2) &= E\left(\left[\varphi(Y_1)(K_b(Y_1-Y_2)-g(Y_1))\right]^2\right) \\
&= \int\int \varphi(y)^2(K_b(y-z)-g(y))^2 g(y)g(z)dydz \\
&= \int\int \varphi(y)^2 K_b(y-z)^2 g(y)g(z)dydz \\
&- 2\int\int \varphi(y)^2 K_b(y-z)^2 g(y)^2 g(z)dydz \\
&+ \int\int \varphi(y)^2 g(y)^3 g(z)dydz = \int \varphi(y)^2 g(y)\left(\int K_b(y-z)^2 g(z)dz\right)dy \\
&- 2\int \varphi(y)^2 g(y)^2\left(\int K_b(y-z)g(z)dz\right)dy + \int \varphi(y)^2 g(y)^3 dy \\
&= \int \delta(y)\left(\int \frac{1}{b}K(t)^2 g(y-bt)dt\right)dy \\
&- 2\int \alpha(y)\left(\int K(t)g(y-bt)dt\right)dy + \int \beta(y)dy \\
&= \frac{\mu_0(K^2)}{b}\int \alpha(y)dy + \frac{\mu_2(K^2)}{2}b\int \delta(y)g''(y)dy \\
&- \mu_2(K)b^2\int \alpha(y)g''(y)dy - \int \beta(y)dy + O(b^3). \qquad \text{(A.42)}
\end{aligned}$$

Using (A.37) and (A.42) we obtain (A.41). □

Lemma A.1.6. *The covariance of τ_{12} and τ_{21} is*

$$
\begin{aligned}
Cov(\tau_{12},\tau_{21}) = \frac{\mu_0(K^2)}{b}\int\alpha(y)dy - 2\int\beta(y)dy + \frac{\mu_2(K^2)}{2}b\int\gamma(y)\gamma''(y)dy \quad \text{(A.43)}\\
- \quad \mu_2(K)b^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy + \frac{\mu_4(K^2)}{24}b^3\int\gamma(y)\gamma^{(4)}(y)dy + O(b^4).
\end{aligned}
$$

Proof. Let us now consider $Cov(\tau_{12},\tau_{21})$:

$$
Cov(\tau_{12},\tau_{21}) = E(\tau_{12}\tau_{21}) - E(\tau_{12})E(\tau_{21}) = E(\tau_{12}\tau_{21}) - [E(\tau_{12})]^2,
$$

where

$$
\begin{aligned}
E(\tau_{12}\tau_{21}) &= E\left[\varphi(Y_1)(K_b(Y_1-Y_2)-g(Y_1))\varphi(Y_2)(K_b(Y_2-Y_1)-g(Y_2))\right]\\
&= \int\int\varphi(y)(K_b(y-z)-g(y))\varphi(z)(K_b(z-y)-g(z))g(y)g(z)dydz\\
&= \int\int\varphi(y)K_b(y-z)^2\varphi(z)g(y)g(z)dydz\\
&- \int\int\varphi(y)K_b(y-z)\varphi(z)g(y)g(z)^2dydz\\
&- \int\int\varphi(y)K_b(y-z)\varphi(z)g(y)^2g(z)dydz + \int\int\varphi(y)\varphi(z)g(y)^2g(z)^2dydz\\
&= \int\int K_b(y-z)^2\gamma(y)\gamma(z)dydz - 2\int\int K_b(y-z)f(y)(v(y)-\mu_v)\gamma(z)dydz\\
&+ \left(\int f(y)(v(y)-\mu_v)dy\right)^2 = \int\gamma(y)\left(\int K_b(y-z)^2\gamma(z)dz\right)dy\\
&- 2\int f(y)(v(y)-\mu_v)\left(\int K_b(y-z)\gamma(z)dz\right)dy. \qquad \text{(A.44)}
\end{aligned}
$$

The last integral in (A.44) has been dealt with in equation (A.38):

$$
\begin{aligned}
\int K_b(y-z)\gamma(z)dz &= \int K_b(z-y)\gamma(z)dz\\
&= \gamma(z) + \frac{\mu_2(K)}{2}b^2\gamma''(z) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z) + O(b^6),
\end{aligned}
$$

and consequently

$$
\begin{aligned}
2\int f(y)(v(y)-\mu_v)\left(\int K_b(y-z)\gamma(z)dz\right)dy = 2\int\beta(y)dy\\
+ \quad \mu_2(K)b^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy\\
+ \quad \frac{\mu_4(K)}{12}b^4\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy + O(b^6). \qquad \text{(A.45)}
\end{aligned}
$$

On the other hand, the first integral in (A.44) is

$$\begin{aligned}
&\int \gamma(y)\left(\int K_b(y-z)^2\gamma(z)dz\right)dy = \int \gamma(y)\left(\int \frac{1}{b}K(t)^2\gamma(y-bt)dt\right)dy \\
=& \int \gamma(y)\left[\frac{\mu_0(K^2)}{b}\gamma(y)+\frac{\mu_2(K^2)}{2}b\gamma''(y)+\frac{\mu_4(K^2)}{24}b^3\gamma^{(4)}(y)+O(b^5)\right]dy \\
=& \frac{\mu_0(K^2)}{b}\int \alpha(y)dy+\frac{\mu_2(K^2)}{2}b\int \gamma(y)\gamma''(y)dy \\
+& \frac{\mu_4(K^2)}{24}b^3\int \gamma(y)\gamma^{(4)}(y)dy+O(b^5). \qquad \text{(A.46)}
\end{aligned}$$

Now, using (A.45) and (A.46) in (A.44) we get

$$\begin{aligned}
E(\tau_{12}\tau_{21}) =& \frac{\mu_0(K^2)}{b}\int \alpha(y)dy-2\int \beta(y)dy+\frac{\mu_2(K^2)}{2}b\int \gamma(y)\gamma''(y)dy \\
-& \mu_2(K)b^2\int \gamma''(y)f(y)(v(y)-\mu_v)dy+\frac{\mu_4(K^2)}{24}b^3\int \gamma(y)\gamma^{(4)}(y)dy \\
-& \frac{\mu_4(K)}{12}b^4\int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy+O(b^5). \qquad \text{(A.47)}
\end{aligned}$$

As a consequence of expression (A.37), we have:

$$E(\tau_{12}) = \frac{\mu_2(K)}{2}b^2\int \gamma(y)g''(y)dy+\frac{\mu_4(K)}{24}b^4\int \gamma(y)g^{(4)}(y)dy+O(b^6).$$

Using the previous expression and expression (A.47), we have (A.43). □

Lemma A.1.7. *The covariance of* τ_{12} *and* τ_{11} *is*

$$\begin{aligned}
Cov(\tau_{12},\tau_{11}) =& \frac{\mu_2(K)K(0)}{2}b\left[\int \delta(y)g''(y)dy-\left(\int \gamma(y)g''(y)dy\right)\left(\int \gamma(y)dy\right)\right] \\
-& \frac{\mu_2(K)}{2}b^2\int \alpha(y)g''(y)dy+\frac{\mu_4(K)K(0)}{24}b^3\left[\int \delta(y)g^{(4)}(y)dy\right. \qquad \text{(A.48)} \\
-& \left.\left(\int \gamma(y)g^{(4)}(y)dy\right)\left(\int \gamma(y)dy\right)\right]-\frac{\mu_4(K)}{24}b^4\int \alpha(y)g^{(4)}(y)dy+O(b^5).
\end{aligned}$$

Proof. Let's deal now with the term:

$$\begin{aligned}
Cov(\tau_{12},\tau_{11}) &= Cov\left(E\left(\tau_{12}|Y_1\right),E\left(\tau_{11}|Y_1\right)\right)+E\left[Cov(\tau_{12},\tau_{11}|Y_1)\right] \\
&= Cov\left(E\left(\varphi(Y_1)(K_b(Y_1-Y_2)-g(Y_1))|Y_1\right),\tau_{11}\right),
\end{aligned}$$

since

$$\tau_{11} = \varphi(Y_1)(K_b(Y_1-Y_1)-g(Y_1)) = \varphi(Y_1)\left(\frac{K(0)}{b}-g(Y_1)\right) \qquad \text{(A.49)}$$

is a measurable function of Y_1 and then $Cov(\tau_{12},\tau_{11}|Y_1)=0$.

On the other hand,

$$\begin{aligned} E\left[\varphi(Y_1)(K_b(Y_1-Y_2)-g(Y_1))|Y_1\right] &= \varphi(Y_1)(E(K_b(Y_1-Y_2)|Y_1)-g(Y_1)) \\ &= \varphi(Y_1)\left[(K_b*g)(Y_1)-g(Y_1)\right]. \end{aligned}$$

Using the previous expressions we can obtain a simpler expression for $Cov(\tau_{12},\tau_{11})$:

$$\begin{aligned} Cov(\tau_{12},\tau_{11}) = E&\left[\varphi(Y_1)\left((K_b*g)(Y_1)-g(Y_1)\right)\varphi(Y_1)\left(\frac{K(0)}{b}-g(Y_1)\right)\right] \\ - \quad & E\left[\varphi(Y_1)\left((K_b*g)(Y_1)-g(Y_1)\right)\right]E\left[\varphi(Y_1)\left(\frac{K(0)}{b}-g(Y_1)\right)\right] \\ = \quad & \int\varphi(y)^2\left((K_b*g)(y)-g(y)\right)\left(\frac{K(0)}{b}-g(y)\right)g(y)dy \\ - \quad & \left[\frac{\mu_2(K)}{2}b^2\int\gamma(y)g''(y)dy+\frac{\mu_4(K)}{24}b^4\int\gamma(y)g^{(4)}(y)dy+O(b^6)\right] \\ \cdot \quad & \left[\frac{K(0)}{b}E\left[\varphi(Y_1)\right]-E\left[\gamma(Y_1)\right]\right]. \end{aligned} \tag{A.50}$$

But

$$\begin{aligned} &\int\varphi(y)^2\left((K_b*g)(y)-g(y)\right)\left(\frac{K(0)}{b}-g(y)\right)g(y)dy \\ &\quad = \int\varphi(y)^2\left[\frac{\mu_2(K)}{2}b^2g''(y)+\frac{\mu_4(K)}{24}b^4g^{(4)}(y)+O(b^6)\right] \\ &\quad \cdot \left(\frac{K(0)}{b}-g(y)\right)g(y)dy = \frac{\mu_2(K)K(0)}{2}b\int\delta(y)g''(y)dy \\ &\quad - \frac{\mu_2(K)}{2}b^2\int\alpha(y)g''(y)dy+\frac{\mu_4(K)K(0)}{24}b^3\int\delta(y)g^{(4)}(y)dy \\ &\quad - \frac{\mu_4(K)}{24}b^4\int\alpha(y)g^{(4)}(y)dy+O(b^5) \end{aligned} \tag{A.51}$$

and

$$\begin{aligned} \frac{K(0)}{b}E\left[\varphi(Y_1)\right]-E\left[\gamma(Y_1)\right] &= \frac{K(0)}{b}\int\varphi(y)g(y)dy-\int f(y)(v(y)-\mu_v)dy \\ &= \frac{K(0)}{b}\int\gamma(y)dy. \end{aligned} \tag{A.52}$$

Using (A.51) and (A.52) in (A.50) we get (A.48). □

Lemma A.1.8. *The covariance of τ_{12} and τ_{22} is*

$$\begin{aligned}
Cov(\tau_{12},\tau_{22}) = \frac{K(0)}{b}\int \alpha(y)dy - \int \beta(y)dy + \frac{\mu_2(K)K(0)}{2}b\Bigg[\int \gamma(y)\gamma''(y)dy \\
- \quad \left(\int \gamma''(y)g(y)dy\right)\left(\int \gamma(y)dy\right)\Bigg] - \frac{\mu_2(K)}{2}b^2\int \gamma''(y)f(y)(v(y)-\mu_v)dy \\
+ \quad \frac{\mu_4(K)K(0)}{24}b^3\left[\int \gamma(y)\gamma^{(4)}(y)dy - \left(\int \gamma^{(4)}(y)g(y)dy\right)\cdot\left(\int \gamma(y)dy\right)\right] \\
- \quad \frac{\mu_4(K)}{24}b^4\int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy + O(b^5). \qquad \text{(A.53)}
\end{aligned}$$

Proof.

$$\begin{aligned}
Cov(\tau_{12},\tau_{22}) &= Cov\left(E\left(\tau_{12}|Y_2\right),E\left(\tau_{22}|Y_2\right)\right) + E\left[Cov(\tau_{12},\tau_{22}|Y_2)\right] \\
&= Cov\left(E\left(\varphi(Y_1)(K_b(Y_1-Y_2)-g(Y_1))|Y_2\right),\tau_{22}\right),
\end{aligned}$$

since $Cov(\tau_{12},\tau_{22}|Y_2)=0$ (because τ_{22} is a measurable function of Y_2).

On the other hand,

$$\begin{aligned}
&E\left[\varphi(Y_1)(K_b(Y_1-Y_2)-g(Y_1))|Y_2\right] = E\left[\varphi(Y_1)K_b(Y_1-Y_2)|Y_2\right] - E\left[\gamma(Y_1)|Y_2\right] \\
&= \quad \int \varphi(y)K_b(y-Y_2)g(y)dy - E\left[\gamma(Y_1)\right] = \int \gamma(y)K_b(Y_2-y)dy \\
&- \quad \int f(y)(v(y)-\mu_v)dy = (\gamma * K_b)(Y_2).
\end{aligned}$$

Using the previous expressions and $\tau_{22}=\varphi(Y_2)\left(\frac{K(0)}{b}-g(Y_2)\right)$, we have:

$$\begin{aligned}
Cov(\tau_{12},\tau_{22}) &= Cov\left((\gamma * K_b)(Y_2),\varphi(Y_2)\left(\frac{K(0)}{b}-g(Y_2)\right)\right) \\
&= E\left[(\gamma * K_b)(Y_2)\varphi(Y_2)\left(\frac{K(0)}{b}-g(Y_2)\right)\right] \\
&- E\left[(\gamma * K_b)(Y_2)\right]E\left[\varphi(Y_2)\left(\frac{K(0)}{b}-g(Y_2)\right)\right], \qquad \text{(A.54)}
\end{aligned}$$

where

$$E\left[(\gamma * K_b)(Y_2)\varphi(Y_2)\left(\frac{K(0)}{b}-g(Y_2)\right)\right] = \int (\gamma * K_b)(y)\varphi(y)\left(\frac{K(0)}{b}-g(y)\right)g(y)dy$$

with

$$\begin{aligned}
(\gamma * K_b)(y) &= \int \gamma(z)K_b(y-z)dz = \int \gamma(y-bt)K(t)dt \\
&= \gamma(y) + \frac{\mu_2(K)}{2}b^2\gamma''(y) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(y) + O(b^6).
\end{aligned}$$

Consequently,

$$
\begin{aligned}
E\left[(\gamma * K_b)(Y_2)\varphi(Y_2)\left(\frac{K(0)}{b} - g(Y_2)\right)\right] &= \int \gamma(y)\varphi(y)\left(\frac{K(0)}{b} - g(y)\right) g(y)dy \\
+ & \frac{\mu_2(K)}{2}b^2 \int \gamma''(y)\varphi(y)\left(\frac{K(0)}{b} - g(y)\right) g(y)dy \\
+ & \frac{\mu_4(K)}{24}b^4 \int \gamma^{(4)}(y)\varphi(y)\left(\frac{K(0)}{b} - g(y)\right) g(y)dy + O(b^5) \\
= & \frac{K(0)}{b}\int \alpha(y)dy - \int \beta(y)dy + \frac{\mu_2(K)K(0)}{2}b\int \gamma(y)\gamma''(y)dy \\
- & \frac{\mu_2(K)}{2}b^2 \int \gamma''(y)f(y)(v(y)-\mu_v)dy + \frac{\mu_4(K)K(0)}{24}b^3\int \gamma(y)\gamma^{(4)}(y)dy \\
- & \frac{\mu_4(K)}{24}b^4 \int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy + O(b^5) \qquad \text{(A.55)}
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[(\gamma * K_b)(Y_2)\right] &= \int (\gamma * K_b)(y)g(y)dy = \int \gamma(y)g(y)dy + \frac{\mu_2(K)}{2}b^2\int \gamma''(y)g(y)dy \\
+ & \frac{\mu_4(K)}{24}b^4\int \gamma^{(4)}(y)g(y)dy + O(b^6) = \int f(y)(v(y)-\mu_v)dy \\
+ & \frac{\mu_2(K)}{2}b^2\int \gamma''(y)g(y)dy + \frac{\mu_4(K)}{24}b^4\int \gamma^{(4)}(y)g(y)dy + O(b^6) \\
= & \frac{\mu_2(K)}{2}b^2\int \gamma''(y)g(y)dy + \frac{\mu_4(K)}{24}b^4\int \gamma^{(4)}(y)g(y)dy + O(b^6). \qquad \text{(A.56)}
\end{aligned}
$$

The last factor in (A.54) is exactly the same as the equation (A.52). Now, using (A.55), (A.56) and (A.52) in (A.54) results in (A.53). □

Lemma A.1.9. *The variance of* τ_{11} *is*

$$
\begin{aligned}
Var(\tau_{11}) &= \frac{K(0)^2}{b^2}\left[\int \delta(y)dy - \left(\int \gamma(y)dy\right)^2\right] \\
&- \frac{2K(0)}{b}\int \alpha(y)dy + \int \beta(y)dy. \qquad \text{(A.57)}
\end{aligned}
$$

Proof. Let us consider

$$Var(\tau_{11}) = E(\tau_{11}^2) - E(\tau_{11})^2,$$

where the last term can be directly obtained from equation (A.52):

$$E(\tau_{11})^2 = \left[E\left(\varphi(Y_1)\left(\frac{K(0)}{b} - g(Y_1)\right)\right)\right]^2 = \frac{K(0)^2}{b^2}\left(\int \gamma(y)dy\right)^2. \qquad \text{(A.58)}$$

For the term $E(\tau_{11}^2)$ we use the expression (A.49) for τ_{11} to get:

$$\begin{aligned}
E(\tau_{11}^2) &= E\left[\varphi(Y_1)^2\left(\frac{K(0)}{b} - g(Y_1)\right)^2\right] \\
&= \frac{K(0)^2}{b^2}E\left[\varphi(Y_1)^2\right] - 2\frac{K(0)}{b}E\left[\delta(Y_1)\right] + E\left[\alpha(Y_1)\right] \\
&= \frac{K(0)^2}{b^2}\int \delta(y)dy - \frac{2K(0)}{b}\int \alpha(y)dy + \int \beta(y)dy. \qquad \text{(A.59)}
\end{aligned}$$

Using (A.58) and (A.59) we obtain (A.57). □

We will proceed now with the covariance terms in (4.14).

Lemma 4.5.7. *The covariance between $\widehat{A}_1$ and $\widehat{A}_2$ is*

$$Cov\left(\widehat{A}_1, \widehat{A}_2\right) = D_{27}\frac{h^2}{N} + D_{28}\frac{h^4}{N} + O\left(\frac{h^6}{N}\right), \qquad (4.18)$$

where

$$\begin{aligned}
D_{27} &:= \frac{\mu_2(K)}{2}\int \theta(y)f''(y)dy, \\
D_{28} &:= \frac{\mu_4(K)}{24}\int \theta(y)f^{(4)}(y)dy.
\end{aligned}$$

Proof. Let us define

$$\omega_i := \frac{f(Y_i)}{g(Y_i)}(v(Y_i) - \mu_v).$$

Using this definition and the expressions for $\widehat{A}_1$ in (A.17) we have

$$\widehat{A}_1 = \frac{1}{N}\sum_{i=1}^{N}\omega_i.$$

Let us first consider $Cov(\widehat{A}_1, \widehat{A}_2)$:

$$\begin{aligned}
Cov(\widehat{A}_1, \widehat{A}_2) &= Cov\left(\frac{1}{N}\sum_{i=1}^{N}\omega_i, \frac{1}{Nn}\sum_{j=1}^{N}\sum_{k=1}^{n}\eta_{jk}\right) \\
&= \frac{1}{N^2 n}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{n}Cov(\omega_i, \eta_{jk}) = \frac{1}{N}Cov(\omega_1, \eta_{11}), \qquad \text{(A.60)}
\end{aligned}$$

since $Cov(\omega_i, \eta_{jk}) = 0$ for $i \neq j$ because ω_i and η_{jk} are independent for $i \neq j$.

On the other hand,

$$
\begin{aligned}
Cov(\omega_1, \eta_{11}) &= Cov\left(\gamma(Y_1), \frac{K_h(Y_1 - X_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right) \\
&= E\left[\frac{\gamma(Y_1)}{g(Y_1)}\left(K_h(Y_1 - X_1) - f(Y_1)\right)(v(Y_1) - \mu_v)\right] \\
&- E\left(\gamma(Y_1)\right) E\left[\frac{K_h(Y_1 - X_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right] \\
&= E\left[\varphi(Y_1)\left(K_h(Y_1 - X_1) - f(Y_1)\right)(v(Y_1) - \mu_v)\right],
\end{aligned}
$$

since

$$E\left(\gamma(Y_1)\right) = \int \gamma(y)g(y)dy = \int f(y)(v(y) - \mu_v) = 0$$

and

$$
\begin{aligned}
&E\left[\frac{K_h(Y_1 - X_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right] \\
&= E\left[E\left[\frac{K_h(Y_1 - X_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)|Y_1\right]\right] \\
&= E\left[\frac{E\left[K_h(Y_1 - X_1)|Y_1\right] - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right] \\
&= E\left[\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right] \\
&= \int \frac{(K_h * f)(y) - f(y)}{g(y)}(v(y) - \mu_v)g(y)dy \\
&= \int (K_h * f)(y)(y - \mu)dy - \int f(y)(v(y) - \mu_v)dy \\
&= \int \left(\int K_h(y - z)f(z)dz\right)(v(y) - \mu_v)dy \\
&= \int f(z)\left(\int K(t)(v(z + ht) - \mu_v)dt\right)dz \\
&= \int f(z)\left(\int K(t)v(z + ht)dt\right)dz - \mu_v \\
&= \frac{h^2}{2}\mu_2(K)\int v''(z)f(z)dz + \frac{h^4}{24}\mu_4(K)\int v^{(4)}(z)f(z)dz + O(h^6).
\end{aligned}
$$

Thus:

$$E\left(\gamma(Y_1)\right) E\left[\frac{K_h(Y_1 - X_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right] = 0.$$

Consequently,

$$
\begin{aligned}
Cov(\omega_1, \eta_{11}) &= E\left[\varphi(Y_1)\left(K_h(Y_1 - X_1) - f(Y_1)\right)(v(Y_1) - \mu_v)\right] \\
&= E\left[E\left[\varphi(Y_1)\left(K_h(Y_1 - X_1) - f(Y_1)\right)(v(Y_1) - \mu_v)|Y_1\right]\right] \\
&= E\left[\varphi(Y_1)\left(E\left[K_h(Y_1 - X_1)|Y_1\right] - f(Y_1)\right)(v(Y_1) - \mu_v)\right] \\
&= E\left[\varphi(Y_1)\left((K_h * f)(Y_1) - f(Y_1)\right)(v(Y_1) - \mu_v)\right] \\
&= \int \varphi(y)\left((K_h * f)(y) - f(y)\right)(v(y) - \mu_v)g(y)dy \\
&= \int \gamma(y)\left(\frac{\mu_2(K)}{2}h^2 f''(y) + \frac{\mu_4(K)}{24}h^4 f^{(4)}(y) + O(h^6)\right)(v(y) - \mu_v)dy \\
&= \frac{\mu_2(K)}{2}h^2 \int \gamma(y) f''(y)(v(y) - \mu_v)dy \\
&+ \frac{\mu_4(K)}{24}h^4 \int \gamma(y) f^{(4)}(y)(v(y) - \mu_v)dy + O(h^6).
\end{aligned}
$$

Using the previous expression in (A.60) gives (4.18). □

Lemma 4.5.8. *The covariance between $\widehat{A}_1$ and $\widehat{A}_3$ is*

$$
\begin{aligned}
Cov\left(\widehat{A}_1, \widehat{A}_3\right) &= D_{29}\frac{1}{N} + D_{30}\frac{1}{N^2 b} + D_{31}\frac{1}{N^2} + D_{32}\frac{b^2}{N} \\
&+ D_{33}\frac{b^4}{N} + O\left(\frac{b^2}{N^2}\right) + O\left(\frac{b^6}{N}\right),
\end{aligned}
\tag{4.19}
$$

where

$$
\begin{aligned}
D_{29} &:= \int \beta(y)dy = D_{18}, \\
D_{30} &:= K(0)\int \alpha(y)dy, \\
D_{31} &:= -2\int \beta(y)dy = -2D_{18}, \\
D_{32} &:= \frac{\mu_2(K)}{2}\left[\int \alpha(y)g''(y)dy + \int \gamma''(y)f(y)(v(y) - \mu_v)dy\right], \\
D_{33} &:= \frac{\mu_4(K)}{24}\left[\int \alpha(y)g^{(4)}(y)dy + \int \gamma^{(4)}(y)f(y)(v(y) - \mu_v)dy\right].
\end{aligned}
$$

Proof. Let us now consider $Cov(\widehat{A}_1, \widehat{A}_3)$:

$$
\begin{aligned}
Cov(\widehat{A}_1, \widehat{A}_3) &= Cov\left(\frac{1}{N}\sum_{i=1}^{N}\omega_i, \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\tau_{jk}\right) = \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} Cov(\omega_i, \tau_{jk}) \\
&= \frac{N-1}{N^2}Cov(\omega_1, \tau_{12}) + \frac{N-1}{N^2}Cov(\omega_1, \tau_{21}) + \frac{1}{N^2}Cov(\omega_1, \tau_{11}),
\end{aligned}
\tag{A.61}
$$

since ω_1 and τ_{23} are independent and ω_1 and τ_{22} are also independent.

Let us study now the terms in (A.61):

$$Cov(\omega_1, \tau_{12}) = E\left[Cov(\omega_1, \tau_{12}|Y_1)\right] + Cov(E(\omega_1|Y_1), E(\tau_{12}|Y_1)).$$

But $Cov(\omega_1, \tau_{12}|Y_1) = 0$ since ω_1 is a function of Y_1, $E(\omega_1|Y_1) = \omega_1 = \gamma(Y_1)$, and

$$\begin{aligned} E(\tau_{12}|Y_1) &= E\left[\varphi(Y_1)\left(K_b(Y_1 - Y_2) - g(Y_1)\right)|Y_1\right] \\ &= \varphi(Y_1)\left(E\left[K_b(Y_1 - Y_2)|Y_1\right] - g(Y_1)\right) = \varphi(Y_1)\left((K_b * g)(Y_1) - g(Y_1)\right) \end{aligned}$$

and

$$E(\omega_1) = E(\gamma(Y_1)) = \int \gamma(y)g(y)dy = \int f(y)(v(y) - \mu_v) = 0.$$

Thus

$$\begin{aligned} Cov(\omega_1, \tau_{12}) &= Cov(E(\omega_1|Y_1), E(\tau_{12}|Y_1)) = E\left[E(\omega_1|Y_1)E(\tau_{12}|Y_1)\right] \\ &- E\left[E(\omega_1|Y_1)\right]E\left[E(\tau_{12}|Y_1)\right] = E\left[\omega_1 E(\tau_{12}|Y_1)\right] - E(\omega_1)E(\tau_{12}) \\ &= E\left[\gamma(Y_1)\varphi(Y_1)\left((K_b * g)(Y_1) - g(Y_1)\right)\right] \\ &= \int \delta(y)\left((K_b * g)(y) - g(y)\right)g(y)dy \\ &= \int \alpha(y)\left(\frac{\mu_2(K)}{2}b^2 g''(y) + \frac{\mu_4(K)}{24}b^4 g^{(4)}(y) + O(b^6)\right)dy \\ &= \frac{\mu_2(K)}{2}b^2 \int \alpha(y)g''(y)dy + \frac{\mu_4(K)}{24}b^4 \int \alpha(y)g^{(4)}(y)dy + O(b^6). \quad \text{(A.62)} \end{aligned}$$

Now consider

$$\begin{aligned} Cov(\omega_1, \tau_{21}) &= E\left[Cov(\omega_1, \tau_{21}|Y_1)\right] + Cov(E(\omega_1|Y_1), E(\tau_{21}|Y_1)) \\ &= Cov(\omega_1, E(\tau_{21}|Y_1)) = E\left[\omega_1 E(\tau_{21}|Y_1)\right] - E(\omega_1)E\left[E(\tau_{21}|Y_1)\right] \\ &= E\left[\omega_1 E(\tau_{21}|Y_1)\right]. \end{aligned}$$

But

$$E(\tau_{21}|Y_1) = E\left[\varphi(Y_2)\left(K_b(Y_2 - Y_1) - g(Y_2)\right)|Y_1\right] = \int \varphi(y)\left(K_b(y - Y_1) - g(y)\right)g(y)dy.$$

Thus

$$\begin{aligned} Cov(\omega_1, \tau_{21}) &= \int \gamma(z)\left(\int \varphi(y)\left(K_b(y - z) - g(y)\right)g(y)dy\right)g(z)dz \\ &= \int\int \left(K_b(y - z) - g(y)\right)\gamma(y)f(z)(z - \mu)dydz \\ &= \int\int K_b(y - z)\gamma(y)f(z)(v(z) - \mu_v)dydz \\ &- \int\int f(y)(v(y) - \mu_v)f(z)(v(z) - \mu_v)dydz. \quad \text{(A.63)} \end{aligned}$$

But

$$\int\int f(y)(v(y)-\mu_v)f(z)(v(z)-\mu_v)dydz = \left[\int f(y)(v(y)-\mu_v)dy\right]^2 = 0 \quad \text{(A.64)}$$

and

$$\begin{aligned}
&\int\int K_b(y-z)\gamma(y)f(z)(v(z)-\mu_v)dydz \\
= &\int\left(\int K(t)\gamma(z+bt)dt\right)f(z)(v(z)-\mu_v)dz \\
= &\int\left(\gamma(z)+\frac{\mu_2(K)}{2}b^2\gamma''(z)+\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z)+O(b^6)\right)f(z)(v(z)-\mu_v z-\mu)dz \\
= &\int\beta(z)dz+\frac{\mu_2(K)}{2}b^2\int\gamma''(z)f(z)(v(z)-\mu_v)dz \\
+ &\frac{\mu_4(K)}{24}b^4\int\gamma^{(4)}(z)f(z)(v(z)-\mu_v)dz+O(b^6). \qquad \text{(A.65)}
\end{aligned}$$

Using (A.64) and (A.65) in (A.63) we get:

$$\begin{aligned}
Cov(\omega_1,\tau_{21}) &= \int\beta(y)dy+\frac{\mu_2(K)}{2}b^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy \\
&+ \frac{\mu_4(K)}{24}b^4\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy+O(b^6). \qquad \text{(A.66)}
\end{aligned}$$

The last term in (A.61) is:

$$\begin{aligned}
&Cov(\omega_1,\tau_{11}) = E(\omega_1\tau_{11}) - E(\omega_1)E(\tau_{11}) = E(\omega_1\tau_{11}) \\
&= E\left[\gamma(Y_1)\varphi(Y_1)\left(K_b(Y_1-Y_1)-g(Y_1)\right)\right] = \int\gamma(y)\varphi(y)\left(K_b(0)-g(y)\right)g(y)dy \\
&= \int\alpha(y)\left(\frac{K(0)}{b}-g(y)\right)dy = \frac{K(0)}{b}\int\alpha(y)dy-\int\beta(y)dy. \qquad \text{(A.67)}
\end{aligned}$$

Using (A.62), (A.66) and (A.67) in (A.61) gives (4.19). □

Lemma 4.5.9. *The covariance between* $\widehat{A}_2$ *and* $\widehat{A}_3$ *is*

$$\begin{aligned}
Cov\left(\widehat{A}_2,\widehat{A}_3\right) &= D_{34}\frac{h^2}{N}+D_{35}\frac{h^2b^2}{N}+D_{36}\frac{h^4}{N}+D_{37}\frac{h^4b^2}{N}+D_{38}\frac{h^2b^4}{N} \\
&+ D_{39}\frac{h^2}{N^2b}+D_{40}\frac{h^2}{N^2}+D_{41}\frac{h^4}{N^2b}+D_{42}\frac{h^4}{N^2}+D_{43}\frac{h^2b^2}{N^2}+O\left(\frac{h^6}{N}\right) \\
&+ O\left(\frac{h^2b^6}{N}\right)+O\left(\frac{h^4b^4}{N}\right)+O\left(\frac{h^6}{N^2b}\right)+O\left(\frac{h^2b^4}{N^2}\right)+O\left(\frac{h^4b^2}{N^2}\right), \quad \text{(4.20)}
\end{aligned}$$

where

$$\begin{aligned}
D_{34} &:= \frac{\mu_2(K)}{2}\int \theta(y)f''(y)dy,\\
D_{35} &:= \frac{\mu_2(K)^2}{4}\left[\int \rho(y)f''(y)g''(y)dy + \int f''(y)\gamma''(y)(v(y)-\mu_v)dy\right.\\
&- \left. 2\int v''(y)f(y)dy\int \gamma(y)g''(y)dy\right],\\
D_{36} &:= \frac{\mu_4(K)}{24}\int \theta(y)f^{(4)}(y)dy,\\
D_{37} &:= \frac{\mu_2(K)\mu_4(K)}{48}\left[\int \rho(y)f^{(4)}(y)g''(y)dy + \int f^{(4)}(y)\gamma''(y)(v(y)-\mu_v)dy\right.\\
&- \left. 2\int v^{(4)}(y)f(y)dy\int \gamma(y)g''(y)dy\right],\\
D_{38} &:= \frac{\mu_2(K)\mu_4(K)}{48}\left[\int \rho(y)f''(y)g^{(4)}(y)dy + \int f''(y)\gamma^{(4)}(y)(v(y)-\mu_v)dy\right.\\
&- \left. 2\int v''(y)f(y)dy\int \gamma(y)g^{(4)}(y)dy\right],\\
D_{39} &:= \frac{\mu_2(K)K(0)}{2}\left[\int \rho(y)f''(y)dy - \int v''(y)f(y)dy\int \gamma(y)dy\right],\\
D_{40} &:= -\mu_2(K)\int \theta(y)f''(y)dy = -2D_{34},\\
D_{41} &:= \frac{\mu_4(K)K(0)}{24}\left[\int \rho(y)f^{(4)}(y)dy - \int v^{(4)}(y)f(y)dy\int \gamma(y)dy\right],\\
D_{42} &:= -\frac{\mu_4(K)}{12}\int \theta(y)f^{(4)}(y)dy = -2D_{36},\\
D_{43} &:= -D_{35}.
\end{aligned}$$

Proof. Let us consider $Cov(\widehat{A}_2, \widehat{A}_3)$:

$$\begin{aligned}
Cov(\widehat{A}_2, \widehat{A}_3) &= Cov\left(\frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}\eta_{ij}, \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\tau_{kl}\right)\\
&= \frac{1}{N^3 n}\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{N} Cov(\eta_{ij}, \tau_{kl})\\
&= \frac{N-1}{N^2}Cov(\eta_{11}, \tau_{12}) + \frac{N-1}{N^2}Cov(\eta_{11}, \tau_{21}) + \frac{1}{N^2}Cov(\eta_{11}, \tau_{11}), \quad \text{(A.68)}
\end{aligned}$$

since η_{11} only depends on Y_1 and X_1 and τ_{23} only depends on Y_2 and Y_3, then η_{11} and τ_{23} are independent. Similiarly η_{11} and τ_{22} are also independent. Consequently:

$$Cov(\eta_{11}, \tau_{23}) = 0 \quad \text{and} \quad Cov(\eta_{11}, \tau_{22}) = 0.$$

Let us now consider the three terms in (A.68):

$$\begin{aligned} Cov(\eta_{11}, \tau_{12}) &= E\left[Cov(\eta_{11}, \tau_{12}|Y_1)\right] + Cov(E(\eta_{11}|Y_1), E(\tau_{12}|Y_1)) \\ &= Cov(E(\eta_{11}|Y_1), E(\tau_{12}|Y_1)), \end{aligned}$$

since $Cov(\eta_{11}, \tau_{12}|Y_1) = 0$ because η_{11} and τ_{12} are conditionally independent given Y_1 (η_{11} is only function of Y_1 and X_1, τ_{12} is only function of Y_1 and Y_2 and X_1 and Y_2 are independent).

On the other hand, using expressions (A.23) and (A.33), we have:

$$\begin{aligned} &Cov(\eta_{11}, \tau_{12}) = Cov\left(E(\eta_{11}|Y_1), E(\tau_{12}|Y_1)\right) \\ &= Cov\left(\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v), \varphi(Y_1)\left((K_b * g)(Y_1) - g(Y_1)\right)\right) \\ &= E\left(\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\varphi(Y_1)\left((K_b * g)(Y_1) - g(Y_1)\right)\right) \\ &- E(\eta_{11})E(\tau_{12}) \qquad \text{(A.69)} \\ &= E\left[\sigma(Y_1)\left[(K_h * f)(Y_1) - f(Y_1)\right]\left[(K_b * g)(Y_1) - g(Y_1)\right]\right] - E(\eta_{11})E(\tau_{12}), \end{aligned}$$

where

$$\sigma(y) := \frac{\varphi(y)}{g(y)}(v(y) - \mu_v) = \frac{f(y)}{g(y)^3}(v(y) - \mu_v)^2.$$

The first term in (A.69) is

$$\begin{aligned} &E\left[\sigma(Y_1)\left[(K_h * f)(Y_1) - f(Y_1)\right]\left[(K_b * g)(Y_1) - g(Y_1)\right]\right] \\ &= \int \sigma(y)\left[(K_h * f)(y) - f(y)\right]\left[(K_b * g)(y) - g(y)\right]g(y)dy \\ &= \int \sigma(y)\left[\frac{\mu_2(K)}{2}h^2 f''(y) + \frac{\mu_4(K)}{24}h^4 f^{(4)}(y) + O(h^6)\right] \\ &\cdot \left[\frac{\mu_2(K)}{2}b^2 g''(y) + \frac{\mu_4(K)}{24}b^4 g^{(4)}(y) + O(b^6)\right]g(y)dy \\ &= \frac{\mu_2(K)^2}{4}h^2 b^2 \int \rho(y) f''(y) g''(y)dy \\ &+ \frac{\mu_2(K)\mu_4(K)}{48}\left[h^2 b^4 \int \rho(y) f''(y) g^{(4)}(y)dy + h^4 b^2 \int \rho(y) f^{(4)}(y) g''(y)dy\right] \\ &+ O(h^6 b^2) + O(h^4 b^4) + O(h^2 b^6), \qquad \text{(A.70)} \end{aligned}$$

where

$$\rho(y) := \sigma(y)g(y) = \frac{f(y)}{g(y)^2}(v(y) - \mu_v)^2.$$

On the other hand, since

$$\begin{aligned} E(\eta_{11}) &= E\left[E(\eta_{11}|Y_1)\right] = E\left[\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\right] \\ &= \frac{h^2}{2}\mu_2(K)\int v''(z)f(z)dz + \frac{h^4}{24}\mu_4(K)\int v^{(4)}(z)f(z)dz + O(h^6) \end{aligned}$$

and

$$E(\tau_{12}) = \frac{\mu_2(K)}{2}b^2\int \gamma(y)g''(y)dy + \frac{\mu_4(K)}{24}b^4\int \gamma(y)g^{(4)}(y)dy + O(b^6),$$

the second term in (A.69) is

$$\begin{aligned} E(\eta_{11})E(\tau_{12}) = &\frac{\mu_2(K)^2}{4}h^2b^2\int v''(y)f(y)dy\int \gamma(y)g''(y)dy \\ + &\frac{\mu_2(K)\mu_4(K)}{48}\left[h^2b^4\int v''(y)f(y)dy\int \gamma(y)g^{(4)}(y)dy\right. \qquad \text{(A.71)} \\ + &\left. h^4b^2\int v^{(4)}(y)f(y)dy\int \gamma(y)g''(y)dy\right] + O(h^6b^2) + O(h^4b^4) + O(h^2b^6). \end{aligned}$$

Plugging (A.70) and (A.71) into (A.69) leads to:

$$\begin{aligned} Cov(\eta_{11}, \tau_{12}) = &\frac{\mu_2(K)^2}{4}h^2b^2\left[\int \rho(y)f''(y)g''(y)dy - \int v''(y)f(y)dy\int \gamma(y)g''(y)dy\right] \\ + &\frac{\mu_2(K)\mu_4(K)}{48}\left[h^2b^4\left(\int \rho(y)f''(y)g^{(4)}(y)dy - \int v''(y)f(y)dy\int \gamma(y)g^{(4)}(y)dy\right)\right. \\ + &\left. h^4b^2\left(\int \rho(y)f^{(4)}(y)g''(y)dy - \int v^{(4)}(y)f(y)dy\int \gamma(y)g''(y)dy\right)\right] + O(h^6b^2) \\ + &O(h^4b^4) + O(h^2b^6). \qquad \text{(A.72)} \end{aligned}$$

Let's deal now with the term $Cov(\eta_{11}, \tau_{21})$ in (A.68):

$$Cov(\eta_{11}, \tau_{21}) = E(\eta_{11}\tau_{21}) - E(\eta_{11})E(\tau_{21}) = E(\eta_{11}\tau_{21}) - E(\eta_{11})E(\tau_{12}),$$

since $E(\tau_{12}) = E(\tau_{21})$.

Thus, using that τ_{21} is a function of only Y_1 and Y_2 we get:

$$\begin{aligned} Cov(\eta_{11}, \tau_{21}) &= E(\eta_{11}\tau_{21}) - E(\eta_{11})E(\tau_{12}) = E\left[E(\eta_{11}\tau_{21}|Y_1, Y_2)\right] - E(\eta_{11})E(\tau_{12}) \\ &= E\left[E(\eta_{11}|Y_1, Y_2)\tau_{21}\right] - E(\eta_{11})E(\tau_{12}) \\ &= E\left[E(\eta_{11}|Y_1)\tau_{21}\right] - E(\eta_{11})E(\tau_{12}), \qquad \text{(A.73)} \end{aligned}$$

since η_{11} is independent of Y_2.

Using once more the expression (A.23) for $E(\eta_{11}|Y_1)$, we have:

$$\begin{aligned}
&E\left[E(\eta_{11}|Y_1)\tau_{21}\right] \\
&\quad= E\left[\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\varphi(Y_2)\left(K_b(Y_2-Y_1)-g(Y_2)\right)\right] \\
&\quad= E\left[\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\varphi(Y_2)K_b(Y_2-Y_1)\right] \\
&\quad- E\left[\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\gamma(Y_2)\right]. \qquad \text{(A.74)}
\end{aligned}$$

Let us consider the two terms in (A.74):

$$\begin{aligned}
&E\left[\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\varphi(Y_2)K_b(Y_2-Y_1)\right] \\
&\quad= \int\int\frac{(K_h*f)(y)-f(y)}{g(y)}(v(y)-\mu_v)\varphi(z)K_b(z-y)g(y)g(z)dydz \\
&\quad= \int((K_h*f)(y)-f(y))(v(y)-\mu_v)\left(\int\gamma(z)K_b(z-y)dz\right)dy \\
&\quad= \int((K_h*f)(y)-f(y))(v(y)-\mu_v)\left(\int\gamma(y+bt)K(t)dt\right)dy \\
&\quad= \int((K_h*f)(y)-f(y))(v(y)-\mu_v)\left(\gamma(y)+\frac{\mu_2(K)}{2}b^2\gamma''(y)\right. \\
&\quad+ \left.\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(y)+O(b^6)\right)dy \\
&\quad= \int\left(\frac{\mu_2(K)}{2}h^2f''(y)+\frac{\mu_4(K)}{24}h^4f^{(4)}(y)+O(h^6)\right)(v(y)-\mu_v) \\
&\quad\cdot \left(\gamma(y)+\frac{\mu_2(K)}{2}b^2\gamma''(y)+\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(y)+O(b^6)\right)dy \\
&\quad= \frac{\mu_2(K)}{2}h^2\int\theta(y)f''(y)dy+\frac{\mu_4(K)}{24}h^4\int\theta(y)f^{(4)}(y)dy \qquad \text{(A.75)} \\
&\quad+ \frac{\mu_2(K)^2}{4}h^2b^2\int f''(y)\gamma''(y)(v(y)-\mu_v)dy \\
&\quad+ \frac{\mu_2(K)\mu_4(K)}{48}h^2b^4\int f''(y)\gamma^{(4)}(y)(v(y)-\mu_v)dy \\
&\quad+ \frac{\mu_2(K)\mu_4(K)}{48}h^4b^2\int f^{(4)}(y)\gamma''(y)(v(y)-\mu_v)dy \\
&\quad+ O(h^6)+O(h^4b^4)+O(h^2b^6).
\end{aligned}$$

The second term in (A.74) is:

$$\begin{aligned}
&E\left[\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\gamma(Y_2)\right] \\
&\quad= E\left[\frac{(K_h*f)(Y_1)-f(Y_1)}{g(Y_1)}(v(Y_1)-\mu_v)\right]E\left[\gamma(Y_2)\right]=0,
\end{aligned}$$

since

$$E\left[\gamma(Y_2)\right] = \int \gamma(y)g(y)dy = \int f(y)(v(y) - \mu_v)dy = 0.$$

Plugging equation (A.75) and (A.71) into (A.73) gives:

$$\begin{aligned}
Cov(\eta_{11}, \tau_{21}) &= \frac{\mu_2(K)}{2}h^2 \int \theta(y)f''(y)dy + \frac{\mu_4(K)}{24}h^4 \int \theta(y)f^{(4)}(y)dy \\
&+ \frac{\mu_2(K)^2}{4}h^2b^2 \left[\int f''(y)\gamma''(y)(v(y) - \mu_v)dy\right. \\
&- \left.\int v''(y)f(y)dy \int \gamma(y)g''(y)dy\right] \\
&+ \frac{\mu_2(K)\mu_4(K)}{48}h^2b^4 \left[\int f''(y)\gamma^{(4)}(y)(v(y) - \mu_v)dy\right. \\
&- \left.\int v''(y)f(y)dy \int \gamma(y)g^{(4)}(y)dy\right] \\
&+ \frac{\mu_2(K)\mu_4(K)}{48}h^4b^2 \left[\int f^{(4)}(y)\gamma''(y)(v(y) - \mu_v)dy\right. \\
&- \left.\int v^{(4)}(y)f(y)dy \int \gamma(y)g''(y)dy\right] \\
&+ O(h^6) + O(h^4b^4) + O(h^2b^6). \qquad (A.76)
\end{aligned}$$

We finally deal with the last term in (A.68):

$$Cov(\eta_{11}, \tau_{11}) = E(\eta_{11}\tau_{11}) - E(\eta_{11})E(\tau_{11})$$

where the first term is

$$\begin{aligned}
E(\eta_{11}\tau_{11}) &= E\left[E(\eta_{11}|Y_1)\tau_{11}\right] \\
&= E\left[\frac{(K_h * f)(Y_1) - f(Y_1)}{g(Y_1)}(v(Y_1) - \mu_v)\varphi(Y_1)(K_b(0) - g(Y_1))\right] \\
&= \int \frac{(K_h * f)(y) - f(y)}{g(y)}(v(y) - \mu_v)\varphi(y)\left(\frac{K(0)}{b} - g(y)\right)g(y)dy \\
&= \int ((K_h * f)(y) - f(y))\rho(y)\left(\frac{K(0)}{b} - g(y)\right)dy \\
&= \int \left(\frac{\mu_2(K)}{2}h^2f''(y) + \frac{\mu_4(K)}{24}h^4f^{(4)}(y) + O(h^6)\right)\rho(y)\left(\frac{K(0)}{b} - g(y)\right)dy \\
&= \frac{\mu_2(K)K(0)}{2}\frac{h^2}{b}\int \rho(y)f''(y)dy - \frac{\mu_2(K)}{2}h^2\int \theta(y)f''(y)dy \\
&+ \frac{\mu_4(K)K(0)}{24}\frac{h^4}{b}\int \rho(y)f^{(4)}(y)dy - \frac{\mu_4(K)}{24}h^4\int \theta(y)f^{(4)}(y)dy + O\left(\frac{h^6}{b}\right),
\end{aligned}$$

and the second one

$$\begin{aligned}
E(\eta_{11})E(\tau_{11}) &= \left(\frac{h^2}{2}\mu_2(K)\int v''(y)f(y)dy + \frac{h^4}{24}\mu_4(K)\int v^{(4)}(y)f(y)dy + O(h^6)\right) \\
&\cdot \left(\frac{K(0)}{b}\int \gamma(y)dy\right) = \frac{\mu_2(K)K(0)}{2}\frac{h^2}{b}\int v''(y)f(y)dy\int \gamma(y)dy \\
&+ \frac{\mu_4(K)K(0)}{24}\frac{h^4}{b}\int v^{(4)}(y)f(y)dy\int \gamma(y)dy + O\left(\frac{h^6}{b}\right).
\end{aligned}$$

Consequently:

$$\begin{aligned}
Cov(\eta_{11}, \tau_{11}) &= \frac{\mu_2(K)K(0)}{2}\frac{h^2}{b}\left[\int \rho(y)f''(y)dy - \int v''(y)f(y)dy\int \gamma(y)dy\right] \\
&- \frac{\mu_2(K)}{2}h^2\int \theta(y)f''(y)dy \\
&+ \frac{\mu_4(K)K(0)}{24}\frac{h^4}{b}\left[\int \rho(y)f^{(4)}(y)dy - \int v^{(4)}(y)f(y)dy\int \gamma(y)dy\right] \\
&- \frac{\mu_4(K)}{24}h^4\int \theta(y)f^{(4)}(y)dy + O\left(\frac{h^6}{b}\right). \qquad \text{(A.77)}
\end{aligned}$$

We now plug (A.72), (A.76) and (A.77) into (A.68) to obtain (4.20). □

Lemma 4.5.10. *The variance of $\widehat{A}$ is*

$$\begin{aligned}
Var\left(\widehat{A}\right) &\simeq D_7\frac{1}{n} + D_8\frac{1}{Nn} - 4D_6\frac{1}{N^2} + D_{26}\frac{1}{N^3} + D_9\frac{1}{Nnh} + D_{44}\frac{1}{N^2b} + D_{24}\frac{1}{N^3b^2} \\
&+ D_{25}\frac{1}{N^3b} - 2D_{39}\frac{h^2}{N^2b} - 2D_{41}\frac{h^4}{N^2b} + D_{10}\frac{h^2}{n} + D_{11}\frac{h^4}{n} + D_{12}\frac{h^4}{N} \\
&+ D_{45}\frac{b^4}{N} - 2D_{35}\frac{h^2b^2}{N} - 2D_{37}\frac{h^4b^2}{N} - 2D_{38}\frac{h^2b^4}{N} + D_{13}\frac{h}{Nn} + D_{14}\frac{h^2}{Nn} \\
&+ D_{15}\frac{h^3}{Nn} + D_{23}\frac{b}{N^2} + 4D_{27}\frac{h^2}{N^2} + 4D_{28}\frac{h^4}{N^2} + 2D_{35}\frac{h^2b^2}{N^2} + O\left(\frac{h^6}{n}\right) \\
&+ O\left(\frac{b^6}{N}\right) + O\left(\frac{h^4b^4}{N}\right) + O\left(\frac{h^4}{Nn}\right) + O\left(\frac{b^2}{N^2}\right) + O\left(\frac{h^6}{N^2b}\right),
\end{aligned}$$

being

$$\begin{aligned}
D_{44} &:= 2\mu_0(K^2)\int \alpha(y)dy, \\
D_{45} &:= \frac{\mu_2(K)^2}{4}\left[\int \delta(y)g''(y)^2dy - 4\left(\int \gamma(y)g''(y)dy\right)^2\right. \\
&+ \left.\int \gamma''(y)^2g(y)dz + 2\int \gamma(y)\gamma''(y)g''(y)dy\right].
\end{aligned}$$

Proof. Consequence of Lemmas 4.5.2, 4.5.4, 4.5.5, 4.5.6, 4.5.7, 4.5.8 and 4.5.9. □

Theorem 4.2.1. *Under the classical conditions on the bandwiths and the sample sizes, i.e. $h \to 0$, $b \to 0$, $nh \to \infty$, $Nb \to \infty$ and $N/n \to \infty$, if Conditions A1, A6 and A7 are fulfilled, then the asymptotic mean squared error of $\hat{\mu}_v$ is*

$$\begin{aligned} AMSE\left(\hat{\mu}_v\right) &= \left(C_1 b^2 + \frac{C_2}{Nb} + C_3 h^2\right)^2 + \frac{C_4}{n} + \frac{C_5}{Nn} + \frac{C_6}{N^2} + \frac{C_7}{Nnh} \\ &+ \frac{C_8}{N^2 b} + \frac{C_9 h^2}{n} + \frac{C_{10} h^2}{N^2 b} + \frac{C_{11} h^4}{N} + \frac{C_{12} h^2 b^2}{N}, \end{aligned} \tag{4.7}$$

where the first three terms come from the squared bias and the rest of them from the variance of the estimator. The constants $C_1, \ldots, C_{12}$ are defined in the sketch of the proofs (Subsection 4.5.1).

Proof. Consequence of Lemmas 4.5.3 and 4.5.10, considering $C_1 := D_2$, $C_2 := D_1$, $C_3 := D_4$, $C_4 := D_7$, $C_5 := D_8$, $C_6 := -4D_6$, $C_7 := D_9$, $C_8 := D_{44}$, $C_9 := D_{10}$, $C_{10} := -2D_{39}$, $C_{11} := D_{12}$ and $C_{12} := -2D_{35}$. □

A.1.4 Proof of Theorem 4.2.2

Using Lemma 4.5.1, in this case $\widehat{A}$ can be expressed as:

$$\widehat{A} = \widehat{A}_1^* + \widehat{A}_2^* - \widehat{A}_3^* - \widehat{A}_4^* + \widehat{A}_5^*,$$

where

$$\begin{aligned} \widehat{A}_1^* &:= \frac{1}{N}\sum_{i=1}^N \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v), \\ \widehat{A}_2^* &:= \frac{1}{N}\sum_{i=1}^N \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v), \\ \widehat{A}_3^* &:= \frac{1}{N}\sum_{i=1}^N \frac{(K_h * f)(Y_i)(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v), \\ \widehat{A}_4^* &:= \frac{1}{N}\sum_{i=1}^N \frac{(\hat{f}_h(Y_i) - (K_h * f)(Y_i))(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v), \\ \widehat{A}_5^* &:= \frac{1}{N}\sum_{i=1}^N \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_b * g)(Y_i)}\right)^2 (v(Y_i) - \mu_v), \end{aligned}$$

being $\widehat{A}_4^*$ and $\widehat{A}_5^*$ negligible terms. Thus we will consider

$$\widehat{A}^* := \widehat{A}_1^* + \widehat{A}_2^* - \widehat{A}_3^*.$$

The proof of Theorem 4.2.2 follows parallel lines to that of Theorem 4.2.1.

Lemma A.1.10. *The expectation and variance of* $\widehat{A}^*$ *are*

$$\begin{aligned} E\left(\widehat{A}^*\right) &= E\left(\widehat{A}_1^*\right) + E\left(\widehat{A}_2^*\right) - E\left(\widehat{A}_3^*\right), && \text{(A.78)} \\ Var\left(\widehat{A}^*\right) &= Var\left(\widehat{A}_1^*\right) + Var\left(\widehat{A}_2^*\right) + Var\left(\widehat{A}_3^*\right) \\ &+ 2Cov\left(\widehat{A}_1^*, \widehat{A}_2^*\right) - 2Cov\left(\widehat{A}_1^*, \widehat{A}_3^*\right) - 2Cov\left(\widehat{A}_2^*, \widehat{A}_3^*\right). && \text{(A.79)} \end{aligned}$$

Proof. We may write some expression for the ratio $\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}$:

$$\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} = \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} \left(\frac{\hat{g}_b(Y_i)}{(K_b * g)(Y_i)} + 1 - \frac{\hat{g}_b(Y_i)}{(K_b * g)(Y_i)} \right),$$

which gives:

$$\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} = \frac{\hat{f}_h(Y_i)}{(K_b * g)(Y_i)} + \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} \left(\frac{(K_b * g)(Y_i) - \hat{g}_b(Y_i)}{(K_b * g)(Y_i)} \right).$$

As a consequence,

$$\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} - \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)} = \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)} + \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} \left(\frac{(K_b * g)(Y_i) - \hat{g}_b(Y_i)}{(K_b * g)(Y_i)} \right).$$

Applying similar techniques once more to the term $\frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}$ in the right-hand side of the previous expression gives:

$$\begin{aligned} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} - \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)} &= \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)} \\ &+ \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)} \left(\frac{(K_b * g)(Y_i) - \hat{g}_b(Y_i)}{(K_b * g)(Y_i)} \right) \\ &+ \left(\frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)} \right) \left(\frac{(K_b * g)(Y_i) - \hat{g}_b(Y_i)}{(K_b * g)(Y_i)} \right) \\ &+ \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} \left(\frac{(K_b * g)(Y_i) - \hat{g}_b(Y_i)}{(K_b * g)(Y_i)} \right)^2 \end{aligned}$$

which can also be expressed as:

$$\begin{aligned} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} - \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)} &= \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)} \\ &- \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)} \left(\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_b * g)(Y_i)} \right) \\ &- \frac{(\hat{f}_h(Y_i) - (K_h * f)(Y_i))(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2} \\ &+ \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)} \left(\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_b * g)(Y_i)} \right)^2. && \text{(A.80)} \end{aligned}$$

Using (A.80) in the definition of $\widehat{A}$ (expression (4.11)) gives:

$$\widehat{A} = \widehat{A}_1^* + \widehat{A}_2^* - \widehat{A}_3^* - \widehat{A}_4^* + \widehat{A}_5^*,$$

where

$$\widehat{A}_1^* := \frac{1}{N}\sum_{i=1}^{N} \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v), \tag{A.81}$$

$$\widehat{A}_2^* := \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v), \tag{A.82}$$

$$\widehat{A}_3^* := \frac{1}{N}\sum_{i=1}^{N} \frac{(K_h * f)(Y_i)(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v), \tag{A.83}$$

$$\widehat{A}_4^* := \frac{1}{N}\sum_{i=1}^{N} \frac{(\hat{f}_h(Y_i) - (K_h * f)(Y_i))(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v), \tag{A.84}$$

$$\widehat{A}_5^* := \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{f}_h(Y_i)}{\hat{g}_b(Y_i)}\left(\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_b * g)(Y_i)}\right)^2 (v(Y_i) - \mu_v). \tag{A.85}$$

Since the terms $\widehat{A}_4^*$ and $\widehat{A}_5^*$ have some factors of quadratic nature inside the sum (i.e. $(\hat{f}_h(Y_i) - (K_h * f)(Y_i))(\hat{g}_b(Y_i) - (K_b * g)(Y_i))$ and $(\hat{g}_b(Y_i) - (K_b * g)(Y_i))^2$) it is expected that applying results of the type by Mack & Silverman (1982) one could prove negligibility of the terms (A.84) and (A.85).

Thus we will consider

$$\widehat{A}^* := \widehat{A}_1^* + \widehat{A}_2^* - \widehat{A}_3^*.$$

Since we want to obtain the mean and variance of $\widehat{A}^*$, we proceed as shown in (A.78) and (A.79). □

We now consider the terms in the right-hand side of (A.78).

Lemma 4.5.11. *The expectation of* $\widehat{A}^*$ *is*

$$E(\widehat{A}^*) \quad = \quad D_1^* + D_2^* \frac{1}{N}, \tag{4.21}$$

where

$$\begin{aligned} D_1^* &:= \int \gamma^*(y) g(y) dy, \\ D_2^* &:= D_1^* - \frac{K(0)}{b} \int \varphi^*(y) g(y) dy, \end{aligned}$$

with

$$\begin{aligned}\gamma^*(y) &:= \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v), \\ \varphi^*(y) &:= \frac{(K_h * f)(y)}{(K_b * g)(y)^2}(v(y) - \mu_v).\end{aligned}$$

Proof.

$$\begin{aligned}E\left(\widehat{A}_1^*\right) &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{(K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)\right] = E\left[\frac{(K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)\right] \\ &= \int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy = D_1^*. \qquad \text{(A.86)}\end{aligned}$$

Since the random variables

$$\eta_i^* := \Psi\left(X_1, \ldots, X_n, Y_i\right) := \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v),$$

for $i = 1, 2, \ldots, N$, are identical distributed (but not independent!!) then

$$E\left(\widehat{A}_2^*\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(\eta_i^*\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(\eta_1^*\right) = E\left(\eta_1^*\right).$$

But

$$\begin{aligned}E\left(\eta_1^*\right) &= E\left[E\left(\eta_1^*|Y_1\right)\right] = E\left[\frac{E(\hat{f}_h(Y_1)|Y_1) - (K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)\right] \\ &= E\left[\frac{(K_h * f)(Y_1) - (K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)\right] = 0.\end{aligned}$$

As a consequence,

$$E\left(\widehat{A}_2^*\right) = E\left(\eta_1^*\right) = 0. \qquad \text{(A.87)}$$

In view of (A.78),

$$\begin{aligned}E(\widehat{A}_3^*) &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{(K_h * f)(Y_i)(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v)\right] \\ &= E\left[\frac{(K_h * f)(Y_1)(\hat{g}_b(Y_1) - (K_b * g)(Y_1))}{(K_b * g)(Y_1)^2}(v(Y_1) - \mu_v)\right] \\ &= E\left[E\left(\frac{(K_h * f)(Y_1)(\hat{g}_b(Y_1) - (K_b * g)(Y_1))}{(K_b * g)(Y_1)^2}(v(Y_1) - \mu_v)|Y_1\right)\right] \\ &= E\left[\frac{(K_h * f)(Y_1)(E\left[\hat{g}_b(Y_1)|Y_1\right] - (K_b * g)(Y_1))}{(K_b * g)(Y_1)^2}(v(Y_1) - \mu_v)\right].\end{aligned}$$

But since

$$\begin{aligned}
\hat{g}_b(Y_1) &= \frac{1}{N}\sum_{i=1}^{N} K_b(Y_1 - Y_i) = \frac{1}{N}\left(K_b(0) + \sum_{i=2}^{N} K_b(Y_1 - Y_i)\right) \\
&= \frac{K(0)}{Nb} + \frac{N-1}{N}\frac{1}{N-1}\sum_{i=2}^{N} K_b(Y_1 - Y_i) = \frac{K(0)}{Nb} + \frac{N-1}{N}\hat{g}_b^{(-1)}(Y_1),
\end{aligned}$$

we get

$$E\left[\hat{g}_b(Y_1)|Y_1\right] = \frac{K(0)}{Nb} + \frac{N-1}{N}E\left[\hat{g}_b^{(-1)}(Y_1)|Y_1\right] = \frac{K(0)}{Nb} + \frac{N-1}{N}(K_b * g)(Y_1).$$

Using this expression above we have:

$$\begin{aligned}
E(\widehat{A}_3^*) &= E\left[\frac{(K_h * f)(Y_1)(v(Y_1) - \mu_v)}{(K_b * g)(Y_1)^2}\right. \\
&\cdot \left.\left(\frac{K(0)}{Nb} + \frac{N-1}{N}(K_b * g)(Y_1) - (K_b * g)(Y_1)\right)\right] \\
&= E\left[\frac{(K_h * f)(Y_1)(v(Y_1) - \mu_v)}{(K_b * g)(Y_1)^2}\left(\frac{K(0)}{Nb} - \frac{1}{N}(K_b * g)(Y_1)\right)\right] \\
&= \int \frac{(K_h * f)(y)(v(y) - \mu_v)}{(K_b * g)(y)^2}\left(\frac{K(0)}{Nb} - \frac{1}{N}(K_b * g)(y)\right) g(y)dy \\
&= \frac{K(0)}{Nb}\int \frac{(K_h * f)(y)}{(K_b * g)(y)^2}(v(y) - \mu_v)g(y)dy \\
&- \frac{1}{N}\int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy = -D_2^*\frac{1}{N}. \qquad (A.88)
\end{aligned}$$

From (A.86), (A.87) and (A.88), we obtain (4.21). □

We now consider the terms in the right-hand side of (A.79):

Lemma 4.5.12. *The variance of $\widehat{A}_1^*$ is*

$$Var\left(\widehat{A}_1^*\right) = \frac{D_3^*}{N}, \qquad (4.22)$$

where

$$D_3^* := \int \alpha^*(y)g(y)dy - D_1^{*2},$$

with

$$\alpha^*(y) := \frac{(K_h * f)(y)^2}{(K_b * g)(y)^2}(v(y) - \mu_v)^2.$$

Proof.

$$
\begin{aligned}
Var\left(\widehat{A}_1^*\right) &= \frac{1}{N^2}\sum_{i=1}^{N} Var\left[\frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i)-\mu_v)\right]\\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var\left[\frac{(K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1)-\mu_v)\right]\\
&= \frac{1}{N} Var\left[\frac{(K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1)-\mu_v)\right]\\
&= \frac{1}{N}\left\{E\left[\frac{(K_h * f)(Y_1)^2}{(K_b * g)(Y_1)^2}(v(Y_1)-\mu_v)^2\right]\right.\\
&- \left.\left(E\left[\frac{(K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1)-\mu_v)\right]\right)^2\right\}\\
&= \frac{1}{N}\left\{\int \frac{(K_h * f)(y)^2}{(K_b * g)(y)^2}(v(y)-\mu_v)^2 g(y)dy\right.\\
&- \left.\left(\int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y)-\mu_v)g(y)dy\right)^2\right\}.
\end{aligned}
$$

Shortening the previous expression, we obtain (4.22). □

Lemma 4.5.13. *The variance of $\widehat{A}_2^*$ is*

$$
Var\left(\widehat{A}_2^*\right) = D_4^*\frac{1}{n} + D_5^*\frac{1}{Nn}, \tag{4.23}
$$

where

$$
\begin{aligned}
D_4^* &:= \int\left(\int \frac{K_h(y-z)}{(K_b * g)(y)}(v(y)-\mu_v)g(y)dy - \int \gamma^*(y)g(y)dy\right)^2 f(z)dz,\\
D_5^* &:= \int \frac{((K_h)^2 * f)(y)}{(K_b * g)(y)^2}(v(y)-\mu_v)^2 g(y)dy - \int \alpha^*(y)g(y)dy - D_4^*.
\end{aligned}
$$

Proof. In order to compute the variance of $\widehat{A}_2^*$, let us rewrite the terms η_i^* as follows:

$$
\begin{aligned}
\eta_i^* &= \frac{\hat{f}_h(Y_i) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i)-\mu_v)\\
&= \frac{\frac{1}{n}\sum_{j=1}^{n} K_h(Y_i - X_j) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i)-\mu_v)\\
&= \frac{1}{n}\sum_{j=1}^{n}\frac{K_h(Y_i - X_j) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i)-\mu_v) = \frac{1}{n}\sum_{j=1}^{n}\eta_{ij}^*,
\end{aligned}
$$

with

$$
\eta_{ij}^* := \frac{K_h(Y_i - X_j) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i)-\mu_v), \tag{A.89}
$$

for $i = 1, \dots, N; j = 1, \dots, n$.

Now, using (A.89), $\widehat{A}_2^*$ can be written as

$$\widehat{A}_2^* = \frac{1}{Nn} \sum_{i=1}^{N} \sum_{j=1}^{n} \eta_{ij}^*. \tag{A.90}$$

Thus,

$$Var(\widehat{A}_2^*) = \frac{1}{N^2 n^2} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{N} \sum_{l=1}^{n} Cov(\eta_{ij}^*, \eta_{kl}^*). \tag{A.91}$$

Collecting all the equal terms in (A.91) gives:

$$\begin{aligned} Var(\widehat{A}_2^*) &= \frac{1}{N^2 n^2} \left[Nn(n-1) Cov(\eta_{11}^*, \eta_{12}^*) \right. \\ &+ \left. nN(N-1) Cov(\eta_{11}^*, \eta_{21}^*) + NnVar(\eta_{11}^*) \right]. \end{aligned} \tag{A.92}$$

We now work these covariance terms out:

$$Cov(\eta_{11}^*, \eta_{12}^*) = Cov(E(\eta_{11}^* | Y_1), E(\eta_{12}^* | Y_1)) + E\left[Cov(\eta_{11}^*, \eta_{12}^* | Y_1) \right].$$

But $Cov(\eta_{11}^*, \eta_{12}^* | Y_1) = 0$, since

$$\eta_{11}^* = \frac{K_h(Y_1 - X_1) - (K_h * f)(Y_1)}{(K_b * g)(Y_1)} (v(Y_1) - \mu_v)$$

and

$$\eta_{12}^* = \frac{K_h(Y_1 - X_2) - (K_h * f)(Y_1)}{(K_b * g)(Y_1)} (v(Y_1) - \mu_v)$$

are conditionally independent on Y_1 (because X_1 and X_2 are independent).

On the other hand,

$$\begin{aligned} E(\eta_{11}^* | Y_1) &= E(\eta_{12}^* | Y_1) = \frac{E\left[K_h(Y_1 - X_1) | Y_1 \right] - (K_h * f)(Y_1)}{(K_b * g)(Y_1)} (v(Y_1) - \mu_v) \\ &= \frac{E\left[K_h(X_1 - Y_1) | Y_1 \right] - (K_h * f)(Y_1)}{(K_b * g)(Y_1)} (v(Y_1) - \mu_v) \\ &= \frac{(K_h * f)(Y_1) - (K_h * f)(Y_1)}{(K_b * g)(Y_1)} (v(Y_1) - \mu_v) = 0. \end{aligned}$$

So,

$$Cov(\eta_{11}^*, \eta_{12}^*) = Cov(E(\eta_{11}^* | Y_1), E(\eta_{12}^* | Y_1)) = Var(E(\eta_{11}^* | Y_1)) = 0. \tag{A.93}$$

Now we deal with the term $Cov(\eta_{11}^*, \eta_{21}^*)$ in (A.92):

$$\begin{aligned} Cov(\eta_{11}^*, \eta_{21}^*) &= Cov(E(\eta_{11}^*|X_1), E(\eta_{21}^*|X_1)) + E[Cov(\eta_{11}^*, \eta_{21}^*|X_1)] \\ &= Cov(E(\eta_{11}^*|X_1), E(\eta_{21}^*|X_1)) = E(E(\eta_{11}^*|X_1) E(\eta_{21}^*|X_1)) \\ &- E(E(\eta_{11}^*|X_1)) E(E(\eta_{21}^*|X_1)) = E(E(\eta_{11}^*|X_1) E(\eta_{21}^*|X_1)) \\ &- E(\eta_{11}^*) E(\eta_{21}^*) = E(E(\eta_{11}^*|X_1) E(\eta_{21}^*|X_1)), \end{aligned}$$

since $Cov(\eta_{11}^*, \eta_{21}^*|X_1) = 0$ because

$$\eta_{11}^* = \frac{K_h(Y_1 - X_1) - (K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)$$

and

$$\eta_{21}^* = \frac{K_h(Y_2 - X_1) - (K_h * f)(Y_2)}{(K_b * g)(Y_2)}(v(Y_2) - \mu_v)$$

are conditionally independent given X_1 (since Y_1 and Y_2 are independent) and

$$E(\eta_{21}^*) = E(E(\eta_{21}^*|Y_2)) = 0.$$

But

$$E(\eta_{11}^*|X_1) = E(\eta_{21}^*|X_1) = \int \frac{K_h(y - X_1) - (K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy$$

then

$$\begin{aligned} Cov(\eta_{11}^*, \eta_{21}^*) &= E\left(\left(\int \frac{K_h(y - X_1) - (K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy\right)^2\right) \\ &= \int \left(\int \frac{K_h(y - z) - (K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy\right)^2 f(z)dz \\ &= \int \left(\int \frac{K_h(y - z)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy - \int \gamma^*(y)g(y)dy\right)^2 f(z)dz. \quad \text{(A.94)} \end{aligned}$$

We now examine the term $Var(\eta_{11}^*)$ in (A.92):

$$Var(\eta_{11}^*) = E\left(\eta_{11}^{*2}\right) + E(\eta_{11}^*)^2 = E\left(\eta_{11}^{*2}\right) = E\left[E(\eta_{11}^{*2}|Y_1)\right],$$

since $E(\eta_{11}^*) = E[E(\eta_{11}^*|Y_1)] = 0$.

So,

$$\begin{aligned} Var(\eta_{11}^*) &= E\left[E\left(\eta_{11}^{*2}|Y_1\right)\right] \\ &= E\left[E\left(\left(\frac{K_h(Y_1 - X_1) - (K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)\right)^2 |Y_1\right)\right] \\ &= \int \frac{((K_h)^2 * f)(y) - (K_h * f)(y)^2}{(K_b * g)(y)^2}(v(y) - \mu_v)^2 g(y)dy \\ &= \int \frac{((K_h)^2 * f)(y)}{(K_b * g)(y)^2}(v(y) - \mu_v)^2 g(y)dy - \int \alpha^*(y)g(y)dy. \quad \text{(A.95)} \end{aligned}$$

Now, using (A.93), (A.94) and (A.95) in (A.92) gives:

$$\begin{aligned}
Var(\widehat{A}_2^*) =& \frac{N-1}{Nn}\left[\int\left(\int\frac{K_h(y-z)}{(K_b*g)(y)}(v(y)-\mu_v)g(y)dy-\int\gamma^*(y)g(y)dy\right)^2 f(z)dz\right]\\
+ & \frac{1}{Nn}\left[\int\frac{((K_h)^2*f)(y)}{(K_b*g)(y)^2}(v(y)-\mu_v)^2 g(y)dy-\int\alpha^*(y)g(y)dy\right]\\
= & \frac{1}{n}\left[\int\left(\int\frac{K_h(y-z)}{(K_b*g)(y)}(v(y)-\mu_v)g(y)dy-\int\gamma^*(y)g(y)dy\right)^2 f(z)dz\right]\\
+ & \frac{1}{Nn}\left[\int\frac{((K_h)^2*f)(y)}{(K_b*g)(y)^2}(v(y)-\mu_v)^2 g(y)dy-\int\alpha^*(y)g(y)dy\right.\\
- & \left.\int\left(\int\frac{K_h(y-z)}{(K_b*g)(y)}(v(y)-\mu_v)g(y)dy-\int\gamma^*(y)g(y)dy\right)^2 f(z)dz\right]. \qquad \text{(A.96)}
\end{aligned}$$

In order to shorten (A.96), we obtain (4.23).

□

Lemma 4.5.14. *The variance of $\widehat{A}_3^*$ is*

$$Var\left(\widehat{A}_3^*\right) = D_6^*\frac{1}{N}+D_7^*\frac{1}{N^2}+D_8^*\frac{1}{N^3}, \qquad (4.24)$$

where

$$D_6^* := \int\left[\int\varphi^*(y)K_b(y-z)g(y)dy-\int\gamma^*(y)g(y)dy\right]^2 g(z)dz,$$

$$\begin{aligned}
D_7^* := & \frac{2K(0)}{b}\left[\int\int\varphi^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz\right.\\
- & \left.\int\gamma^*(y)g(y)dy\int\varphi^*(y)g(y)dy\right]\\
- & 4\int\int\gamma^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz+3\left(\int\gamma^*(y)g(y)dy\right)^2\\
- & \int\alpha^*(y)g(y)dy+\int\varphi^*(y)^2((K_b)^2*g)(y)g(y)dy\\
+ & \int\int\varphi^*(y)\varphi^*(z)K_b(y-z)^2 g(y)g(z)dydz\\
- & 3\int\left[\int\varphi^*(y)K_b(y-z)g(y)dy-\int\gamma^*(y)g(y)dy\right]^2 g(z)dz,
\end{aligned}$$

$$\begin{aligned}
D_8^* &:= \frac{K(0)^2}{b^2}\left[\int \varphi^*(y)^2 g(y)dy - \left(\int \varphi^*(y)g(y)dy\right)^2\right] \\
&+ \frac{2K(0)}{b}\left[-\int\int \varphi^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz\right. \\
&+ \left. 2\int \gamma^*(y)g(y)dy \int \varphi^*(y)g(y)dy - \int \delta^*(y)g(y)dy\right] \\
&+ 4\int\int \gamma^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz - 4\left(\int \gamma^*(y)g(y)dy\right)^2 \\
&+ 2\int \alpha^*(y)g(y)dy - \int \varphi^*(y)^2((K_b)^2 * g)(y)g(y)dy \\
&- \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)^2 g(y)g(z)dydz \\
&+ 2\int\left[\int \varphi^*(y)K_b(y-z)g(y)dy - \int \gamma^*(y)g(y)dy\right]^2 g(z)dz.
\end{aligned}$$

Proof. To deal with the variance of $\widehat{A}_3^*$ we first consider

$$\tau_i^* := \frac{(K_h * f)(Y_i)(\hat{g}_b(Y_i) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v), \quad i = 1, \ldots, N$$

and write

$$\hat{g}_b(Y_i) = \frac{1}{N}\sum_{j=1}^N K_b(Y_i - Y_j).$$

As a consequence,

$$\tau_i^* = \frac{(K_h * f)(Y_i)\left[\frac{1}{N}\sum_{j=1}^N K_b(Y_i - Y_j) - (K_b * g)(Y_i)\right]}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v) = \frac{1}{N}\sum_{j=1}^N \tau_{ij}^*,$$

where

$$\tau_{ij}^* := \frac{(K_h * f)(Y_i)(K_b(Y_i - Y_j) - (K_b * g)(Y_i))}{(K_b * g)(Y_i)^2}(v(Y_i) - \mu_v),$$

for $i, j = 1, \ldots, N$. Then

$$\widehat{A}_3^* = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \tau_{ij}^*. \tag{A.97}$$

To compute the variance of $\widehat{A}_3^*$ we consider

$$\begin{aligned}
Var(\widehat{A}_3^*) &= Cov(\widehat{A}_3^*, \widehat{A}_3^*) = Cov\left(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \tau_{ij}^*, \frac{1}{N^2}\sum_{k=1}^N\sum_{l=1}^N \tau_{kl}^*\right) \\
&= \frac{1}{N^4}\sum_{i,j,k,l=1}^N Cov(\tau_{ij}^*, \tau_{kl}^*).
\end{aligned}$$

Thus, the variance of $\widehat{A}_3$ can be written as:

$$\begin{aligned} Var(\widehat{A}_3^*) &= \frac{(N-1)(N-2)}{N^3} Cov(\tau_{12}^*, \tau_{13}^*) + \frac{2(N-1)(N-2)}{N^3} Cov(\tau_{12}^*, \tau_{31}^*) \\ &+ \frac{(N-1)(N-2)}{N^3} Cov(\tau_{12}^*, \tau_{32}^*) + \frac{N-1}{N^3} Var(\tau_{12}^*) \\ &+ \frac{N-1}{N^3} Cov(\tau_{12}^*, \tau_{21}^*) + \frac{2(N-1)}{N^3} Cov(\tau_{12}^*, \tau_{11}^*) \\ &+ \frac{2(N-1)}{N^3} Cov(\tau_{12}^*, \tau_{22}^*) + \frac{1}{N^3} Var(\tau_{11}^*). \end{aligned} \tag{A.98}$$

Let's deal with every one of these terms; but first, in order to save space, let us write τ_{ij}^* in a more compact form:

$$\tau_{ij}^* = \varphi^*(Y_i)(K_b(Y_i - Y_j) - (K_b * g)(Y_i)), \quad i, j = 1, \ldots, N.$$

Let us now study the terms in (A.98):

$$Cov(\tau_{12}^*, \tau_{13}^*) = E\left[Cov(\tau_{12}^*, \tau_{13}^*|Y_1)\right] + Cov\left(E\left(\tau_{12}^*|Y_1\right), E\left(\tau_{13}^*|Y_1\right)\right) = Var\left(E\left(\tau_{12}^*|Y_1\right)\right)$$

because

$$\tau_{12}^* = \varphi^*(Y_1)(K_b(Y_1 - Y_2) - (K_b * g)(Y_1))$$

and

$$\tau_{13}^* = \varphi^*(Y_1)(K_b(Y_1 - Y_3) - (K_b * g)(Y_1))$$

are conditionally independent given Y_1 (so $Cov(\tau_{12}^*, \tau_{13}^*|Y_1) = 0$). But

$$\begin{aligned} E\left(\tau_{12}^*|Y_1\right) &= E\left(\tau_{13}^*|Y_1\right) = \varphi^*(Y_1)\left[E\left[K_b(Y_1 - Y_2)|Y_1\right] - (K_b * g)(Y_1)\right] \\ &= \varphi^*(Y_1)\left[(K_b * g)(Y_1) - (K_b * g)(Y_1)\right] = 0, \end{aligned}$$

then

$$Cov(\tau_{12}^*, \tau_{13}^*) = 0. \tag{A.99}$$

Now consider

$$\begin{aligned} Cov(\tau_{12}^*, \tau_{31}^*) = E\left[Cov(\tau_{12}^*, \tau_{31}^*|Y_1)\right] + Cov\left(E\left(\tau_{12}^*|Y_1\right), E\left(\tau_{31}^*|Y_1\right)\right) \\ = Cov\left(E\left(\tau_{12}^*|Y_1\right), E\left(\tau_{31}^*|Y_1\right)\right) \\ = E\left[E\left(\tau_{12}^*|Y_1\right) E\left(\tau_{31}^*|Y_1\right)\right] - E\left[E\left(\tau_{12}^*|Y_1\right)\right] E\left[E\left(\tau_{31}^*|Y_1\right)\right] = 0. \end{aligned} \tag{A.100}$$

since $E\left(\tau_{12}^*|Y_1\right) = 0$.

We now deal with the covariance:

$$\begin{aligned}
Cov(\tau_{12}^*, \tau_{32}^*) &= E\left[Cov(\tau_{12}^*, \tau_{32}^*|Y_2)\right] + Cov\left(E\left(\tau_{12}^*|Y_2\right), E\left(\tau_{32}^*|Y_2\right)\right) \\
&= \quad Cov\left(E\left(\tau_{12}^*|Y_2\right), E\left(\tau_{32}^*|Y_2\right)\right) E\left[E\left(\tau_{12}^*|Y_2\right) E\left(\tau_{32}^*|Y_2\right)\right] \\
&- \quad E\left[E\left(\tau_{12}^*|Y_2\right)\right] E\left[E\left(\tau_{32}^*|Y_2\right)\right] = E\left[E\left(\tau_{12}^*|Y_2\right) E\left(\tau_{32}^*|Y_2\right)\right] \\
&- \quad E(\tau_{12}^*) E(\tau_{32}^*) = E\left[E\left(\tau_{12}^*|Y_2\right) E\left(\tau_{32}^*|Y_2\right)\right] = E\left[\left[E\left(\tau_{12}^*|Y_2\right)\right]^2\right]
\end{aligned}$$

since $E(\tau_{12}^*) = E\left[E\left(\tau_{12}^*|Y_1\right)\right] = 0$.

But

$$\begin{aligned}
E\left(\tau_{12}^*|Y_2\right) &= E\left[\varphi^*(Y_1)(K_b(Y_1 - Y_2) - (K_b * g)(Y_1))\right] \\
&= \int \varphi^*(y)\left[K_b(y - Y_2) - (K_b * g)(y)\right] g(y)dy
\end{aligned}$$

and consequently

$$\begin{aligned}
Cov(\tau_{12}^*, \tau_{32}^*) &= E\left[\left[E\left(\tau_{12}^*|Y_2\right)\right]^2\right] \\
&= \quad E\left[\left(\int \varphi^*(y)\left[K_b(y - Y_2) - (K_b * g)(y)\right] g(y)dy\right)^2\right] \\
&= \quad \int \left[\int \varphi^*(y)\left(K_b(y - z) - (K_b * g)(y)\right) g(y)dy\right]^2 g(z)dz \\
&= \quad \int \left[\int \varphi^*(y) K_b(y - z) g(y)dy - \int \gamma^*(y) g(y)dy\right]^2 g(z)dz. \qquad \text{(A.101)}
\end{aligned}$$

Let's deal now with $Var(\tau_{12}^*)$:

$$\begin{aligned}
Var(\tau_{12}^*) &= E(\tau_{12}^{*2}) - \left[E(\tau_{12}^*)\right]^2 = E(\tau_{12}^{*2}) = E\left[E(\eta_{12}^{*2}|Y_1)\right] \\
&= \quad E\left[E\left(\left[\varphi^*(Y_1)(K_b(Y_1 - Y_2) - (K_b * g)(Y_1))\right]^2 |Y_1\right)\right] \\
&= \quad E\left[\varphi^*(Y_1)^2\left(E\left[K_b(Y_1 - Y_2)^2|Y_1\right]\right.\right. \\
&- \quad \left.\left.2(K_b * g)(Y_1) E\left[K_b(Y_1 - Y_2)|Y_1\right] + (K_b * g)(Y_1)^2\right)\right] \\
&= \quad E\left[\varphi^*(Y_1)^2\left(((K_b)^2 * g)(Y_1) - 2(K_b * g)(Y_1)^2 + (K_b * g)(Y_1)^2\right)\right] \\
&= \quad E\left[\varphi^*(Y_1)^2\left(((K_b)^2 * g)(Y_1) - (K_b * g)(Y_1)^2\right)\right] \\
&= \quad \int \varphi^*(y)^2\left(((K_b)^2 * g)(y) - (K_b * g)(y)^2\right) g(y)dy \\
&= \quad \int \varphi^*(y)^2((K_b)^2 * g)(y) g(y)dy - \int \alpha^*(y) g(y)dy. \qquad \text{(A.102)}
\end{aligned}$$

Let us now consider $Cov(\tau_{12}^*, \tau_{21}^*)$:

$$\begin{aligned}
Cov(\tau_{12}^*, \tau_{21}^*) &= E(\tau_{12}^*\tau_{21}^*) - E(\tau_{12}^*)E(\tau_{21}^*) = E(\tau_{12}^*\tau_{21}^*) - [E(\tau_{12}^*)]^2 \\
&= E(\tau_{12}^*\tau_{21}^*) = E\left[\varphi^*(Y_1)\left(K_b(Y_1 - Y_2) - (K_b * g)(Y_1)\right)\right. \\
&\cdot \left.\varphi^*(Y_2)\left(K_b(Y_2 - Y_1) - (K_b * g)(Y_2)\right)\right] \\
&= \int\int \varphi^*(y)(K_b(y-z) - (K_b * g)(y)) \\
&\cdot \varphi^*(z)\left(K_b(z-y) - (K_b * g)(z)\right) g(y)g(z)dydz \\
&= \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)^2 g(y)g(z)dydz \\
&- \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)(K_b * g)(z)g(y)g(z)dydz \\
&- \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)(K_b * g)(y)g(y)g(z)dydz \\
&+ \int\int \varphi^*(y)(K_b * g)(y)\varphi^*(z)(K_b * g)(z)g(y)g(z)dydz \\
&= \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)^2 g(y)g(z)dydz \\
&+ \left(\int \varphi^*(y)(K_b * g)(y)g(y)dy\right)^2 \\
&- 2\int\int \varphi^*(y)\varphi^*(z)K_b(y-z)(K_b * g)(y)g(y)g(z)dydz \\
&= \int\int \varphi^*(y)\varphi^*(z)K_b(y-z)^2 g(y)g(z)dydz \\
&- 2\int\int \gamma^*(y)\varphi^*(z)K_b(y-z)g(y)g(z)dydz \\
&+ \left(\int \gamma^*(y)g(y)dy\right)^2. \qquad \text{(A.103)}
\end{aligned}$$

Let's deal now with the term:

$$\begin{aligned}
Cov(\tau_{12}^*, \tau_{11}^*) &= Cov\left(E\left(\tau_{12}^*|Y_1\right), E\left(\tau_{11}^*|Y_1\right)\right) + E\left[Cov(\tau_{12}^*, \tau_{11}^*|Y_1)\right] \\
&= Cov\left(E\left(\tau_{12}^*|Y_1\right), E\left(\tau_{11}^*|Y_1\right)\right) = 0, \qquad \text{(A.104)}
\end{aligned}$$

since $E\left(\tau_{12}^*|Y_1\right) = 0$.

We now compute the covariance $Cov(\tau_{12}^*, \tau_{22}^*)$:

$$\begin{aligned}
Cov(\tau_{12}^*, \tau_{22}^*) &= Cov\left(E\left(\tau_{12}^*|Y_2\right), E\left(\tau_{22}^*|Y_2\right)\right) + E\left[Cov(\tau_{12}^*, \tau_{22}^*|Y_2)\right] \\
&= Cov\left(E\left(\tau_{12}^*|Y_2\right), E\left(\tau_{22}^*|Y_2\right)\right) = E\left(E\left(\tau_{12}^*|Y_2\right)\tau_{22}^*\right) - E\left[E\left(\tau_{12}^*|Y_2\right)\right]E\left(\tau_{22}^*\right) \\
&= E\left(E\left(\varphi^*(Y_1)(K_b(Y_1 - Y_2) - (K_b * g)(Y_1))|Y_2\right)\tau_{22}^*\right),
\end{aligned}$$

since $Cov(\tau_{12}^*, \tau_{22}^*|Y_2) = 0$ (because τ_{22}^* is a measurable function of Y_2) and since

$$E\left[E\left(\tau_{12}^*|Y_2\right)\right] = E(\tau_{12}^*) = E\left[E\left(\tau_{12}^*|Y_1\right)\right] = 0.$$

On the other hand,

$$\begin{aligned} & E\left[\varphi^*(Y_1)(K_b(Y_1 - Y_2) - (K_b * g)(Y_1))|Y_2\right] \\ = & \int \varphi^*(y)(K_b(y - Y_2) - (K_b * g)(y))g(y)dy \end{aligned}$$

Using the previous expressions and $\tau_{22}^* = \varphi^*(Y_2)\left(\frac{K(0)}{b} - (K_b * g)(Y_2)\right)$, we have:

$$\begin{aligned} Cov(\tau_{12}^*, \tau_{22}^*) &= E\left[\int \varphi^*(y)(K_b(y - Y_2) - (K_b * g)(y))g(y)dy\right. \\ &\cdot \left.\varphi^*(Y_2)\left(\frac{K(0)}{b} - (K_b * g)(Y_2)\right)\right] \\ &= \int\int \varphi^*(y)(K_b(y - z) - (K_b * g)(y)) \\ &\cdot \varphi^*(z)\left(\frac{K(0)}{b} - (K_b * g)(z)\right)g(y)g(z)dydz \\ &= \frac{K(0)}{b}\int\int \varphi^*(y)\varphi^*(z)K_b(y - z)g(y)g(z)dydz \\ &- \int\int \varphi^*(y)\varphi^*(z)K_b(y - z)(K_b * g)(z)g(y)g(z)dydz \\ &- \frac{K(0)}{b}\int\int \varphi^*(y)\varphi^*(z)(K_b * g)(y)g(y)g(z)dydz \\ &+ \left(\int \varphi^*(y)(K_b * g)(y)g(y)dy\right)^2 \\ &= \frac{K(0)}{b}\int\int \varphi^*(y)\varphi^*(z)K_b(y - z)g(y)g(z)dydz \\ &+ \left(\int \gamma^*(y)g(y)dy\right)^2 - \int\int \varphi^*(y)\gamma^*(z)K_b(y - z)g(y)g(z)dydz \\ &- \frac{K(0)}{b}\int\int \gamma^*(y)\varphi^*(z)g(y)g(z)dydz. \end{aligned} \tag{A.105}$$

Finally, we study the term $Var(\tau_{11}^*)$:

$$Var(\tau_{11}^*) = E(\tau_{11}^{*2}) - E(\tau_{11}^*)^2,$$

where the first term $E(\tau_{11}^{*2})$ is:

$$\begin{aligned}
E(\tau_{11}^{*2}) &= E\left[\varphi^*(Y_1)^2\left(\frac{K(0)}{b}-(K_b*g)(Y_1)\right)^2\right] \\
&= \frac{K(0)^2}{b^2}E\left[\varphi^*(Y_1)^2\right]-2\frac{K(0)}{b}E\left[\varphi^*(Y_1)^2(K_b*g)(Y_1)\right] \\
&+ E\left[\varphi^*(Y_1)^2(K_b*g)(Y_1)^2\right]=\frac{K(0)^2}{b^2}\int\varphi^*(y)^2g(y)dy \\
&- \frac{2K(0)}{b}\int\delta^*(y)g(y)dy+\int\alpha^*(y)g(y)dy, \qquad \text{(A.106)}
\end{aligned}$$

where

$$\delta^*(y):=\frac{(K_h*f)(y)^2}{(K_b*g)(y)^3}(v(y)-\mu_v)^2.$$

And the last term:

$$\begin{aligned}
E(\tau_{11}^{*})^2 &= \left[E\left(\varphi^*(Y_1)\left(\frac{K(0)}{b}-(K_b*g)(Y_1)\right)\right)\right]^2 \\
&= \left(\int\varphi^*(y)\left(\frac{K(0)}{b}-(K_b*g)(y)\right)g(y)dy\right)^2 \\
&= \frac{K(0)^2}{b^2}\left(\int\varphi^*(y)g(y)dy\right)^2-2\frac{K(0)}{b}\left(\int\varphi^*(y)g(y)dy\right) \\
&\cdot \left(\int\gamma^*(y)g(y)dy\right)+\left(\int\gamma^*(y)g(y)dy\right)^2. \qquad \text{(A.107)}
\end{aligned}$$

Using (A.106) and (A.107) we obtain

$$\begin{aligned}
Var(\tau_{11}^*) &= \frac{K(0)^2}{b^2}\left[\int\varphi^*(y)^2g(y)dy-\left(\int\varphi^*(y)g(y)dy\right)^2\right] \\
&- \frac{2K(0)}{b}\left[\int\varphi^*(y)^2(K_b*g)(y)g(y)dy\right. \\
&- \left.\left(\int\varphi^*(y)g(y)dy\right)\left(\int\varphi^*(y)(K_b*g)(y)g(y)dy\right)\right] \\
&+ \left[\int\varphi^*(y)^2(K_b*g)(y)^2g(y)dy-\left(\int\varphi^*(y)(K_b*g)(y)g(y)dy\right)^2\right] \\
&= \frac{K(0)^2}{b^2}\left[\int\varphi^*(y)^2g(y)dy-\left(\int\varphi^*(y)g(y)dy\right)^2\right] \\
&- \frac{2K(0)}{b}\left[\int\delta^*(y)g(y)dy-\left(\int\varphi^*(y)g(y)dy\right)\left(\int\gamma^*(y)g(y)dy\right)\right] \\
&+ \left[\int\alpha^*(y)g(y)dy-\left(\int\gamma^*(y)g(y)dy\right)^2\right]. \qquad \text{(A.108)}
\end{aligned}$$

Expressions (A.99), (A.100), (A.101), (A.102), (A.103), (A.104), (A.105) and (A.108) can be used in (A.98) to obtain (4.24). □

We will proceed now with the covariance terms in (A.79).

Lemma 4.5.15. *The covariance between* $\widehat{A}_1^*$ *and* $\widehat{A}_2^*$ *is*

$$Cov\left(\widehat{A}_1^*, \widehat{A}_2^*\right) = 0.$$

Proof. Let us define

$$\omega_i^* := \frac{(K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v).$$

Using this definition and the expressions for $\widehat{A}_1^*$ in (A.81), $\widehat{A}_2^*$ in (A.82) and in (A.90) and for $\widehat{A}_3^*$ in (A.83) and in (A.97), we have

$$\widehat{A}_1^* = \frac{1}{N}\sum_{i=1}^N \omega_i^*,$$

$$\widehat{A}_2^* = \frac{1}{Nn}\sum_{i=1}^N\sum_{j=1}^n \eta_{ij}^*, \quad \text{with} \quad \eta_{ij}^* = \frac{K_h(Y_i - X_j) - (K_h * f)(Y_i)}{(K_b * g)(Y_i)}(v(Y_i) - \mu_v)$$

and

$$\widehat{A}_3^* = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \tau_{ij}^*, \quad \text{with} \quad \tau_{ij}^* = \varphi^*(Y_i)(K_b(Y_i - Y_j) - (K_b * g)(Y_i)).$$

Let us first consider $Cov(\widehat{A}_1^*, \widehat{A}_2^*)$:

$$\begin{aligned} Cov(\widehat{A}_1^*, \widehat{A}_2^*) &= Cov\left(\frac{1}{N}\sum_{i=1}^N \omega_i^*, \frac{1}{Nn}\sum_{j=1}^N\sum_{k=1}^n \eta_{jk}^*\right) \\ &= \frac{1}{N^2 n}\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^n Cov(\omega_i^*, \eta_{jk}^*) = \frac{1}{N^2 n} Nn Cov(\omega_1^*, \eta_{11}^*) \\ &= \frac{1}{N} Cov(\omega_1^*, \eta_{11}^*), \end{aligned}$$

since $Cov(\omega_i^*, \eta_{jk}^*) = 0$ for $i \neq j$ because ω_i^* and η_{jk}^* are independent for $i \neq j$.

But

$$\begin{aligned} Cov(\omega_1^*, \eta_{11}^*) &= Cov\left(E\left(\omega_1^* | Y_1\right), E\left(\eta_{11}^* | Y_1\right)\right) + E\left[Cov(\omega_1^*, \eta_{11}^* | Y_1)\right] \\ &= Cov\left(E\left(\omega_1^* | Y_1\right), E\left(\eta_{11}^* | Y_1\right)\right) = 0, \end{aligned}$$

since $E\left(\eta_{11}^* | Y_1\right) = 0$. Consequently:

$$Cov(\widehat{A}_1^*, \widehat{A}_2^*) = 0.$$

□

Lemma 4.5.16. *The covariance between $\widehat{A}_1^*$ and $\widehat{A}_3^*$ is*

$$Cov\left(\widehat{A}_1^*, \widehat{A}_3^*\right) = D_9^* \frac{1}{N} + D_{10}^* \frac{1}{N^2}, \tag{4.25}$$

where

$$\begin{aligned} D_9^* &:= \int\int \gamma^*(z)\varphi^*(y)K_b(y-z)g(y)g(z)dydz - \left(\int \gamma^*(y)g(y)dy\right)^2, \\ D_{10}^* &:= \frac{K(0)}{b}\left[\int \delta^*(y)g(y)dy - \left(\int \varphi^*(y)g(y)dy\right)\left(\int \gamma^*(y)g(y)dy\right)\right] \\ &- \int \alpha^*(y)g(y)dy + 2\left(\int \gamma^*(y)g(y)dy\right)^2 \\ &- \int\int \gamma^*(z)\varphi^*(y)K_b(y-z)g(y)g(z)dydz. \end{aligned}$$

Proof. Let us now consider $Cov(\widehat{A}_1^*, \widehat{A}_3^*)$:

$$\begin{aligned} Cov(\widehat{A}_1^*, \widehat{A}_3^*) = Cov\left(\frac{1}{N}\sum_{i=1}^N \omega_i^*, \frac{1}{N^2}\sum_{j=1}^N\sum_{k=1}^N \tau_{jk}^*\right) &= \frac{1}{N^3}\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N Cov(\omega_i^*, \tau_{jk}^*) \\ &= \frac{1}{N^3}\left[N(N-1)(N-2)Cov(\omega_1^*, \tau_{23}^*) + N(N-1)Cov(\omega_1^*, \tau_{12}^*)\right. \\ &+ \left. N(N-1)Cov(\omega_1^*, \tau_{21}^*) + N(N-1)Cov(\omega_1^*, \tau_{22}^*) + NCov(\omega_1^*, \tau_{11}^*)\right] \\ &= \frac{N-1}{N^2}Cov(\omega_1^*, \tau_{12}^*) + \frac{N-1}{N^2}Cov(\omega_1^*, \tau_{21}^*) + \frac{1}{N^2}Cov(\omega_1^*, \tau_{11}^*), \end{aligned} \tag{A.109}$$

since ω_1^* and τ_{23}^* are independent and ω_1^* and τ_{22}^* are also independent.

Let us study now the terms in (A.109):

$$\begin{aligned} Cov(\omega_1^*, \tau_{12}^*) &= E\left[Cov(\omega_1^*, \tau_{12}^*|Y_1)\right] + Cov(E(\omega_1^*|Y_1), E(\tau_{12}^*|Y_1)) \\ &= Cov(E(\omega_1^*|Y_1), E(\tau_{12}^*|Y_1)) = 0, \end{aligned} \tag{A.110}$$

since $E(\tau_{12}^*|Y_1) = 0$.

Now consider

$$\begin{aligned} Cov(\omega_1^*, \tau_{21}^*) &= E\left[Cov(\omega_1^*, \tau_{21}^*|Y_1)\right] + Cov(E(\omega_1^*|Y_1), E(\tau_{21}^*|Y_1)) \\ &= Cov(\omega_1^*, E(\tau_{21}^*|Y_1)) = E\left[\omega_1^* E(\tau_{21}^*|Y_1)\right] - E(\omega_1^*)E\left[E(\tau_{21}^*|Y_1)\right] \\ &= E\left[\omega_1^* E(\tau_{21}^*|Y_1)\right]. \end{aligned}$$

But

$$\begin{aligned} E(\tau_{21}^*|Y_1) &= E\left[\varphi^*(Y_2)\left(K_b(Y_2 - Y_1) - (K_b * g)(Y_2)\right)|Y_1\right] \\ &= \int \varphi^*(y)\left(K_b(y - Y_1) - (K_b * g)(y)\right)g(y)dy. \end{aligned}$$

Thus

$$
\begin{aligned}
Cov(\omega_1^*, \tau_{21}^*) &= \int \frac{(K_h * f)(z)}{(K_b * g)(z)}(v(z) - \mu_v) \\
&\cdot \left(\int \varphi^*(y)\left(K_b(y-z) - (K_b * g)(y)\right) g(y)dy\right) g(z)dz \\
&= \int\int \frac{(K_h * f)(z)}{(K_b * g)(z)}(v(z) - \mu_v)\varphi^*(y)\left(K_b(y-z) - (K_b * g)(y)\right) g(y)g(z)dydz \\
&= \int\int \gamma^*(z)\varphi^*(y)K_b(y-z)g(y)g(z)dydz - \left(\int \gamma^*(y)g(y)dy\right)^2. \qquad (A.111)
\end{aligned}
$$

The last term in (A.109) is:

$$
\begin{aligned}
Cov(\omega_1^*, \tau_{11}^*) &= E\left(\omega_1^*\tau_{11}^*\right) - E(\omega_1^*)E(\tau_{11}^*) \\
&= E\left[\frac{(K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)\varphi^*(Y_1)\left(K_b(0) - (K_b * g)(Y_1)\right)\right] \\
&- E\left[\frac{(K_h * f)(Y_1)}{(K_b * g)(Y_1)}(v(Y_1) - \mu_v)\right] E\left[\varphi^*(Y_1)\left(K_b(0) - (K_b * g)(Y_1)\right)\right] \\
&= \int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)\varphi^*(y)\left(K_b(0) - (K_b * g)(y)\right) g(y)dy \\
&- \left(\int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy\right) \\
&\cdot \left(\int \varphi^*(y)\left(K_b(0) - (K_b * g)(y)\right) g(y)dy\right) \\
&= \frac{K(0)}{b}\left[\int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)\varphi^*(y)g(y)dy\right. \\
&- \left.\left(\int \varphi^*(y)g(y)dy\right)\left(\int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy\right)\right] \\
&- \left[\int (K_h * f)(y)(v(y) - \mu_v)\varphi^*(y)g(y)dy\right. \\
&- \left(\int \varphi^*(y)(K_b * g)(y)g(y)dy\right) \\
&\cdot \left.\left(\int \frac{(K_h * f)(y)}{(K_b * g)(y)}(v(y) - \mu_v)g(y)dy\right)\right] \\
&= \frac{K(0)}{b}\left[\int \delta^*(y)g(y)dy - \left(\int \varphi^*(y)g(y)dy\right)\left(\int \gamma^*(y)g(y)dy\right)\right] \\
&- \int \alpha^*(y)g(y)dy + \left(\int \gamma^*(y)g(y)dy\right)^2. \qquad (A.112)
\end{aligned}
$$

Using (A.110), (A.111) and (A.112) in (A.109) gives (4.25). □

Lemma 4.5.17. *The covariance between $\widehat{A}_2^*$ and $\widehat{A}_3^*$ is*

$$
Cov\left(\widehat{A}_2^*, \widehat{A}_3^*\right) = 0.
$$

Proof. Finally, we consider $Cov(\widehat{A}_2^*, \widehat{A}_3^*)$:

$$\begin{aligned}
Cov(\widehat{A}_2^*, \widehat{A}_3^*) &= Cov\left(\frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}\eta_{ij}^*, \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\tau_{kl}^*\right)\\
&= \frac{1}{N^3n}\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{N}Cov(\eta_{ij}^*, \tau_{kl}^*) = \frac{1}{N^3n}\left[nN(N-1)(N-2)Cov(\eta_{11}^*, \tau_{23}^*)\right.\\
&+ nN(N-1)Cov(\eta_{11}^*, \tau_{12}^*) + nN(N-1)Cov(\eta_{11}^*, \tau_{21}^*)\\
&+ \left. nN(N-1)Cov(\eta_{11}^*, \tau_{22}^*) + nNCov(\eta_{11}^*, \tau_{11}^*)\right]\\
&= \frac{N-1}{N^2}Cov(\eta_{11}^*, \tau_{12}^*) + \frac{N-1}{N^2}Cov(\eta_{11}^*, \tau_{21}^*) + \frac{1}{N^2}Cov(\eta_{11}^*, \tau_{11}^*), \qquad \text{(A.113)}
\end{aligned}$$

since η_{11}^* only depends on Y_1 and X_1 and τ_{23}^* only depends on Y_2 and Y_3, then η_{11}^* and τ_{23}^* are independent. Similiarly η_{11}^* and τ_{22}^* are also independent. Consequently:

$$Cov(\eta_{11}^*, \tau_{23}^*) = 0 \quad \text{and} \quad Cov(\eta_{11}^*, \tau_{22}^*) = 0.$$

Let us now consider the other three terms in (A.113):

$$\begin{aligned}
Cov(\eta_{11}^*, \tau_{12}^*) &= E\left[Cov(\eta_{11}^*, \tau_{12}^*|Y_1)\right] + Cov(E(\eta_{11}^*|Y_1), E(\tau_{12}^*|Y_1))\\
&= Cov(E(\eta_{11}^*|Y_1), E(\tau_{12}^*|Y_1)) = 0, \qquad \text{(A.114)}
\end{aligned}$$

since $E(\eta_{11}^*|Y_1) = 0$ and $E(\tau_{12}^*|Y_1) = 0$.

Let's deal now with the term $Cov(\eta_{11}^*, \tau_{21}^*)$ in (A.113):

$$\begin{aligned}
Cov(\eta_{11}^*, \tau_{21}^*) &= E\left[Cov(\eta_{11}^*, \tau_{21}^*|Y_1)\right] + Cov(E(\eta_{11}^*|Y_1), E(\tau_{21}^*|Y_1))\\
&= Cov(E(\eta_{11}^*|Y_1), E(\tau_{21}^*|Y_1)) = 0, \qquad \text{(A.115)}
\end{aligned}$$

since $E(\eta_{11}^*|Y_1) = 0$.

We finally deal with the last term in (A.113):

$$\begin{aligned}
Cov(\eta_{11}^*, \tau_{11}^*) &= E\left[Cov(\eta_{11}^*, \tau_{11}^*|Y_1)\right] + Cov(E(\eta_{11}^*|Y_1), E(\tau_{11}^*|Y_1))\\
&= Cov(E(\eta_{11}^*|Y_1), E(\tau_{11}^*|Y_1)) = 0, \qquad \text{(A.116)}
\end{aligned}$$

since $E(\eta_{11}^*|Y_1) = 0$.

As a consequence of (A.114), (A.115) and (A.116) in (A.113), we obtain:

$$Cov(\widehat{A}_2^*, \widehat{A}_3^*) = 0.$$

□

Lemma 4.5.18. *The variance of $\widehat{A}$ is*

$$Var\left(\widehat{A}^{*}\right) = D_4^{*}\frac{1}{n} + D_{11}^{*}\frac{1}{N} + D_5^{*}\frac{1}{Nn} + D_{12}^{*}\frac{1}{N^2} + D_8^{*}\frac{1}{N^3},$$

where

$$\begin{aligned} D_{11}^{*} &:= D_3^{*} + D_6^{*} - 2D_9^{*}, \\ D_{12}^{*} &:= D_7^{*} - 2D_{10}^{*}. \end{aligned}$$

Proof. Consequence of Lemmas A.1.10, 4.5.12, 4.5.13, 4.5.14, 4.5.15, 4.5.16 and 4.5.17. □

Theorem 4.2.2. *Let us assume $h \to h_0 > 0$, $b \to b_0 > 0$, $n \to \infty$, $N/n \to \infty$, and Conditions A1 and A8-A10. The asymptotic mean squared error for the estimator $\hat{\mu}_v$ in (4.4) is given by*

$$AMSE\left(\hat{\mu}_v\right) = C_1^{*} + \frac{C_2^{*}}{n} + \frac{C_3^{*}}{N} + \frac{C_4^{*}}{Nn} + \frac{C_5^{*}}{N^2} + \frac{C_6^{*}}{N^3},$$

where the first two constants are

$$\begin{aligned} C_1^{*} &= \left(\int \frac{K_h * f(y)}{K_b * g(y)}(v(y) - \mu_v)g(y)dy\right)^2 \\ C_2^{*} &= \int\left(\int \frac{K_h(y-z)}{K_b * g(y)}(v(y) - \mu_v)g(y)dy - C_1^{*1/2}\right)^2 f(z)dz \end{aligned}$$

and C_3^{*}, C_4^{*}, C_5^{*} and C_6^{*} are constants depending on populational functions reported in the sketch of the proofs (Subsection 4.5.2).

Proof. Consequence of Lemmas 4.5.11 and 4.5.18, considering $C_1^{*} := D_1^{*2}$, $C_2^{*} := D_4^{*}$, $C_3^{*} := 2D_1^{*}D_2^{*} + D_{11}^{*}$, $C_4^{*} := D_5^{*}$, $C_5^{*} := D_2^{*2} + D_{12}^{*}$ and $C_6^{*} := D_8^{*}$. □

A.2 Proofs of the results in Chapter 5

A.2.1 Proof of Theorem 5.2.1

Let us first state an auxiliary lemma:

Lemma 5.5.1. *The difference $\hat{\hat{\mu}}_v - \mu_v$ can be expressed as follows*

$$\hat{\hat{\mu}}_v - \mu_v = \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}} \simeq \frac{\widehat{A}^{\bullet}}{c}, \tag{5.5}$$

where

$$\widehat{A}^{\bullet} = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}(v(Y_i) - \mu_v) \tag{5.6}$$

and

$$\widehat{B}^{\bullet} = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}, \tag{5.7}$$

considering c given by the relation

$$\frac{g(y)}{m(y)} = c \cdot \frac{f(y)}{g(y)}. \tag{5.8}$$

Proof. Considering the estimator $\hat{\hat{\mu}}_v$ defined in (5.4), the difference $\hat{\hat{\mu}}_v - \mu_v$ can be expressed as follows:

$$\hat{\hat{\mu}}_v - \mu_v = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}(v(Y_i)-\mu_v)}{\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}} = \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}},$$

being $\widehat{A}^{\bullet}$ and $\widehat{B}^{\bullet}$ the terms defined in (5.6) and (5.7), respectively.

Considering the relation in (5.8), $\widehat{B}^{\bullet}$ is a consistent estimator of

$$B^{\bullet} = E\left(\frac{g(Y)}{m(Y)}\right) = c \cdot E\left(\frac{f(Y)}{g(Y)}\right) = c\int \frac{f(y)}{g(y)}g(y)dy = c\int f(y)dy = c, \tag{A.117}$$

and we have that

$$\begin{aligned} A^{\bullet} &= E\left(\frac{g(Y)}{m(Y)}(v(Y)-\mu_v)\right) = c \cdot E\left(\frac{f(Y)}{g(Y)}(v(Y)-\mu_v)\right) \\ &= = c\int \frac{f(y)}{g(y)}(v(y)-\mu_v)g(y)dy = c\int f(y)(v(y)-\mu_v)dy \\ &= c\int f(y)v(y)dy - c\cdot\mu_v\int f(y)dy = c\cdot\mu_v - c\cdot\mu_v = 0, \end{aligned} \tag{A.118}$$

so we can express

$$\begin{aligned} \hat{\hat{\mu}}_v - \mu_v &= \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}} = \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}}\left(\frac{\widehat{B}^{\bullet}}{B^{\bullet}} + \left(1 - \frac{\widehat{B}^{\bullet}}{B^{\bullet}}\right)\right) = \frac{\widehat{A}^{\bullet}}{B^{\bullet}} + \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}}\cdot\frac{B^{\bullet}-\widehat{B}^{\bullet}}{\widehat{B}^{\bullet}} \\ &= \frac{\widehat{A}^{\bullet}}{B^{\bullet}} + \left(\frac{\widehat{A}^{\bullet}}{B^{\bullet}} + \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}}\cdot\frac{B^{\bullet}-\widehat{B}^{\bullet}}{\widehat{B}^{\bullet}}\right)\cdot\frac{B^{\bullet}-\widehat{B}^{\bullet}}{\widehat{B}^{\bullet}} \\ &= \frac{\widehat{A}^{\bullet}}{B^{\bullet}} + \frac{\widehat{A}^{\bullet}}{B^{\bullet}}\cdot\frac{B^{\bullet}-\widehat{B}^{\bullet}}{\widehat{B}^{\bullet}} + \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}}\cdot\frac{(\widehat{B}^{\bullet}-B^{\bullet})^2}{\widehat{B}^{\bullet 2}} \\ &= \frac{\widehat{A}^{\bullet}}{B^{\bullet}} + \frac{A^{\bullet}}{B^{\bullet}}\cdot\frac{B^{\bullet}-\widehat{B}^{\bullet}}{\widehat{B}^{\bullet}} + \frac{\widehat{A}^{\bullet}-A^{\bullet}}{B^{\bullet}}\cdot\frac{B^{\bullet}-\widehat{B}^{\bullet}}{\widehat{B}^{\bullet}} + \frac{\widehat{A}^{\bullet}}{\widehat{B}^{\bullet}}\cdot\frac{(\widehat{B}^{\bullet}-B^{\bullet})^2}{\widehat{B}^{\bullet 2}} \simeq \frac{\widehat{A}^{\bullet}}{B^{\bullet}}. \end{aligned}$$

As a consequence of (A.117) and (A.118), we obtain (5.5). □

The term $\widehat{A}^{\bullet}$ defined in (5.6) can be splitted in different terms:

$$\widehat{A}^{\bullet} = \widehat{A}_1^{\bullet} - \widehat{A}_2^{\bullet} + \widehat{A}_3^{\bullet} - \widehat{A}_4^{\bullet} + \widehat{A}_5^{\bullet},$$

where

$$\begin{aligned}
\widehat{A}_1^{\bullet} &:= \frac{1}{N}\sum_{i=1}^{N} \frac{g(Y_i)}{m(Y_i)}(v(Y_i) - \mu_v), \\
\widehat{A}_2^{\bullet} &:= \frac{1}{N}\sum_{i=1}^{N} \frac{g(Y_i)(\hat{m}_h(Y_i) - m(Y_i))}{m(Y_i)^2}(v(Y_i) - \mu_v), \\
\widehat{A}_3^{\bullet} &:= \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{g}_b(Y_i) - g(Y_i)}{m(Y_i)}(v(Y_i) - \mu_v), \\
\widehat{A}_4^{\bullet} &:= \frac{1}{N}\sum_{i=1}^{N} \frac{(\hat{g}_b(Y_i) - g(Y_i))(\hat{m}_h(Y_i) - m(Y_i))}{m(Y_i)^2}(v(Y_i) - \mu_v), \\
\widehat{A}_5^{\bullet} &:= \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} \left(\frac{\hat{m}_h(Y_i) - m(Y_i)}{m(Y_i)}\right)^2 (v(Y_i) - \mu_v).
\end{aligned}$$

Since the terms $\widehat{A}_4^{\bullet}$ and $\widehat{A}_5^{\bullet}$ have some factors of quadratic nature whitin the sum (i.e. $(\hat{g}_b(Y_i) - g(Y_i))(\hat{m}_h(Y_i) - m(Y_i))$ and $(\hat{m}_h(Y_i) - m(Y_i))^2$) it is expected that one could prove negligibility of this terms.

Thus we will consider

$$\widehat{A}^{\bullet} \simeq \widehat{A}_1^{\bullet} - \widehat{A}_2^{\bullet} + \widehat{A}_3^{\bullet}.$$

Lemma 5.5.2. *The expectation and variance of* $\widehat{A}^{\bullet}$ *can be approximated by*

$$\begin{aligned}
E\left(\widehat{A}^{\bullet}\right) &\simeq E\left(\widehat{A}_1^{\bullet}\right) - E\left(\widehat{A}_2^{\bullet}\right) + E\left(\widehat{A}_3^{\bullet}\right), \qquad (5.9)\\
Var\left(\widehat{A}^{\bullet}\right) &\simeq Var\left(\widehat{A}_1^{\bullet}\right) + Var\left(\widehat{A}_2^{\bullet}\right) + Var\left(\widehat{A}_3^{\bullet}\right) \\
&- 2Cov\left(\widehat{A}_1^{\bullet}, \widehat{A}_2^{\bullet}\right) + 2Cov\left(\widehat{A}_1^{\bullet}, \widehat{A}_3^{\bullet}\right) - 2Cov\left(\widehat{A}_2^{\bullet}, \widehat{A}_3^{\bullet}\right). \qquad (5.10)
\end{aligned}$$

Proof. We may rewrite the ratio $\hat{g}_b(Y_i)/\hat{m}_h(Y_i)$ involved in the definition of $\widehat{A}^{\bullet}$ in (5.6):

$$\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} = \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\left(\frac{\hat{m}_h(Y_i)}{m(Y_i)} + 1 - \frac{\hat{m}_h(Y_i)}{m(Y_i)}\right)$$

which gives:

$$\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} = \frac{\hat{g}_b(Y_i)}{m(Y_i)} + \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\frac{m(Y_i) - \hat{m}_h(Y_i)}{m(Y_i)}.$$

As a consequence,

$$\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} - \frac{g(Y_i)}{m(Y_i)} = \frac{\hat{g}_b(Y_i) - g(Y_i)}{m(Y_i)} + \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} \frac{m(Y_i) - \hat{m}_h(Y_i)}{m(Y_i)}.$$

Applying similar techniques once more to the term $\hat{g}_b(Y_i)/\hat{m}_h(Y_i)$ in the right-hand side of the previous expression gives:

$$\begin{aligned}\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} - \frac{g(Y_i)}{m(Y_i)} &= \frac{\hat{g}_b(Y_i) - g(Y_i)}{m(Y_i)} + \frac{g(Y_i)}{m(Y_i)} \frac{m(Y_i) - \hat{m}_h(Y_i)}{m(Y_i)} \\ &+ \frac{\hat{g}_b(Y_i) - g(Y_i)}{m(Y_i)} \frac{m(Y_i) - \hat{m}_h(Y_i)}{m(Y_i)} + \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} \left(\frac{m(Y_i) - \hat{m}_h(Y_i)}{m(Y_i)} \right)^2,\end{aligned}$$

which can also be expressed as:

$$\begin{aligned}\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} - \frac{g(Y_i)}{m(Y_i)} &= \frac{\hat{g}_b(Y_i) - g(Y_i)}{m(Y_i)} - \frac{g(Y_i)}{m(Y_i)} \frac{\hat{m}_h(Y_i) - m(Y_i)}{m(Y_i)} \\ &- \frac{(\hat{g}_b(Y_i) - g(Y_i))(\hat{m}_h(Y_i) - m(Y_i))}{m(Y_i)^2} + \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} \left(\frac{\hat{m}_h(Y_i) - m(Y_i)}{m(Y_i)} \right)^2.\end{aligned}$$

Therefore, the term $\widehat{A}^\bullet$ defined in (5.6) can be splitted in different terms:

$$\widehat{A}^\bullet = \widehat{A}_1^\bullet - \widehat{A}_2^\bullet + \widehat{A}_3^\bullet - \widehat{A}_4^\bullet + \widehat{A}_5^\bullet,$$

where

$$\begin{aligned}\widehat{A}_1^\bullet &:= \frac{1}{N} \sum_{i=1}^N \frac{g(Y_i)}{m(Y_i)} (v(Y_i) - \mu_v), \\ \widehat{A}_2^\bullet &:= \frac{1}{N} \sum_{i=1}^N \frac{g(Y_i)(\hat{m}_h(Y_i) - m(Y_i))}{m(Y_i)^2} (v(Y_i) - \mu_v), \\ \widehat{A}_3^\bullet &:= \frac{1}{N} \sum_{i=1}^N \frac{\hat{g}_b(Y_i) - g(Y_i)}{m(Y_i)} (v(Y_i) - \mu_v), \\ \widehat{A}_4^\bullet &:= \frac{1}{N} \sum_{i=1}^N \frac{(\hat{g}_b(Y_i) - g(Y_i))(\hat{m}_h(Y_i) - m(Y_i))}{m(Y_i)^2} (v(Y_i) - \mu_v), \\ \widehat{A}_5^\bullet &:= \frac{1}{N} \sum_{i=1}^N \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} \left(\frac{\hat{m}_h(Y_i) - m(Y_i)}{m(Y_i)} \right)^2 (v(Y_i) - \mu_v).\end{aligned}$$

Since the terms $\widehat{A}_4^\bullet$ and $\widehat{A}_5^\bullet$ have some factors of quadratic nature whitin the sum (i.e. $(\hat{g}_b(Y_i) - g(Y_i))(\hat{m}_h(Y_i) - m(Y_i))$ and $(\hat{m}_h(Y_i) - m(Y_i))^2$) it is expected that one could prove negligibility of this terms.

Thus we will consider

$$\widehat{A}^{\bullet} \simeq \widehat{A}_1^{\bullet} - \widehat{A}_2^{\bullet} + \widehat{A}_3^{\bullet}.$$

Since we want to obtain the mean and variance of $\widehat{A}^{\bullet}$, we proceed as shown in (5.9) and (5.10). □

The proof of Theorem 5.2.1 proceeds by analyzing the expectations and variances involved.

Lemma 5.5.3. *The expectation of $\widehat{A}^{\bullet}$ is*

$$\begin{aligned} E(\widehat{A}^{\bullet}) &\simeq D_1^{\bullet}\frac{1}{Nb} + D_2^{\bullet}b^2 + D_3^{\bullet}b^4 - D_2^{\bullet}\frac{b^2}{N} - D_3^{\bullet}\frac{b^4}{N} \\ &+ D_4^{\bullet}h^2 + D_5^{\bullet}h^4 + O(b^6) + O(h^6), \end{aligned} \tag{5.11}$$

where

$$\begin{aligned} D_1^{\bullet} &:= K(0)c\int \gamma(y)dy, \\ D_2^{\bullet} &:= \frac{\mu_2(K)}{2}c\int \gamma(y)g''(y)dy, \\ D_3^{\bullet} &:= \frac{\mu_4(K)}{24}c\int \gamma(y)g^{(4)}dy, \\ D_4^{\bullet} &:= -\frac{\mu_2(K)}{2}c^2B^{\bullet}\left(\Omega''\right), \\ D_5^{\bullet} &:= -\frac{\mu_4(K)}{24}c^2B^{\bullet}(\Omega^{(4)}), \end{aligned}$$

the operator $B^{\bullet}$ is defined by

$$B^{\bullet}(\phi) := \int \phi(x)m(x)dx,$$

$$\gamma(y) := \frac{f(y)}{g(y)}(v(y) - \mu_v)$$

and

$$\Omega(y) := \frac{f(y)^2}{g(y)^2}(v(y) - \mu_v).$$

Proof. We now consider the terms in the right-hand side of (5.9).

$$\begin{aligned} E\left(\widehat{A}_1^{\bullet}\right) &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{g(Y_i)}{m(Y_i)}(v(Y_i) - \mu_v)\right] = \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{g(Y_1)}{m(Y_1)}(v(Y_1) - \mu_v)\right] \\ &= E\left[\frac{g(Y_1)}{m(Y_1)}(v(Y_1) - \mu_v)\right] = \int \frac{g(y)}{m(y)}(v(y) - \mu_v)g(y)dy \\ &= c\cdot\int \frac{f(y)}{g(y)}(v(y) - \mu_v)g(y)dy = c\cdot\int v(y)f(y)dy - c\cdot\mu_v\int f(y)dy \\ &= c\cdot\mu_v - c\cdot\mu_v = 0. \end{aligned}$$

Since the random variables

$$\eta_i^\bullet := \Psi(X_1,\ldots,X_n,Y_i) := \frac{g(Y_i)\left[\hat{m}_h(Y_i)-m(Y_i)\right]}{m(Y_i)^2}(v(Y_i)-\mu_v), \quad i=1,2,\ldots,N$$

are identically distributed (but not independent) then

$$\begin{aligned}
E(\widehat{A}_2^\bullet) &= \frac{1}{N}\sum_{i=1}^{N}E(\eta_i^\bullet)=\frac{1}{N}\sum_{i=1}^{N}E(\eta_1^\bullet)=E(\eta_1^\bullet)=E\left[E(\eta_1^\bullet|Y_1)\right]\\
&= E\left[E\left(\frac{g(Y_1)(\hat{m}_h(Y_1)-m(Y_1))}{m(Y_1)^2}(v(Y_1)-\mu_v)\middle|Y_1\right)\right]\\
&= E\left[\frac{g(Y_1)(E\left[\hat{m}_h(Y_1)|Y_1\right]-m(Y_1))}{m(Y_1)^2}(v(Y_1)-\mu_v)\right]\\
&= E\left[\frac{g(Y_1)((K_h*m)(Y_1)-m(Y_1))}{m(Y_1)^2}(v(Y_1)-\mu_v)\right]\\
&= \int\frac{((K_h*m)(y)-m(y))}{m(y)^2}g(y)(v(y)-\mu_v)g(y)dy\\
&= \int\left(\int K_h(y-z)m(z)dz\right)\frac{g(y)^2(v(y)-\mu_v)}{m(y)^2}dy\\
&- c\int\frac{f(y)}{g(y)}g(y)(v(y)-\mu_v)dy\\
&= c^2\int m(z)\left(\int K(t)\frac{f(z+ht)^2(v(z+ht)-\mu_v)}{g(z+ht)^2}dt\right)dz\\
&= c^2\int m(z)\left(\int K(t)\Omega(z+ht)dt\right)dz\\
&= c^2\int\Omega(z)m(z)dz+\frac{h^2\mu_2(K)}{2}c^2\int\Omega''(z)m(z)dz\\
&+ \frac{h^4\mu_4(K)}{24}c^2\int\Omega^{(4)}(z)m(z)dz+O(b^6)\\
&= \frac{h^2\mu_2(K)}{2}c^2B^\bullet(\Omega'')+\frac{h^4\mu_4(K)}{24}c^2B^\bullet(\Omega^{(4)})+O(h^6), \qquad \text{(A.119)}
\end{aligned}$$

since $\int\Omega(z)m(z)dz=0$ and

$$m(y)=\frac{g(y)^2}{cf(y)}.$$

Finally,

$$\begin{aligned}
E\left(\widehat{A}_3^\bullet\right) &= \frac{1}{N}\sum_{i=1}^{N}E\left[\frac{\hat{g}_b(Y_i)-g(Y_i)}{m(Y_i)}(v(Y_i)-\mu_v)\right]=E\left[\frac{\hat{g}_b(Y_1)-g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)\right]\\
&= E\left[E\left(\frac{\hat{g}_b(Y_1)-g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)|Y_1\right)\right]\\
&= E\left[\frac{E(\hat{g}_b(Y_1)|Y_1)-g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)\right]
\end{aligned}$$

$$
\begin{aligned}
&= E\left[\frac{\frac{K(0)}{Nb}+\frac{N-1}{N}(K_b*g)(Y_1)-g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)\right] \\
&= \int \frac{\frac{K(0)}{Nb}+\frac{N-1}{N}(K_b*g)(y)-g(y)}{m(y)}(v(y)-\mu_v)g(y)dy \\
&= \frac{K(0)}{Nb}c\int\frac{f(y)}{g(y)}(v(y)-\mu_v)dy+\frac{N-1}{N}c\int\frac{f(y)(K_b*g)(y)}{g(y)}(v(y)-\mu_v)dy \\
&- c\int f(y)(v(y)-\mu_v)dy = \frac{K(0)}{Nb}c\int\frac{f(y)}{g(y)}(v(y)-\mu_v)dy \\
&+ \frac{N-1}{N}c\int\frac{f(y)(K_b*g)(y)}{g(y)}(v(y)-\mu_v)dy=\frac{K(0)}{Nb}c\int\frac{f(y)}{g(y)}(v(y)-\mu_v)dy \\
&+ \frac{N-1}{N}c\int\frac{f(y)\left[g(y)+\frac{\mu_2(K)}{2}b^2g''(y)+\frac{\mu_4(K)}{4!}b^4g^{(4)}(y)+O(b^6)\right]}{g(y)}(v(y)-\mu_v)dy \\
&= \frac{K(0)}{Nb}c\int\gamma(y)dy+\frac{\mu_2(K)}{2}\frac{(N-1)b^2}{N}c\int\gamma(y)g''(y)dy \\
&+ \frac{\mu_4(K)}{24}\frac{(N-1)b^4}{N}c\int\gamma(y)g^{(4)}(y)dy+O(b^6). \qquad \text{(A.120)}
\end{aligned}
$$

From (A.119) and (A.120), since $E(\widehat{A}^{\bullet})\simeq -E(\widehat{A}_2^{\bullet})+E(\widehat{A}_3^{\bullet})$, we get (5.11).

□

We now consider the terms in the right-hand side of (5.10):

Lemma 5.5.4. *The variance of* $\widehat{A}_1^{\bullet}$ *is*

$$Var\left(\widehat{A}_1^{\bullet}\right) = \frac{D_6^{\bullet}}{N},$$

where

$$D_6^{\bullet} := c^2\int\beta(y)dy$$

with

$$\beta(y) := \frac{f(y)^2}{g(y)}(v(y)-\mu_v)^2.$$

Proof.

$$Var\left(\widehat{A}_1^{\bullet}\right) = \frac{1}{N^2}\sum_{i=1}^{N}Var\left[\frac{g(Y_i)}{m(Y_i)}(v(Y_i)-\mu_v)\right]=\frac{1}{N}Var\left[\frac{g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)\right]$$

$$
\begin{aligned}
&= \frac{1}{N}\left\{E\left[\frac{g(Y_1)^2}{m(Y_1)^2}(v(Y_1)-\mu_v)^2\right]-\left(E\left[\frac{g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)\right]\right)^2\right\}\\
&= \frac{1}{N}\left\{\int\frac{g(y)^2}{m(y)^2}(v(y)-\mu_v)^2g(y)dy-\left(\int\frac{g(y)}{m(y)}(v(y)-\mu_v)g(y)dy\right)^2\right\}\\
&= \frac{1}{N}\left\{c^2\int\frac{f(y)^2}{g(y)^2}(v(y)-\mu_v)^2g(y)dy-\left(c\int f(y)(v(y)-\mu_v)dy\right)^2\right\}\\
&= \frac{c^2}{N}\int\frac{f(y)^2}{g(y)}(v(y)-\mu_v)^2dy=\frac{c^2}{N}\int\beta(y)dy.
\end{aligned}
$$

□

Lemma 5.5.5. *The variance of $\widehat{A}_2^\bullet$ is*

$$
\begin{aligned}
Var\left(\widehat{A}_2^\bullet\right) &= D_7^\bullet\frac{1}{n}+D_8^\bullet\frac{1}{Nn}+D_9^\bullet\frac{1}{Nnh}+D_{10}^\bullet\frac{h^2}{n}+D_{11}^\bullet\frac{h^4}{n}+D_{12}^\bullet\frac{h^4}{N}+D_{13}^\bullet\frac{h}{Nn}\\
&+ D_{14}^\bullet\frac{h^2}{Nn}+D_{15}^\bullet\frac{h^3}{Nn}+D_{16}^\bullet\frac{h^6}{n}+D_{17}^\bullet\frac{h^6}{N}+O\left(\frac{h^8}{n}\right)+O\left(\frac{h^4}{Nn}\right), \qquad (5.12)
\end{aligned}
$$

where

$$
\begin{aligned}
D_7^\bullet &:= c^4B^\bullet(\Omega^2),\\
D_8^\bullet &:= -D_6^\bullet-D_7^\bullet,\\
D_9^\bullet &:= \mu_0(K^2)c^3\int\psi(y)dy,\\
D_{10}^\bullet &:= \mu_2(K)c^4B^\bullet(\Omega\cdot\Omega''),\\
D_{11}^\bullet &:= \frac{\mu_2(K)^2}{4}c^4\left[B^\bullet(\Omega''^2)-B^\bullet(\Omega'')^2\right]+\frac{\mu_4(K)}{12}c^4B^\bullet(\Omega\cdot\Omega^{(4)}),\\
D_{12}^\bullet &:= \frac{\mu_2(K)^2}{4}c^4\left[\int\xi(y)m''(y)^2dy-B^\bullet(\Omega'')^2\right],\\
D_{13}^\bullet &:= \frac{\mu_2(K^2)}{2}c^4\int\xi(y)m''(y)dy,\\
D_{14}^\bullet &:= -\mu_2(K)c^4B^\bullet(\Omega\cdot\Omega'')-\mu_2(K)c^3\int\psi(y)m''(y)dy,\\
D_{15}^\bullet &:= \frac{\mu_4(K^2)}{24}c^4\int\xi(y)m^{(4)}(y)dy,\\
D_{16}^\bullet &:= \frac{\mu_2(K)\mu_4(K)}{24}c^4\left[B^\bullet(\Omega''\cdot\Omega^{(4)})-B^\bullet(\Omega'')B^\bullet(\Omega^{(4)})\right]+\frac{\mu_6(K)}{360}c^4B^\bullet(\Omega\cdot\Omega^{(6)}),\\
D_{17}^\bullet &:= \frac{\mu_2(K)\mu_4(K)}{24}c^4\left[\int\xi(y)m''(y)m^{(4)}(y)dy-B^\bullet(\Omega'')B^\bullet(\Omega^{(4)})\right],
\end{aligned}
$$

being

$$\begin{aligned}\psi(y) &:= \frac{f(y)^3}{g(y)^3}(v(y)-\mu_v)^2,\\ \xi(y) &:= \frac{f(y)^4}{g(y)^5}(v(y)-\mu_v)^2.\end{aligned}$$

Proof. In order to compute the variance of $\widehat{A}_2^\bullet$, let us rewrite the terms $\eta_i^\bullet$ as follows:

$$\eta_i^\bullet = \frac{1}{n}\sum_{j=1}^n \frac{g(Y_i)(K_h(Y_i-X_j)-m(Y_i))}{m(Y_i)^2}(v(Y_i)-\mu_v) = \frac{1}{n}\sum_{j=1}^n \eta_{ij}^\bullet$$

with

$$\eta_{ij}^\bullet := \frac{g(Y_i)(K_h(Y_i-X_j)-m(Y_i))}{m(Y_i)^2}(v(Y_i)-\mu_v), \quad i=1,\ldots,N; j=1,\ldots,n. \tag{A.121}$$

Using (A.121), $\widehat{A}_2^\bullet$ can be written as

$$\widehat{A}_2^\bullet = \frac{1}{Nn}\sum_{i=1}^N\sum_{j=1}^n \eta_{ij}^\bullet$$

and

$$\begin{aligned}Var(\widehat{A}_2^\bullet) &= Cov(\widehat{A}_2^\bullet, \widehat{A}_2^\bullet) = Cov\left(\frac{1}{Nn}\sum_{i=1}^N\sum_{j=1}^n \eta_{ij}^\bullet, \frac{1}{Nn}\sum_{k=1}^N\sum_{l=1}^n \eta_{kl}^\bullet\right)\\ &= \frac{1}{N^2n^2}\sum_{i=1}^N\sum_{j=1}^n\sum_{k=1}^N\sum_{l=1}^n Cov(\eta_{ij}^\bullet, \eta_{kl}^\bullet). \end{aligned} \tag{A.122}$$

Collecting all the equal terms in (A.122) gives:

$$Var(\widehat{A}_3^\bullet) = \frac{n-1}{Nn}Cov(\eta_{11}^\bullet, \eta_{12}^\bullet) + \frac{N-1}{Nn}Cov(\eta_{11}^\bullet, \eta_{21}^\bullet) + \frac{1}{Nn}Var(\eta_{11}^\bullet). \tag{A.123}$$

We now work these covariance terms out:

$$Cov(\eta_{11}^\bullet, \eta_{12}^\bullet) = Cov(E(\eta_{11}^\bullet|Y_1), E(\eta_{12}^\bullet|Y_1)) + E\left[Cov(\eta_{11}^\bullet, \eta_{12}^\bullet|Y_1)\right].$$

But $Cov(\eta_{11}^\bullet, \eta_{12}^\bullet|Y_1) = 0$, since

$$\eta_{11}^\bullet := \frac{g(Y_1)(K_h(Y_1-X_1)-m(Y_1))}{m(Y_1)^2}(v(Y_1)-\mu_v)$$

and

$$\eta_{12}^\bullet := \frac{g(Y_1)(K_h(Y_1-X_2)-m(Y_1))}{m(Y_1)^2}(v(Y_1)-\mu_v)$$

are conditionally independent on Y_1 (because X_1 and X_2 are independent).

On the other hand,

$$\begin{aligned}
E\left(\eta^{\bullet}_{11}|Y_1\right) &= E\left(\eta^{\bullet}_{12}|Y_1\right) = \frac{g(Y_1)\left(E\left[K_h(Y_1-X_1)|Y_1\right]-m(Y_1)\right)}{m(Y_1)^2}(v(Y_1)-\mu_v) \\
&= \frac{g(Y_1)\left(E\left[K_h(X_1-Y_1)|Y_1\right]-m(Y_1)\right)}{m(Y_1)^2}(v(Y_1)-\mu_v) \\
&= \frac{g(Y_1)\left((K_h*m)(Y_1)-m(Y_1)\right)}{m(Y_1)^2}(v(Y_1)-\mu_v). \qquad \text{(A.124)}
\end{aligned}$$

Now

$$\begin{aligned}
&Cov(E\left(\eta^{\bullet}_{11}|Y_1\right), E\left(\eta^{\bullet}_{12}|Y_1\right)) = Var(E\left(\eta^{\bullet}_{11}|Y_1\right)) \\
&= E\left[\left(\frac{g(Y_1)\left((K_h*m)(Y_1)-m(Y_1)\right)}{m(Y_1)^2}(v(Y_1)-\mu_v)\right)^2\right] \\
&- \left[E\left(\frac{g(Y_1)\left((K_h*m)(Y_1)-m(Y_1)\right)}{m(Y_1)^2}(v(Y_1)-\mu_v)\right)\right]^2,
\end{aligned}$$

with

$$\begin{aligned}
&E\left[\left(\frac{g(Y_1)\left((K_h*m)(Y_1)-m(Y_1)\right)}{m(Y_1)^2}(v(Y_1)-\mu_v)\right)^2\right] \\
&= \int\left(\frac{g(y)\left((K_h*m)(y)-m(y)\right)}{m(y)^2}(v(y)-\mu_v)\right)^2 g(y)dy \\
&= \int \frac{g(y)^3\left((K_h*m)(y)-m(y)\right)^2}{m(y)^4}(v(y)-\mu_v)^2 dy \\
&= \int \frac{g(y)^3\left(\frac{\mu_2(K)}{2}h^2m''(y)+\frac{\mu_4(K)}{4!}h^4m^{(4)}(y)+O(h^6)\right)^2}{m(y)^4}(v(y)-\mu_v)^2 dy \\
&= \frac{\mu_2(K)^2}{4}h^4\int\frac{g(y)^3m''(y)^2}{m(y)^4}(v(y)-\mu_v)^2 dy \\
&+ \frac{\mu_2(K)\mu_4(K)}{24}h^6\int\frac{g(y)^3m''(y)m^{(4)}(y)}{m(y)^4}(v(y)-\mu_v)^2 dy + O(h^8) \\
&= \frac{\mu_2(K)^2}{4}h^4c^4\int\frac{f(y)^4m''(y)^2}{g(y)^5}(v(y)-\mu_v)^2 dy \\
&+ \frac{\mu_2(K)\mu_4(K)}{24}h^6c^2\int\frac{f(y)^4m''(y)m^{(4)}(y)}{g(y)^5}(v(y)-\mu_v)^2 dy + O(h^8)
\end{aligned}$$

and

$$\begin{aligned}
&\left[E\left(\frac{g(Y_1)\left((K_h * m)(Y_1) - m(Y_1)\right)}{m(Y_1)^2}(v(Y_1) - \mu_v)\right)\right]^2 \\
&= \left[\int \frac{g(y)\left((K_h * m)(y) - m(y)\right)}{m(y)^2}(v(y) - \mu_v)g(y)dy\right]^2 \\
&= \left[\int \frac{g(y)^2\left((K_h * m)(y) - m(y)\right)}{m(y)^2}(v(y) - \mu_v)dy\right]^2 \\
&= \left[c^2 \int \frac{f(y)^2\left((K_h * m)(y) - m(y)\right)}{g(y)^2}(v(y) - \mu_v)dy\right]^2 \\
&= \left[c^2 \int \Omega(y)(K_h * m)(y)dy - c\int f(y)(v(y) - \mu_v)dy\right]^2 \\
&= \left[c^2 \int \Omega(y)(K_h * m)(y)dy\right]^2 \\
&= \left[c^2 \int \Omega(y)\left(\int K_h(y-z)m(z)dz\right)dy\right]^2 \\
&= \left[c^2 \int m(z)\left(\int K(t)\Omega(z+ht)dt\right)dz\right]^2 \\
&= \left[c^2 \int m(z)\Omega(z)dz + \frac{h^2}{2}\mu_2(K)c^2 \int m(z)\Omega''(z)dz\right. \\
&+ \left.\frac{h^4}{24}\mu_4(K)c^2 \int m(z)\Omega^{(4)}(z)dz + O(h^6)\right]^2 \\
&= \left[c \int f(z)(v(z) - \mu_v)dz + \frac{h^2}{2}\mu_2(K)c^2 \int m(z)\Omega''(z)dz\right. \\
&+ \left.\frac{h^4}{24}\mu_4(K)c^2 \int m(z)\Omega^{(4)}(z)dz + O(h^6)\right]^2 \\
&= \left[\frac{h^2}{2}\mu_2(K)c^2 \int m(z)\Omega''(z)dz\right. \\
&+ \left.\frac{h^4}{24}\mu_4(K)c^2 \int m(z)\Omega^{(4)}(z)dz + O(h^6)\right]^2 \\
&= \frac{\mu_2(K)^2}{4}h^4c^4\left(\int m(z)\Omega''(z)dz\right)^2 \\
&+ \frac{\mu_2(K)\mu_4(K)}{24}h^6c^4 \int m(z)\Omega''(z)dz \int m(z)\Omega^{(4)}(z)dz + O(h^8).
\end{aligned}$$

Consequently:

$$
\begin{aligned}
&Cov(\eta^{\bullet}_{11},\eta^{\bullet}_{12}) = Cov(E(\eta^{\bullet}_{11}|Y_1), E(\eta^{\bullet}_{12}|Y_1)) = Var(E(\eta^{\bullet}_{11}|Y_1))\\
= &\ \frac{\mu_2(K)^2}{4}h^4c^4\left(\int \frac{f(y)^4 m''(y)^2}{g(y)^5}(v(y)-\mu_v)^2dy - \left(\int \Omega''(y)m(y)dy\right)^2\right)\\
+ &\ \frac{\mu_2(K)\mu_4(K)}{24}h^6c^4\left(\int \frac{f(y)^4 m''(y)m^{(4)}(y)}{g(y)^5}(v(y)-\mu_v)^2dy\right.\\
- &\left.\int \Omega''(y)m(y)dy\int \Omega^{(4)}(y)m(y)dy\right) + O(h^8)\\
= &\ \frac{\mu_2(K)^2}{4}h^4c^4\left(\int \xi(y)m''(y)^2dy - B^{\bullet}(\Omega'')^2\right) \qquad \text{(A.125)}\\
+ &\ \frac{\mu_2(K)\mu_4(K)}{24}h^6c^4\left(\int \xi(y)m''(y)m^{(4)}(y)dy - B^{\bullet}(\Omega'')B^{\bullet}(\Omega^{(4)})\right) + O(h^8).
\end{aligned}
$$

Now we deal with the term $Cov(\eta^{\bullet}_{11},\eta^{\bullet}_{21})$ in (A.123):

$$
\begin{aligned}
Cov(\eta^{\bullet}_{11},\eta^{\bullet}_{21}) &= Cov(E(\eta^{\bullet}_{11}|X_1), E(\eta^{\bullet}_{21}|X_1)) + E[Cov(\eta^{\bullet}_{11},\eta^{\bullet}_{21}|X_1)]\\
&= Cov(E(\eta^{\bullet}_{11}|X_1), E(\eta^{\bullet}_{21}|X_1)),
\end{aligned}
$$

since $Cov(\eta^{\bullet}_{11},\eta^{\bullet}_{21}|X_1) = 0$ because

$$\eta^{\bullet}_{11} := \frac{g(Y_1)(K_h(Y_1-X_1)-m(Y_1))}{m(Y_1)^2}(v(Y_1)-\mu_v)$$

and

$$\eta^{\bullet}_{21} := \frac{g(Y_2)(K_h(Y_2-X_1)-m(Y_2))}{m(Y_2)^2}(v(Y_2)-\mu_v)$$

are conditionally independent given X_1 (since Y_1 and Y_2 are independent).

But

$$
\begin{aligned}
E(\eta^{\bullet}_{11}|X_1) &= E(\eta^{\bullet}_{21}|X_1) = \int \frac{g(y)^2(K_h(y-X_1)-m(y))}{m(y)^2}(v(y)-\mu_v)dy\\
&= c^2\int \Omega(y)K_h(y-X_1)dy - c\int f(y)(v(y)-\mu_v)dy\\
&= c^2\int K(t)\Omega(X_1+ht)dt.
\end{aligned}
$$

So,

$$
\begin{aligned}
&Cov(E(\eta^{\bullet}_{11}|X_1), E(\eta^{\bullet}_{21}|X_1)) = Var\left(c^2\int K(t)\Omega(X_1+ht)dt\right)\\
= &\ E\left[\left(c^2\int K(t)\Omega(X_1+ht)dt\right)^2\right] - \left[E\left(c^2\int K(t)\Omega(X_1+ht)dt\right)\right]^2.
\end{aligned}
$$

On the one hand,

$$
\begin{aligned}
E\left[c^2\int K(t)\Omega(X_1+ht)dt\right] &= c^2\int\left(\int K(t)\Omega(x+ht)dt\right)m(x)dx \\
= &\ c^2\int\Omega(x)m(x)dx+\frac{h^2}{2}\mu_2(K)c^2\int\Omega''(x)m(x)dx \\
+ &\ \frac{h^4}{24}\mu_4(K)c^2\int\Omega^{(4)}(x)m(x)dx+O(h^6) \\
= &\ \frac{h^2}{2}\mu_2(K)c^2\int\Omega''(x)m(x)dx+\frac{h^4}{24}\mu_4(K)c^2\int\Omega^{(4)}(x)m(x)dx+O(h^6)
\end{aligned}
$$

since $\int\Omega(x)m(x)dx=0$, and consequently,

$$
\begin{aligned}
\left(E\left[c^2\int K(t)\Omega(X_1+ht)dt\right]\right)^2 &= \frac{h^4}{4}\mu_2(K)^2c^4\left(\int\Omega''(x)m(x)dx\right)^2 \\
+ &\ \frac{h^6}{24}\mu_2(K)\mu_4(K)c^4\int\Omega''(x)m(x)dx\int\Omega^{(4)}(x)m(x)dx+O(h^8) \quad (A.126)
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&E\left[\left(c^2\int K(t)\Omega(X_1+ht)dt\right)^2\right] \\
= &\ E\left[c^4\int\int K(s)\Omega(X_1+hs)K(t)\Omega(X_1+ht)dsdt\right] \\
= &\ c^4\int\left(\int\int K(s)\Omega(x+hs)K(t)\Omega(x+ht)dsdt\right)m(x)dx \\
= &\ c^4\int\Omega(x)^2m(x)dx+h^2\mu_2(K)c^4\int\Omega(x)\Omega''(x)m(x)dx \quad (A.127) \\
+ &\ \frac{h^4}{12}\mu_4(K)c^4\int\Omega(x)\Omega^{(4)}(x)m(x)dx+\frac{h^4}{4}\mu_2(K)^2c^4\int\Omega''(x)^2m(x)dx \\
+ &\ \frac{h^6}{360}\mu_6(K)c^4\int\Omega(x)\Omega^{(6)}(x)m(x)dx \\
+ &\ \frac{h^6}{24}\mu_2(K)\mu_4(K)c^4\int\Omega''(x)\Omega^{(4)}(x)m(x)dx+O(h^8).
\end{aligned}
$$

From (A.126) and (A.127), we obtain

$$
\begin{aligned}
Cov(\eta_{11}^\bullet,\eta_{21}^\bullet) &= Var\left(c^2\int K(t)\Omega(X_1+ht)dt\right) \\
= &\ c^4B^\bullet(\Omega^2)+h^2\mu_2(K)c^4B^\bullet(\Omega\cdot\Omega'')+\frac{h^4}{4}\mu_2(K)^2c^4\left[B^\bullet(\Omega''^2)-B^\bullet(\Omega'')^2\right] \\
+ &\ \frac{h^4}{12}\mu_4(K)c^4B^\bullet(\Omega\cdot\Omega^{(4)}) \\
+ &\ \frac{h^6}{24}\mu_2(K)\mu_4(K)c^4\left[B^\bullet(\Omega''\cdot\Omega^{(4)})-B^\bullet(\Omega'')B^\bullet(\Omega^{(4)})\right] \\
+ &\ \frac{h^6}{360}\mu_6(K)c^4B^\bullet(\Omega\cdot\Omega^{(6)})+O(h^8). \quad (A.128)
\end{aligned}
$$

We now examine the term $Var(\eta_{11}^{\bullet})$ in (A.123):

$$\begin{aligned} Var(\eta_{11}^{\bullet}) &= Var\left[E\left(\eta_{11}^{\bullet}|X_1\right)\right] + E\left[Var(\eta_{11}^{\bullet}|X_1)\right] \\ &= Var\left(c^2 \int K(t)\Omega(X_1+ht)dt\right) + E\left[Var(\eta_{11}^{\bullet}|X_1)\right], \end{aligned}$$

since $E\left(\eta_{11}^{\bullet}|X_1\right) = c^2 \int K(t)\Omega(X_1+ht)dt$.

Now

$$Var(\eta_{11}^{\bullet}|X_1) = E\left(\eta_{11}^{\bullet 2}|X_1\right) - \left[E\left(\eta_{11}^{\bullet}|X_1\right)\right]^2 = E\left(\eta_{11}^{\bullet 2}|X_1\right) - \left(c^2 \int K(t)\Omega(X_1+ht)dt\right)^2$$

and since

$$\begin{aligned} E\left[\left[E\left(\eta_{11}^{\bullet}|X_1\right)\right]^2\right] &= E\left[\left(c^2 \int K(t)\Omega(X_1+ht)dt\right)^2\right] = E\left[\left(c^2 \int K(t)\Omega(X_1+ht)dt\right.\right. \\ &- \left.\left. E\left(c^2 \int K(t)\Omega(X_1+ht)dt\right) + E\left(c^2 \int K(t)\Omega(X_1+ht)dt\right)\right)^2\right] \\ &= Var\left(c^2 \int K(t)\Omega(X_1+ht)dt\right) + \left(E\left(c^2 \int K(t)\Omega(X_1+ht)dt\right)\right)^2 \end{aligned}$$

we obtain

$$\begin{aligned} Var(\eta_{11}^{\bullet}) &= E\left(\eta_{11}^{\bullet 2}\right) - E\left(\eta_{11}^{\bullet}\right)^2 = E\left[E\left(\eta_{11}^{\bullet 2}|Y_1\right)\right] - E\left(E\left(\eta_{11}^{\bullet}|X_1\right)\right)^2 \\ &= E\left[E\left(\left(\frac{g(Y_1)\left(K_h(Y_1-X_1)-m(Y_1)\right)}{m(Y_1)^2}(v(Y_1)-\mu_v)\right)^2 \middle| Y_1\right)\right] \\ &- \left[E\left(c^2 \int K(t)\Omega(X_1+ht)dt\right)\right]^2 \\ &= E\left[\frac{((K_h)^2 * m)(Y_1) - 2m(Y_1)(K_h * m)(Y_1) + m(Y_1)^2}{m(Y_1)^4} g(Y_1)^2 (v(Y_1)-\mu_v)^2\right] \\ &- \frac{h^4}{4}\mu_2(K)^2 c^2 B^{\bullet}(\Omega'')^2 + O(h^6) \\ &= \int \frac{((K_h)^2 * m)(y) - 2m(y)(K_h * m)(y) + m(y)^2}{m(y)^4} g(y)^3 (v(y)-\mu_v)^2 dy \\ &- \frac{h^4}{4}\mu_2(K)^2 c^2 B^{\bullet}(\Omega'')^2 + O(h^6). \end{aligned} \tag{A.129}$$

But,

$$\begin{aligned} ((K_h)^2 * m)(y) &= \int (K_h)^2(y-z)m(z)dz = \int \left[K_h(y-z)\right]^2 m(z)dz \\ &= \frac{1}{h}\int K(t)^2 m(y-ht)dt \\ &= \frac{1}{h}\left[\mu_0(K^2)m(y) + \frac{\mu_2(K^2)}{2}h^2 m''(y) + \frac{\mu_4(K^2)}{4!}h^4 m^{(4)}(y) + O(h^6)\right] \\ &= \frac{\mu_0(K^2)}{h}m(y) + \frac{\mu_2(K^2)}{2}hm''(y) + \frac{\mu_4(K^2)}{24}h^3 m^{(4)}(y) + O(h^5). \end{aligned}$$

Now (A.129) becomes:

$$
\begin{aligned}
Var(\eta_{11}^\bullet) &= \int \Bigg[\frac{\mu_0(K^2)}{h} m(y) + \frac{\mu_2(K^2)}{2} h m''(y) + \frac{\mu_4(K^2)}{24} h^3 m^{(4)}(y) + O(h^5) \\
&- 2m(y) \left(m(y) + \frac{\mu_2(K)}{2} h^2 m''(y) + \frac{\mu_4(K)}{24} h^4 m^{(4)}(y) + O(h^6) \right) + m(y)^2 \Bigg] \cdot \\
&\cdot \frac{g(y)^3 (v(y) - \mu_v)^2}{m(y)^4} dy - \frac{h^4}{4} \mu_2(K)^2 c^2 B^\bullet(\Omega'')^2 + O(h^6) \\
&= \frac{\mu_0(K^2)}{h} c^3 \int \psi(y) dy - c^2 \int \beta(y) dy + \frac{\mu_2(K^2)}{2} h c^4 \int \xi(y) m''(y) dy \\
&- \mu_2(K) h^2 c^3 \int \psi(y) m''(y) dy + \frac{\mu_4(K^2)}{24} h^3 c^4 \int \xi(y) m^{(4)}(y) dy \\
&- \frac{\mu_4(K)}{12} h^4 c^3 \int \psi(y) m^{(4)}(y) dy - \frac{h^4}{4} \mu_2(K)^2 c^2 B^\bullet(\Omega'')^2 + O(h^5). \qquad \text{(A.130)}
\end{aligned}
$$

Now, using (A.125), (A.128) and (A.130) in (A.123) gives:

$$
\begin{aligned}
Var(\widehat{A}_2^\bullet) &= \frac{n-1}{Nn} \Bigg[\frac{\mu_2(K)^2}{4} h^4 c^4 \left(\int \xi(y) m''(y)^2 dy - B^\bullet \left(\Omega''\right)^2 \right) \\
&+ \frac{\mu_2(K)\mu_4(K)}{24} h^6 c^4 \left(\int \xi(y) m''(y) m^{(4)}(y) dy - B^\bullet \left(\Omega''\right) B^\bullet \left(\Omega^{(4)}\right) \right) + O(h^8) \Bigg] \\
&+ \frac{N-1}{Nn} \Big[c^4 B^\bullet(\Omega^2) + h^2 \mu_2(K) c^4 B^\bullet(\Omega \cdot \Omega'') \\
&+ \frac{h^4}{4} \mu_2(K)^2 c^4 \left[B^\bullet(\Omega''^2) - B^\bullet(\Omega'')^2 \right] + \frac{h^4}{12} \mu_4(K) c^4 B^\bullet(\Omega \cdot \Omega^{(4)}) \\
&+ \frac{h^6}{24} \mu_2(K) \mu_4(K) c^4 \left[B^\bullet(\Omega'' \cdot \Omega^{(4)}) - B^\bullet(\Omega'') B^\bullet(\Omega^{(4)}) \right] \\
&+ \frac{h^6}{360} \mu_6(K) c^4 B^\bullet(\Omega \cdot \Omega^{(6)}) + O(h^8) \Big] + \frac{1}{Nn} \Bigg[\frac{\mu_0(K^2)}{h} c^3 \int \psi(y) dy \\
&- c^2 \int \beta(y) dy + \frac{\mu_2(K^2)}{2} h c^4 \int \xi(y) m''(y) dy - \mu_2(K) h^2 c^3 \int \psi(y) m''(y) dy \\
&+ \frac{\mu_4(K^2)}{24} h^3 c^4 \int \xi(y) m^{(4)}(y) dy - \frac{\mu_4(K)}{12} h^4 c^3 \int \psi(y) m^{(4)}(y) dy \\
&- \frac{h^4}{4} \mu_2(K)^2 c^2 B^\bullet(\Omega'')^2 + O(h^5) \Bigg]. \qquad \text{(A.131)}
\end{aligned}
$$

Shortening the expression (A.131), we obtain (5.12). □

Lemma 5.5.6. *The variance of* $\widehat{A}_3^\bullet$ *is*

$$
\begin{aligned}
Var\left(\widehat{A}_3^\bullet\right) &= D_{18}^\bullet \frac{1}{N} + D_{19}^\bullet \frac{b^2}{N} + D_{20}^\bullet \frac{b^4}{N} + D_{21}^\bullet \frac{1}{N^2 b} + D_{22}^\bullet \frac{1}{N^2} + D_{23}^\bullet \frac{b}{N^2} \\
&+ D_{24}^\bullet \frac{1}{N^3 b^2} + D_{25}^\bullet \frac{1}{N^3 b} + D_{26}^\bullet \frac{1}{N^3} + O\left(\frac{b^6}{N}\right) + O\left(\frac{b^2}{N^2}\right), \qquad (5.13)
\end{aligned}
$$

where

$$\begin{aligned}
D^\bullet_{18} &:= c^2 \int \beta(y)dy = D^\bullet_6,\\
D^\bullet_{19} &:= \mu_2(K)c^2 \left[\int \alpha(y)g''(y)dy + \int \gamma''(y)f(y)(v(y)-\mu_v)dy\right],\\
D^\bullet_{20} &:= \frac{\mu_4(K)}{12}c^2 \left[\int \alpha(y)g^{(4)}(y)dy + \int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy\right]\\
&+ \frac{\mu_2(K)^2}{4}c^2 \left[\int \delta(y)g''(y)^2dy - 4\left(\int \gamma(y)g''(y)dy\right)^2\right.\\
&+ \left.\int \gamma''(y)^2 g(y)dy + 2\int \gamma(y)\gamma''(y)g''(y)dy\right],\\
D^\bullet_{21} &:= 2c^2\left[\mu_0(K^2)+K(0)\right]\int \alpha(y)dy,\\
D^\bullet_{22} &:= -8c^2\int \beta(y)dy = -8D^\bullet_6,\\
D^\bullet_{23} &:= \mu_2(K)K(0)c^2\left[\int \delta(y)g''(y)dy + \int \gamma(y)\gamma''(y)dy\right.\\
&- \left.\left(\int \gamma(y)g''(y)dy + \int \gamma''(y)g(y)dy\right)\left(\int \gamma(y)dy\right)\right]\\
&+ \frac{\mu_2(K^2)}{2}c^2\left[\int \delta(y)g''(y)dy + \int \gamma(y)\gamma''(y)dy\right],\\
D^\bullet_{24} &:= K(0)^2c^2\left[\int \delta(y)dy - \left(\int \gamma(y)dy\right)^2\right],\\
D^\bullet_{25} &:= -2c^2\left[\mu_0(K^2)+2K(0)\right]\int \alpha(y)dy,\\
D^\bullet_{26} &:= 8c^2\int \beta(y)dy = 8D^\bullet_6,
\end{aligned}$$

with

$$\begin{aligned}
\alpha(y) &:= \frac{f(y)^2}{g(y)^2}(v(y)-\mu_v)^2,\\
\delta(y) &:= \frac{f(y)^2}{g(y)^3}(v(y)-\mu_v)^2.
\end{aligned}$$

Proof. To deal with the variance of $\widehat{A}^\bullet_3$ we first consider

$$\begin{aligned}
\tau^\bullet_i &:= \frac{\hat{g}_b(Y_i)-g(Y_i)}{m(Y_i)}(v(Y_i)-\mu_v)\\
&= \frac{1}{N}\sum_{j=1}^N \frac{K_b(Y_i-Y_j)-g(Y_i)}{m(Y_i)}(v(Y_i)-\mu_v) = \frac{1}{N}\sum_{j=1}^N \tau^\bullet_{ij},
\end{aligned}$$

where

$$\tau^{\bullet}_{ij} := \frac{K_b(Y_i - Y_j) - g(Y_i)}{m(Y_i)}(v(Y_i) - \mu_v).$$

As a consequence, $\widehat{A}^{\bullet}_3$ can be written as

$$\widehat{A}^{\bullet}_3 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau^{\bullet}_{ij}.$$

To compute the variance of $\widehat{A}^{\bullet}_3$ we consider

$$\begin{aligned} Var(\widehat{A}^{\bullet}_3) &= Cov(\widehat{A}^{\bullet}_3, \widehat{A}^{\bullet}_3) = Cov\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau^{\bullet}_{ij}, \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\tau^{\bullet}_{kl}\right) \\ &= \frac{1}{N^4}\sum_{i,j,k,l=1}^{N} Cov(\tau^{\bullet}_{ij}, \tau^{\bullet}_{kl}). \end{aligned}$$

Thus, the variance of $\widehat{A}^{\bullet}_3$ can be written as:

$$\begin{aligned} Var(\widehat{A}^{\bullet}_3) &= \frac{(N-1)(N-2)}{N^3}Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{13}) + \frac{2(N-1)(N-2)}{N^3}Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{31}) \\ &+ \frac{(N-1)(N-2)}{N^3}Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{32}) + \frac{N-1}{N^3}Var(\tau^{\bullet}_{12}) + \frac{N-1}{N^3}Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{21}) \\ &+ \frac{2(N-1)}{N^3}Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{11}) + \frac{2(N-1)}{N^3}Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{22}) + \frac{1}{N^3}Var(\tau^{\bullet}_{11}) \end{aligned} \quad (A.132)$$

Lemmas A.2.1, A.2.2, A.2.3, A.2.4, A.2.5, A.2.6, A.2.7 and A.2.8 can be used in (A.132) to conclude with the asymptotic expression (5.13). □

Let's deal with every one of these terms:

Lemma A.2.1. *The covariance of* $\tau^{\bullet}_{12}$ *and* $\tau^{\bullet}_{13}$ *is*

$$\begin{aligned} Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{13}) &= \frac{\mu_2(K)^2}{4}b^4c^2\left[\int\delta(y)g''(y)^2dy - \left(\int\gamma(y)g''(y)dy\right)^2\right] \\ &+ \frac{\mu_2(K)\mu_4(K)}{24}b^6c^2\left[\int\delta(y)g''(y)g^{(4)}(y)dy\right. \\ &- \left.\left(\int\gamma(y)g''(y)dy\right)\left(\int\gamma(y)g^{(4)}(y)dy\right)\right] + O(b^8). \quad (A.133) \end{aligned}$$

Proof.

$$\begin{aligned} Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{13}) &= E\left[Cov(\tau^{\bullet}_{12}, \tau^{\bullet}_{13}|Y_1)\right] + Cov\left(E\left(\tau^{\bullet}_{12}|Y_1\right), E\left(\tau^{\bullet}_{13}|Y_1\right)\right) \\ &= Var\left(E\left(\tau^{\bullet}_{12}|Y_1\right)\right) = Var\left(\frac{v(Y_1) - \mu_v}{m(Y_1)}\left[(K_b * g)(Y_1) - g(Y_1)\right]\right) \end{aligned}$$

because

$$\tau_{12}^{\bullet} = \frac{v(Y_1) - \mu_v}{m(Y_1)}(K_b(Y_1 - Y_2) - g(Y_1))$$

and

$$\tau_{13}^{\bullet} = \frac{v(Y_1) - \mu_v}{m(Y_1)}(K_b(Y_1 - Y_3) - g(Y_1))$$

are conditionally independent given Y_1 (so $Cov(\tau_{12}^{\bullet}, \tau_{13}^{\bullet}|Y_1) = 0$) and

$$\begin{aligned} E\left(\tau_{12}^{\bullet}|Y_1\right) &= E\left(\tau_{13}^{\bullet}|Y_1\right) = \frac{v(Y_1) - \mu_v}{m(Y_1)}\left[E\left[K_b(Y_1 - Y_2)|Y_1\right] - g(Y_1)\right] \\ &= \frac{v(Y_1) - \mu_v}{m(Y_1)}\left[(K_b * g)(Y_1) - g(Y_1)\right]. \qquad \text{(A.134)} \end{aligned}$$

Now

$$\begin{aligned} Cov(\tau_{12}^{\bullet}, \tau_{13}^{\bullet}) &= E\left[\left(\frac{v(Y_1) - \mu_v}{m(Y_1)}\right)^2 \left[(K_b * g)(Y_1) - g(Y_1)\right]^2\right] \\ &- \left[E\left(\frac{v(Y_1) - \mu_v}{m(Y_1)}\left[(K_b * g)(Y_1) - g(Y_1)\right]\right)\right]^2, \end{aligned}$$

where

$$\begin{aligned} &E\left[\left(\frac{v(Y_1) - \mu_v}{m(Y_1)}\right)^2 \left[(K_b * g)(Y_1) - g(Y_1)\right]^2\right] \\ &= \int \left(\frac{v(y) - \mu_v}{m(y)}\right)^2 \left((K_b * g)(y) - g(y)\right)^2 g(y)dy \\ &= \int \left(\frac{v(y) - \mu_v}{m(y)}\right)^2 \left[\frac{\mu_2(K)}{2}b^2 g''(y) + \frac{\mu_4(K)}{24}b^4 g^{(4)}(y) + O(b^6)\right]^2 g(y)dy \\ &= \frac{\mu_2(K)^2}{4}b^4 c^2 \int \delta(y) g''(y)^2 dy + \frac{\mu_2(K)\mu_4(K)}{24}b^6 c^2 \int \delta(y) g''(y) g^{(4)}(y)dy + O(b^8), \end{aligned}$$

since (considering relation (5.8)),

$$m(y) = \frac{g(y)^2}{cf(y)},$$

and

$$\begin{aligned} &E\left(\frac{v(Y_1) - \mu_v}{m(Y_1)}\left[(K_b * g)(Y_1) - g(Y_1)\right]\right) \\ &= \int \frac{v(y) - \mu_v}{m(y)}\left[\frac{\mu_2(K)}{2}b^2 g''(y) + \frac{\mu_4(K)}{24}b^4 g^{(4)}(y) + O(b^6)\right] g(y)dy \\ &= \frac{\mu_2(K)}{2}b^2 c \int \gamma(y) g''(y)dy + \frac{\mu_4(K)}{24}b^4 c \int \gamma(y) g^{(4)}(y)dy + O(b^6). \text{(A.135)} \end{aligned}$$

Using these two expressions we obtain (A.133). □

Lemma A.2.2. *The covariance of $\tau_{12}^{\bullet}$ and $\tau_{31}^{\bullet}$ is*

$$Cov(\tau_{12}^{\bullet},\tau_{31}^{\bullet}) = \frac{\mu_2(K)}{2}b^2c^2\int\alpha(y)g''(y)dy + b^4c^2\left[\frac{\mu_4(K)}{24}\int\alpha(y)g^{(4)}(y)dy\right.$$
$$+ \quad \left.\frac{\mu_2(K)^2}{4}\left(\int\gamma(y)\gamma''(y)g''(y)dy - \left(\int\gamma(y)g''(y)dy\right)^2\right)\right] + O(b^6).\text{(A.136)}$$

Proof.

$$\begin{aligned} Cov(\tau_{12}^{\bullet},\tau_{31}^{\bullet}) &= E\left[Cov(\tau_{12}^{\bullet},\tau_{31}^{\bullet}|Y_1)\right] + Cov\left(E\left(\tau_{12}^{\bullet}|Y_1\right),E\left(\tau_{31}^{\bullet}|Y_1\right)\right) \\ &= Cov\left(E\left(\tau_{12}^{\bullet}|Y_1\right),E\left(\tau_{31}^{\bullet}|Y_1\right)\right). \end{aligned}$$

We know that

$$E\left(\tau_{12}^{\bullet}|Y_1\right) = \frac{v(Y_1)-\mu_v}{m(Y_1)}\left[(K_b*g)(Y_1) - g(Y_1)\right]$$

and we can also deal with

$$\begin{aligned} E\left(\tau_{31}^{\bullet}|Y_1\right) &= E\left[\frac{v(Y_3)-\mu_v}{m(Y_3)}(K_b(Y_3-Y_1) - g(Y_3))\middle|Y_1\right] \\ &= \int\frac{v(y)-\mu_v}{m(y)}\left[K_b(y-Y_1) - g(y)\right]g(y)dy. \qquad \text{(A.137)} \end{aligned}$$

Now

$$Cov\left(E\left(\tau_{12}^{\bullet}|Y_1\right),E\left(\tau_{31}^{\bullet}|Y_1\right)\right) = E\left[E\left(\tau_{12}^{\bullet}|Y_1\right)E\left(\tau_{31}^{\bullet}|Y_1\right)\right] - E\left[E\left(\tau_{12}^{\bullet}|Y_1\right)\right]E\left[E\left(\tau_{31}^{\bullet}|Y_1\right)\right]$$
$$= \quad E\left[E\left(\tau_{12}^{\bullet}|Y_1\right)E\left(\tau_{31}^{\bullet}|Y_1\right)\right] - E(\tau_{12}^{\bullet})E(\tau_{31}^{\bullet})$$
$$= \quad E\left[E\left(\tau_{12}^{\bullet}|Y_1\right)E\left(\tau_{31}^{\bullet}|Y_1\right)\right] - \left[E(\tau_{12}^{\bullet})\right]^2.$$

But since $E\left(\tau_{12}^{\bullet}|Y_1\right) = \frac{v(Y_1)-\mu_v}{m(Y_1)}\left[(K_b*g)(Y_1) - g(Y_1)\right]$, equation (A.135) lead to:

$$E(\tau_{12}^{\bullet}) = \frac{\mu_2(K)}{2}b^2c\int\gamma(y)g''(y)dy + \frac{\mu_4(K)}{24}b^4c\int\gamma(y)g^{(4)}(y)dy + O(b^6). \quad \text{(A.138)}$$

On the other hand, using the expression (A.137), we have:

$$E\left[E\left(\tau_{12}^{\bullet}|Y_1\right)E\left(\tau_{31}^{\bullet}|Y_1\right)\right]$$
$$= E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}\left[(K_b*g)(Y_1) - g(Y_1)\right]\int\frac{v(y)-\mu_v}{m(y)}\left[K_b(y-Y_1) - g(y)\right]g(y)dy\right]$$
$$= \int\frac{v(z)-\mu_v}{m(z)}\left[(K_b*g)(z) - g(z)\right]\left(\int\frac{v(y)-\mu_v}{m(y)}\left[K_b(y-z) - g(y)\right]g(y)dy\right)g(z)dz$$
$$= \int\frac{v(z)-\mu_v}{m(z)}\left[(K_b*g)(z) - g(z)\right]\left(\int\frac{v(y)-\mu_v}{m(y)}K_b(y-z)g(y)dy\right.$$
$$\left.-\int\frac{v(y)-\mu_v}{m(y)}g(y)^2dy\right)g(z)dz$$
$$= \int\frac{v(z)-\mu_v}{m(z)}\left[(K_b*g)(z) - g(z)\right]\left(\int\frac{v(y)-\mu_v}{m(y)}K_b(y-z)g(y)dy\right)g(z)dz$$
$$-\left(\int\frac{v(y)-\mu_v}{m(y)}g(y)^2dy\right)\left(\int\frac{v(z)-\mu_v}{m(z)}\left[(K_b*g)(z) - g(z)\right]g(z)dz\right).$$

But

$$\int \frac{v(y)-\mu_v}{m(y)} g(y)^2 dy = c\int \frac{f(y)}{g(y)}(v(y)-\mu_v)g(y)dy = c\int f(y)(v(y)-\mu_v)dy = 0,$$

and

$$\begin{aligned}\int \frac{v(y)-\mu_v}{m(y)} K_b(y-z)g(y)dy &= c\int \frac{v(y)-\mu_v}{g(y)} K_b(y-z)f(y)dy \\ &= c\int \gamma(y)K_b(y-z)dy = c\int \gamma(z+bt)K(t)dt \\ &= c\left[\gamma(z)+\frac{\mu_2(K)}{2}b^2\gamma''(z)+\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z)+O(b^6)\right], \qquad \text{(A.139)}\end{aligned}$$

and since

$$(K_b * g)(z) - g(z) = \frac{\mu_2(K)}{2}b^2 g''(z) + \frac{\mu_4(K)}{24}b^4 g^{(4)}(z) + O(b^6),$$

we conclude that

$$\begin{aligned}E\left[E\left(\tau_{12}^{\bullet}|Y_1\right)E\left(\tau_{31}^{\bullet}|Y_1\right)\right] &= c\int \frac{v(z)-\mu_v}{m(z)}\left[\frac{\mu_2(K)}{2}b^2 g''(z)+\frac{\mu_4(K)}{24}b^4 g^{(4)}(z)+O(b^6)\right] \\ &\cdot \left[\gamma(z)+\frac{\mu_2(K)}{2}b^2\gamma''(z)+\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z)+O(b^6)\right]g(z)dz \\ &= \frac{\mu_2(K)}{2}b^2c^2\int \alpha(z)g''(z)dz + \frac{\mu_4(K)}{24}b^4c^2\int \alpha(z)g^{(4)}(z)dz \\ &+ \frac{\mu_2(K)^2}{4}b^4c^2\int \gamma(z)\gamma''(z)g''(z)dz + O(b^6). \qquad \text{(A.140)}\end{aligned}$$

Using (A.138) and (A.140) we get (A.136). □

Lemma A.2.3. *The covariance of* $\tau_{12}^{\bullet}$ *and* $\tau_{32}^{\bullet}$ *is*

$$\begin{aligned}Cov(\tau_{12}^{\bullet},\tau_{32}^{\bullet}) &= c^2\int \beta(y)dy + \mu_2(K)b^2c^2\int \gamma''(y)f(y)(v(y)-\mu_v)dy \\ &+ b^4c^2\left[\frac{\mu_4(K)}{12}\int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy\right. \\ &+ \left.\frac{\mu_2(K)^2}{4}\left[\int \gamma''(y)^2 g(y)dy - \left(\int \gamma(y)g''(y)dy\right)^2\right]\right] + O(b^6). \qquad \text{(A.141)}\end{aligned}$$

Proof.

$$\begin{aligned}Cov(\tau_{12}^{\bullet},\tau_{32}^{\bullet}) &= E\left[Cov(\tau_{12}^{\bullet},\tau_{32}^{\bullet}|Y_2)\right] + Cov\left(E\left(\tau_{12}^{\bullet}|Y_2\right),E\left(\tau_{32}^{\bullet}|Y_2\right)\right) \\ &= Cov\left(E\left(\tau_{12}^{\bullet}|Y_2\right),E\left(\tau_{32}^{\bullet}|Y_2\right)\right) = E\left[E\left(\tau_{12}^{\bullet}|Y_2\right)E\left(\tau_{32}^{\bullet}|Y_2\right)\right] \\ &- E\left[E\left(\tau_{12}^{\bullet}|Y_2\right)\right]E\left[E\left(\tau_{32}^{\bullet}|Y_2\right)\right] = E\left[E\left(\tau_{12}^{\bullet}|Y_2\right)E\left(\tau_{32}^{\bullet}|Y_2\right)\right] - E(\tau_{12}^{\bullet})E(\tau_{32}^{\bullet}) \\ &= E\left[E\left(\tau_{12}^{\bullet}|Y_2\right)E\left(\tau_{32}^{\bullet}|Y_2\right)\right] - \left[E(\tau_{12}^{\bullet})\right]^2 = E\left[\left[E\left(\tau_{12}^{\bullet}|Y_2\right)\right]^2\right] - \left[E(\tau_{12}^{\bullet})\right]^2.\end{aligned}$$

But

$$\begin{aligned} E\left(\tau_{12}^{\bullet}|Y_2\right) &= E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_2)-g(Y_1))\middle|Y_2\right] \\ &= \int \frac{v(y)-\mu_v}{m(y)}\left[K_b(y-Y_2)-g(y)\right]g(y)dy \\ &= c\int \frac{v(y)-\mu_v}{g(y)}\left[K_b(y-Y_2)-g(y)\right]f(y)dy \\ &= c\int \gamma(y)\left[K_b(y-Y_2)-g(y)\right])dy \end{aligned}$$

and

$$E\left[\left[E\left(\tau_{12}^{\bullet}|Y_2\right)\right]^2\right] = E\left[\left(\int \frac{v(y)-\mu_v}{m(y)}\left[K_b(y-Y_2)-g(y)\right]g(y)dy\right)^2\right]$$

$$= E\left[c^2\int\int \gamma(y_1)\left(K_b(y_1-Y_2)-g(y_1)\right)\gamma(y_2)\left(K_b(y_2-Y_2)g(y_2)\right)dy_1dy_2\right]$$

$$= c^2\int\left(\int\int \gamma(y_1)\left(K_b(y_1-z)-g(y_1)\right)\gamma(y_2)\left(K_b(y_2-z)-g(y_2)\right)dy_1dy_2\right)g(z)dz$$

$$= c^2\int\left[\int \gamma(y)\left(K_b(y-z)-g(y)\right)dy\right]^2 g(z)dz,$$

where

$$\begin{aligned} \int \gamma(y)\left(K_b(y-z)-g(y)\right)dy &= \int \gamma(y)K_b(y-z)dy - \int \gamma(y)g(y)dy \\ = & \int \gamma(z+bt)K(t)dt - \int f(y)(v(y)-\mu_v)dy \\ = & \gamma(z) + \frac{\mu_2(K)}{2}b^2\gamma''(z) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z) + O(b^6). \end{aligned}$$

So,

$$E\left[\left[E\left(\tau_{12}^{\bullet}|Y_2\right)\right]^2\right] = c^2\int \gamma(z)^2 g(z)dz + \mu_2(K)b^2c^2\int \gamma(z)\gamma''(z)g(z)dz$$

$$\begin{aligned} + & \frac{\mu_2(K)^2}{4}b^4c^2\int \gamma''(z)^2g(z)dz + \frac{\mu_4(K)}{12}b^4c^2\int \gamma(z)\gamma^{(4)}(z)g(z)dz + O(b^6) \\ = & c^2\int \beta(y)dy + \mu_2(K)b^2c^2\int \gamma''(y)f(y)(v(y)-\mu_v)dy \\ + & \frac{\mu_2(K)^2}{4}b^4c^2\int \gamma''(y)^2g(y)dy + \frac{\mu_4(K)}{12}b^4c^2\int \gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy + O(b^6). \end{aligned}$$

Using now equation (A.138) for $E(\tau_{12}^{\bullet})$ gives (A.141). □

Lemma A.2.4. *The variance of* $\tau_{12}^{\bullet}$ *is*

$$\begin{aligned} Var(\tau_{12}^{\bullet}) &= \frac{\mu_0(K^2)}{b}c^2\int \alpha(y)dy - c^2\int \beta(y)dy + \frac{\mu_2(K^2)}{2}bc^2\int \delta(y)g''(y)dy \\ &- \mu_2(K)b^2c^2\int \alpha(y)g''(y)dy + O(b^3). \end{aligned} \tag{A.142}$$

Proof. Let us consider

$$Var(\tau^{\bullet}_{12}) = E(\tau^{\bullet 2}_{12}) - [E(\tau^{\bullet}_{12})]^2 ,$$

where

$$\begin{aligned}
E(\tau^{\bullet 2}_{12}) &= E\left(\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_2)-g(Y_1))\right]^2\right) \\
&= \int\int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 (K_b(y-z)-g(y))^2 g(y)g(z)dydz \\
&= \int\int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 K_b(y-z)^2 g(y)g(z)dydz \\
&- 2\int\int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 K_b(y-z)g(y)^2 g(z)dydz \\
&+ \int\int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 g(y)^3 g(z)dydz \\
&= \int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 g(y)\left(\int K_b(y-z)^2 g(z)dz\right)dy \\
&- 2\int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 g(y)^2\left(\int K_b(y-z)g(z)dz\right)dy \\
&+ \int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 g(y)^3 dy = c^2\int\delta(y)\left(\int\frac{1}{b}K(t)^2 g(y-bt)dt\right)dy \\
&- 2c^2\int\alpha(y)\left(\int K(t)g(y-bt)dt\right)dy + c^2\int\beta(y)dy + O(b^3) \\
&= \frac{\mu_0(K^2)}{b}c^2\int\alpha(y)dy + \frac{\mu_2(K^2)}{2}bc^2\int\delta(y)g''(y)dy \\
&- \mu_2(K)b^2c^2\int\alpha(y)g''(y)dy - c^2\int\beta(y)dy + O(b^3). \qquad \text{(A.143)}
\end{aligned}$$

Using (A.138) and (A.143) we obtain (A.142). □

Lemma A.2.5. *The covariance of* $\tau^{\bullet}_{12}$ *and* $\tau^{\bullet}_{21}$ *is*

$$\begin{aligned}
Cov(\tau^{\bullet}_{12},\tau^{\bullet}_{21}) &= \frac{\mu_0(K^2)}{b}c^2\int\alpha(y)dy - 2c^2\int\beta(y)dy \\
&+ \frac{\mu_2(K^2)}{2}bc^2\int\gamma(y)\gamma''(y)dy - \mu_2(K)b^2c^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy \\
&+ \frac{\mu_4(K^2)}{24}b^3c^2\int\gamma(y)\gamma^{(4)}(y)dy + O(b^4). \qquad \text{(A.144)}
\end{aligned}$$

Proof. Let us now consider $Cov(\tau^{\bullet}_{12},\tau^{\bullet}_{21})$:

$$Cov(\tau^{\bullet}_{12},\tau^{\bullet}_{21}) = E(\tau^{\bullet}_{12}\tau^{\bullet}_{21}) - E(\tau^{\bullet}_{12})E(\tau^{\bullet}_{21}) = E(\tau^{\bullet}_{12}\tau^{\bullet}_{21}) - [E(\tau^{\bullet}_{12})]^2 ,$$

where

$$
\begin{aligned}
E(\tau_{12}^{\bullet}\tau_{21}^{\bullet}) &= E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_2)-g(Y_1))\frac{v(Y_2)-\mu_v}{m(Y_2)}(K_b(Y_2-Y_1)-g(Y_2))\right] \\
= & \int\int \frac{v(y)-\mu_v}{m(y)}(K_b(y-z)-g(y))\frac{v(z)-\mu_v}{m(z)}(K_b(z-y)-g(z))g(y)g(z)dydz \\
= & \int\int \frac{v(y)-\mu_v}{m(y)}K_b(y-z)^2\frac{v(z)-\mu_v}{m(z)}g(y)g(z)dydz \\
- & \int\int \frac{v(y)-\mu_v}{m(y)}K_b(y-z)\frac{v(z)-\mu_v}{m(z)}g(y)g(z)^2dydz \\
- & \int\int \frac{v(y)-\mu_v}{m(y)}K_b(y-z)\frac{v(z)-\mu_v}{m(z)}g(y)^2g(z)dydz \\
+ & \int\int \frac{v(y)-\mu_v}{m(y)}\frac{v(z)-\mu_v}{m(z)}g(y)^2g(z)^2dydz = c^2\int\int K_b(y-z)^2\gamma(y)\gamma(z)dydz \\
- & 2c^2\int\int K_b(y-z)f(y)(v(y)-\mu_v)\gamma(z)dydz + c^2\left(\int f(y)(v(y)-\mu_v)dy\right)^2 \\
= & c^2\int\gamma(y)\left(\int K_b(y-z)^2\gamma(z)dz\right)dy \\
- & 2c^2\int f(y)(v(y)-\mu_v)\left(\int K_b(y-z)\gamma(z)dz\right)dy. \qquad \text{(A.145)}
\end{aligned}
$$

The last integral in (A.145) has been dealt with in equation (A.139):

$$
\begin{aligned}
\int K_b(y-z)\gamma(z)dz &= \int K_b(z-y)\gamma(z)dz \\
&= \gamma(y)+\frac{\mu_2(K)}{2}b^2\gamma''(y)+\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(y)+O(b^6),
\end{aligned}
$$

and consequently

$$
\begin{aligned}
& 2c^2\int f(y)(v(y)-\mu_v)\left(\int K_b(y-z)\gamma(z)dz\right)dy = 2c^2\int\beta(y)dy \\
+ & \mu_2(K)b^2c^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy \\
+ & \frac{\mu_4(K)}{12}b^4c^2\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy+O(b^6). \qquad \text{(A.146)}
\end{aligned}
$$

On the other hand, the first integral in (A.145) is

$$
\begin{aligned}
& \int\gamma(y)\left(\int K_b(y-z)^2\gamma(z)dz\right)dy = \int\gamma(y)\left(\int\frac{1}{b}K(t)^2\gamma(y-bt)dt\right)dy \\
= & \int\gamma(y)\left[\frac{\mu_0(K^2)}{b}\gamma(y)+\frac{\mu_2(K^2)}{2}b\gamma''(y)+\frac{\mu_4(K^2)}{24}b^3\gamma^{(4)}(y)+O(b^5)\right]dy \\
= & \frac{\mu_0(K^2)}{b}\int\alpha(y)dy+\frac{\mu_2(K^2)}{2}b\int\gamma(y)\gamma''(y)dy \\
+ & \frac{\mu_4(K^2)}{24}b^3\int\gamma(y)\gamma^{(4)}(y)dy+O(b^5). \qquad \text{(A.147)}
\end{aligned}
$$

Now, using (A.146) and (A.147) in (A.145) we get

$$\begin{aligned} E(\tau^\bullet_{12}\tau^\bullet_{21}) &= \frac{\mu_0(K^2)}{b}c^2\int\alpha(y)dy - 2c^2\int\beta(y)dy + \frac{\mu_2(K^2)}{2}bc^2\int\gamma(y)\gamma''(y)dy \\ &- \mu_2(K)b^2c^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy + \frac{\mu_4(K^2)}{24}b^3c^2\int\gamma(y)\gamma^{(4)}(y)dy \\ &- \frac{\mu_4(K)}{12}b^4c^2\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy + O(b^5). \qquad \text{(A.148)} \end{aligned}$$

As a consequence of expression (A.138), we have:

$$E(\tau^\bullet_{12}) = \frac{\mu_2(K)}{2}b^2c\int\gamma(y)g''(y)dy + \frac{\mu_4(K)}{24}b^4c\int\gamma(y)g^{(4)}(y)dy + O(b^6).$$

Using the previous expression and expression (A.148), we have (A.144). □

Lemma A.2.6. *The covariance of* $\tau^\bullet_{12}$ *and* $\tau^\bullet_{11}$ *is*

$$\begin{aligned} Cov(\tau^\bullet_{12},\tau^\bullet_{11}) &= \frac{\mu_2(K)K(0)}{2}bc^2\left[\int\delta(y)g''(y)dy - \left(\int\gamma(y)g''(y)dy\right)\left(\int\gamma(y)dy\right)\right] \\ &- \frac{\mu_2(K)}{2}b^2c^2\int\alpha(y)g''(y)dy + \frac{\mu_4(K)K(0)}{24}b^3c^2\left[\int\delta(y)g^{(4)}(y)dy\right. \qquad \text{(A.149)} \\ &- \left.\left(\int\gamma(y)g^{(4)}(y)dy\right)\left(\int\gamma(y)dy\right)\right] - \frac{\mu_4(K)}{24}b^4c^2\int\alpha(y)g^{(4)}(y)dy + O(b^5). \end{aligned}$$

Proof. Let's deal now with the term:

$$\begin{aligned} Cov(\tau^\bullet_{12},\tau^\bullet_{11}) &= Cov\left(E\left(\tau^\bullet_{12}|Y_1\right), E\left(\tau^\bullet_{11}|Y_1\right)\right) + E\left[Cov(\tau^\bullet_{12},\tau^\bullet_{11}|Y_1)\right] \\ &= Cov\left(E\left(\frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_2)-g(Y_1))\middle| Y_1\right), \tau^\bullet_{11}\right), \end{aligned}$$

since

$$\tau^\bullet_{11} = \frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_1)-g(Y_1)) = \frac{v(Y_1)-\mu_v}{m(Y_1)}\left(\frac{K(0)}{b}-g(Y_1)\right) \qquad \text{(A.150)}$$

is a measurable function of Y_1 and then $Cov(\tau^\bullet_{12},\tau^\bullet_{11}|Y_1) = 0$.

On the other hand,

$$\begin{aligned} & E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_2)-g(Y_1))\middle| Y_1\right] \\ &= \frac{v(Y_1)-\mu_v}{m(Y_1)}(E(K_b(Y_1-Y_2)|Y_1)-g(Y_1)) \\ &= \frac{v(Y_1)-\mu_v}{m(Y_1)}\left[(K_b * g)(Y_1)-g(Y_1)\right]. \end{aligned}$$

Using the previous expressions we can obtain a simpler expression for $Cov(\tau_{12}^{\bullet},\tau_{11}^{\bullet})$:

$$
\begin{aligned}
Cov(\tau_{12}^{\bullet},\tau_{11}^{\bullet}) &= E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}\left((K_b*g)(Y_1)-g(Y_1)\right)\frac{v(Y_1)-\mu_v}{m(Y_1)}\left(\frac{K(0)}{b}-g(Y_1)\right)\right] \\
&- E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}\left((K_b*g)(Y_1)-g(Y_1)\right)\right]E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}\left(\frac{K(0)}{b}-g(Y_1)\right)\right] \\
&= \int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2\left((K_b*g)(y)-g(y)\right)\left(\frac{K(0)}{b}-g(y)\right)g(y)dy \\
&- \left[\frac{\mu_2(K)}{2}b^2c\int\gamma(y)g''(y)dy+\frac{\mu_4(K)}{24}b^4c\int\gamma(y)g^{(4)}(y)dy+O(b^6)\right] \\
&\cdot \left[\frac{K(0)}{b}E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}\right]-cE\left[\gamma(Y_1)\right]\right]. \qquad \text{(A.151)}
\end{aligned}
$$

But

$$
\begin{aligned}
\int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2 & \left((K_b*g)(y)-g(y)\right)\left(\frac{K(0)}{b}-g(y)\right)g(y)dy \\
&= \int\left(\frac{v(y)-\mu_v}{m(y)}\right)^2\left[\frac{\mu_2(K)}{2}b^2g''(y)+\frac{\mu_4(K)}{24}b^4g^{(4)}(y)+O(b^6)\right] \\
&\cdot \left(\frac{K(0)}{b}-g(y)\right)g(y)dy=\frac{\mu_2(K)K(0)}{2}bc^2\int\delta(y)g''(y)dy \\
&- \frac{\mu_2(K)}{2}b^2c^2\int\alpha(y)g''(y)dy+\frac{\mu_4(K)K(0)}{24}b^3c^2\int\delta(y)g^{(4)}(y)dy \\
&- \frac{\mu_4(K)}{24}b^4c^2\int\alpha(y)g^{(4)}(y)dy+O(b^5) \qquad \text{(A.152)}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{K(0)}{b}E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}\right] &- cE\left[\gamma(Y_1)\right]=\frac{K(0)}{b}\int\frac{v(y)-\mu_v}{m(y)}g(y)dy \\
&- c\int f(y)(v(y)-\mu_v)dy=\frac{K(0)}{b}c\int\gamma(y)dy. \qquad \text{(A.153)}
\end{aligned}
$$

Using (A.152) and (A.153) in (A.151) we get (A.149). □

Lemma A.2.7. *The covariance of $\tau_{12}^{\bullet}$ and $\tau_{22}^{\bullet}$ is*

$$
\begin{aligned}
Cov(\tau_{12}^{\bullet},\tau_{22}^{\bullet}) &= \frac{K(0)}{b}c^2\int\alpha(y)dy-c^2\int\beta(y)dy \\
&+ \frac{\mu_2(K)K(0)}{2}bc^2\left[\int\gamma(y)\gamma''(y)dy-\left(\int\gamma''(y)g(y)dy\right)\left(\int\gamma(y)dy\right)\right] \\
&- \frac{\mu_2(K)}{2}b^2c^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy+\frac{\mu_4(K)K(0)}{24}b^3c^2\left[\int\gamma(y)\gamma^{(4)}(y)dy\right. \\
&- \left.\left(\int\gamma^{(4)}(y)g(y)dy\right)\cdot\left(\int\gamma(y)dy\right)\right] \\
&- \frac{\mu_4(K)}{24}b^4c^2\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy+O(b^5). \qquad \text{(A.154)}
\end{aligned}
$$

Proof.

$$\begin{aligned} Cov(\tau_{12}^{\bullet},\tau_{22}^{\bullet}) &= Cov\left(E\left(\tau_{12}^{\bullet}|Y_2\right), E\left(\tau_{22}^{\bullet}|Y_2\right)\right) + E\left[Cov(\tau_{12}^{\bullet},\tau_{22}^{\bullet}|Y_2)\right] \\ &= Cov\left(E\left(\frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_2)-g(Y_1))\bigg|Y_2\right), \tau_{22}^{\bullet}\right), \end{aligned}$$

since $Cov(\tau_{12}^{\bullet},\tau_{22}^{\bullet}|Y_2)=0$ (because $\tau_{22}^{\bullet}$ is a measurable function of Y_2).

On the other hand,

$$\begin{aligned} &E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}(K_b(Y_1-Y_2)-g(Y_1))\bigg|Y_2\right] = E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}K_b(Y_1-Y_2)\bigg|Y_2\right] \\ -&\quad E\left[\gamma(Y_1)|Y_2\right] = \int \frac{v(y)-\mu_v}{m(y)}K_b(y-Y_2)g(y)dy - E\left[\gamma(Y_1)\right] \\ =&\quad c\int \gamma(y)K_b(Y_2-y)dy - c\int f(y)(v(y)-\mu_v)dy = c\cdot(\gamma * K_b)(Y_2). \end{aligned}$$

Using the previous expressions and $\tau_{22}^{\bullet} = \frac{v(Y_2)-\mu_v}{m(Y_2)}\left(\frac{K(0)}{b}-g(Y_2)\right)$, we have:

$$\begin{aligned} Cov(\tau_{12}^{\bullet},\tau_{22}^{\bullet}) &= Cov\left(c\cdot(\gamma * K_b)(Y_2), \frac{v(Y_2)-\mu_v}{m(Y_2)}\left(\frac{K(0)}{b}-g(Y_2)\right)\right) \\ &= E\left[c\cdot(\gamma * K_b)(Y_2)\frac{v(Y_2)-\mu_v}{m(Y_2)}\left(\frac{K(0)}{b}-g(Y_2)\right)\right] \\ &- E\left[c\cdot(\gamma * K_b)(Y_2)\right]E\left[\frac{v(Y_2)-\mu_v}{m(Y_2)}\left(\frac{K(0)}{b}-g(Y_2)\right)\right], \quad \text{(A.155)} \end{aligned}$$

where

$$\begin{aligned} &E\left[c\cdot(\gamma * K_b)(Y_2)\frac{v(Y_2)-\mu_v}{m(Y_2)}\left(\frac{K(0)}{b}-g(Y_2)\right)\right] \\ =&\quad c\int(\gamma * K_b)(y)\frac{v(y)-\mu_v}{m(y)}\left(\frac{K(0)}{b}-g(y)\right)g(y)dy \end{aligned}$$

with

$$\begin{aligned} (\gamma * K_b)(y) &= \int \gamma(z)K_b(y-z)dz = \int \gamma(y-bt)K(t)dt \\ &= \gamma(y) + \frac{\mu_2(K)}{2}b^2\gamma''(y) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(y) + O(b^6). \end{aligned}$$

Consequently,

$$\begin{aligned}
&E\left[c\cdot(\gamma * K_b)(Y_2)\frac{v(Y_2)-\mu_v}{m(Y_2)}\left(\frac{K(0)}{b}-g(Y_2)\right)\right] \\
=\;& c\int\gamma(y)\frac{v(y)-\mu_v}{m(y)}\left(\frac{K(0)}{b}-g(y)\right)g(y)dy \\
+\;& \frac{\mu_2(K)}{2}b^2c\int\gamma''(y)\frac{v(y)-\mu_v}{m(y)}\left(\frac{K(0)}{b}-g(y)\right)g(y)dy \\
+\;& \frac{\mu_4(K)}{24}b^4c\int\gamma^{(4)}(y)\frac{v(y)-\mu_v}{m(y)}\left(\frac{K(0)}{b}-g(y)\right)g(y)dy+O(b^6) \\
=\;& \frac{K(0)}{b}c^2\int\alpha(y)dy-c^2\int\beta(y)dy+\frac{\mu_2(K)K(0)}{2}bc^2\int\gamma(y)\gamma''(y)dy \\
-\;& \frac{\mu_2(K)}{2}b^2c^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy+\frac{\mu_4(K)K(0)}{24}b^3c^2\int\gamma(y)\gamma^{(4)}(y)dy \\
-\;& \frac{\mu_4(K)}{24}b^4c^2\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy+O(b^6) \qquad \text{(A.156)}
\end{aligned}$$

and

$$\begin{aligned}
&E\left[c\cdot(\gamma * K_b)(Y_2)\right]=c\int(\gamma * K_b)(y)g(y)dy=c\int\gamma(y)g(y)dy \\
+\;& \frac{\mu_2(K)}{2}b^2c\int\gamma''(y)g(y)dy+\frac{\mu_4(K)}{24}b^4c\int\gamma^{(4)}(y)g(y)dy+O(b^6) \\
=\;& c\int f(y)(v(y)-\mu_v)dy+\frac{\mu_2(K)}{2}b^2c\int\gamma''(y)g(y)dy \\
+\;& \frac{\mu_4(K)}{24}b^4c\int\gamma^{(4)}(y)g(y)dy+O(b^6) \\
=\;& \frac{\mu_2(K)}{2}b^2c\int\gamma''(y)g(y)dy+\frac{\mu_4(K)}{24}b^4c\int\gamma^{(4)}(y)g(y)dy+O(b^6). \qquad \text{(A.157)}
\end{aligned}$$

The last factor in (A.155) is exactly the same as the equation (A.153). Now, using (A.156), (A.157) and (A.153) in (A.155) results in (A.154). □

Lemma A.2.8. *The variance of* $\tau^{\bullet}_{11}$ *is*

$$\begin{aligned}
Var(\tau^{\bullet}_{11}) \;=\;& \frac{K(0)^2}{b^2}c^2\left[\int\delta(y)dy-\left(\int\gamma(y)dy\right)^2\right] \\
-\;& \frac{2K(0)}{b}c^2\int\alpha(y)dy+c^2\int\beta(y)dy. \qquad \text{(A.158)}
\end{aligned}$$

Proof. Let us consider

$$Var(\tau^{\bullet}_{11})=E(\tau^{\bullet 2}_{11})-E(\tau^{\bullet}_{11})^2,$$

where the last term can be directly obtained from equation (A.153):

$$\begin{aligned} E(\tau_{11}^{\bullet})^2 &= \left[E\left(\frac{v(Y_1)-\mu_v}{m(Y_1)}\left(\frac{K(0)}{b}-g(Y_1)\right)\right)\right]^2 \\ &= \frac{K(0)^2}{b^2}c^2\left(\int\gamma(y)dy\right)^2. \end{aligned} \tag{A.159}$$

For the term $E(\tau_{11}^{\bullet 2})$ we use the expression (A.150) for $\tau_{11}^{\bullet}$ to get:

$$\begin{aligned} E(\tau_{11}^{\bullet 2}) = E&\left[\left(\frac{v(Y_1)-\mu_v}{m(Y_1)}\right)^2\left(\frac{K(0)}{b}-g(Y_1)\right)^2\right] \\ &= \frac{K(0)^2}{b^2}E\left[\left(\frac{v(Y_1)-\mu_v}{m(Y_1)}\right)^2\right]-2\frac{K(0)}{b}c^2E\left[\delta(Y_1)\right]+c^2E\left[\alpha(Y_1)\right] \\ &= \frac{K(0)^2}{b^2}c^2\int\delta(y)dy-\frac{2K(0)}{b}c^2\int\alpha(y)dy+c^2\int\beta(y)dy. \end{aligned} \tag{A.160}$$

Using (A.159) and (A.160) we obtain (A.158). □

We will proceed now with the covariance terms in (5.10).

Lemma 5.5.7. *The covariance of $\widehat{A}_1^{\bullet}$ and $\widehat{A}_2^{\bullet}$ is*

$$Cov\left(\widehat{A}_1^{\bullet},\widehat{A}_2^{\bullet}\right) = D_{27}^{\bullet}\frac{h^2}{N}+D_{28}^{\bullet}\frac{h^4}{N}+O\left(\frac{h^6}{N}\right), \tag{5.14}$$

where

$$\begin{aligned} D_{27}^{\bullet} &:= \frac{\mu_2(K)}{2}c^3\int\psi(y)m''(y)dy, \\ D_{28}^{\bullet} &:= \frac{\mu_4(K)}{24}c^3\int\psi(y)m^{(4)}(y)dy. \end{aligned}$$

Proof. Let us define

$$\omega_i^{\bullet} := \frac{g(Y_i)}{m(Y_i)}(v(Y_i)-\mu_v) = c\cdot\omega_i,$$

where

$$\omega_i := \frac{f(Y_i)}{g(Y_i)}(v(Y_i)-\mu_v).$$

Using this definition and the expressions for $\widehat{A}_1^{\bullet}$ we have

$$\widehat{A}_1^{\bullet} = \frac{1}{N}\sum_{i=1}^{N}\omega_i^{\bullet}.$$

Let us first consider $Cov(\widehat{A}_1^\bullet, \widehat{A}_2^\bullet)$:

$$\begin{aligned}
Cov(\widehat{A}_1^\bullet, \widehat{A}_2^\bullet) &= Cov\left(\frac{1}{N}\sum_{i=1}^{N}\omega_i^\bullet, \frac{1}{Nn}\sum_{j=1}^{N}\sum_{k=1}^{n}\eta_{jk}^\bullet\right) \\
&= \frac{1}{N^2 n}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{n} Cov(\omega_i^\bullet, \eta_{jk}^\bullet) = \frac{1}{N} Cov(\omega_1^\bullet, \eta_{11}^\bullet), \quad \text{(A.161)}
\end{aligned}$$

since $Cov(\omega_i^\bullet, \eta_{jk}^\bullet) = 0$ for $i \neq j$ because $\omega_i^\bullet$ and $\eta_{jk}^\bullet$ are independent for $i \neq j$.

On the other hand,

$$\begin{aligned}
Cov(\omega_1^\bullet, \eta_{11}^\bullet) &= Cov\left(c\gamma(Y_1), \frac{g(Y_1)\left(K_h(Y_1 - X_1) - m(Y_1)\right)}{m(Y_1)^2}(v(Y_1) - \mu_v)\right) \\
&= E\left[\frac{c\gamma(Y_1)g(Y_1)}{m(Y_1)^2}\left(K_h(Y_1 - X_1) - m(Y_1)\right)(v(Y_1) - \mu_v)\right] \\
&- cE\left(\gamma(Y_1)\right)E\left[\frac{g(Y_1)\left(K_h(Y_1 - X_1) - m(Y_1)\right)}{m(Y_1)^2}(v(Y_1) - \mu_v)\right] \\
&= c^3 E\left[\frac{\gamma(Y_1)\Omega(Y_1)}{g(Y_1)}\left(K_h(Y_1 - X_1) - m(Y_1)\right)\right],
\end{aligned}$$

since

$$E\left(\gamma(Y_1)\right) = \int \gamma(y)g(y)dy = \int f(y)(v(y) - \mu_v) = 0$$

and

$$\begin{aligned}
&E\left[\frac{\gamma(Y_1)\Omega(Y_1)}{g(Y_1)}\left(K_h(Y_1 - X_1) - m(Y_1)\right)\right] \\
&= E\left[E\left[\frac{\gamma(Y_1)\Omega(Y_1)}{g(Y_1)}\left(K_h(Y_1 - X_1) - m(Y_1)\right)|Y_1\right]\right] \\
&= E\left[\frac{\gamma(Y_1)\Omega(Y_1)}{g(Y_1)}\left(E\left[K_h(Y_1 - X_1)|Y_1\right] - m(Y_1)\right)\right] \\
&= E\left[\frac{\gamma(Y_1)\Omega(Y_1)}{g(Y_1)}\left((K_h * m)(Y_1) - m(Y_1)\right)\right] \\
&= \int \frac{\gamma(y)\Omega(y)}{g(y)}\left((K_h * m)(y) - m(y)\right)g(y)dy \\
&= \int \gamma(y)\Omega(y)\left((K_h * m)(y) - m(y)\right)dy \\
&= \int \gamma(y)\Omega(y)\left(\frac{h^2}{2}\mu_2(K)m''(y) + \frac{h^4}{24}\mu_4(K)m^{(4)}(y) + O(h^6)\right)dy \\
&= \int \psi(y)\left(\frac{h^2}{2}\mu_2(K)m''(y) + \frac{h^4}{24}\mu_4(K)m^{(4)}(y) + O(h^6)\right)dy \\
&= \frac{h^2}{2}\mu_2(K)\int \psi(y)m''(y)dy + \frac{h^4}{24}\mu_4(K)\int \psi(y)m^{(4)}(y)dy + O(h^6).
\end{aligned}$$

Consequently,

$$\begin{aligned} Cov(\omega_1^\bullet, \eta_{11}^\bullet) &= \frac{h^2}{2}\mu_2(K)c^3 \int \psi(y)m''(y)dy \\ &+ \frac{h^4}{24}\mu_4(K)c^3 \int \psi(y)m^{(4)}(y)dy + O(h^6) \end{aligned}$$

Using the previous expression in (A.161) gives (5.14). □

Lemma 5.5.8. *The covariance of* $\widehat{A}_1^\bullet$ *and* $\widehat{A}_3^\bullet$ *is*

$$\begin{aligned} Cov\left(\widehat{A}_1^\bullet, \widehat{A}_3^\bullet\right) &= D_{29}^\bullet \frac{1}{N} + D_{30}^\bullet \frac{1}{N^2 b} + D_{31}^\bullet \frac{1}{N^2} + D_{32}^\bullet \frac{b^2}{N} \\ &+ D_{33}^\bullet \frac{b^4}{N} + O\left(\frac{b^2}{N^2}\right) + O\left(\frac{b^6}{N}\right), \end{aligned} \tag{5.15}$$

where

$$\begin{aligned} D_{29}^\bullet &:= c^2 \int \beta(y)dy = D_6^\bullet, \\ D_{30}^\bullet &:= K(0)c^2 \int \alpha(y)dy, \\ D_{31}^\bullet &:= -2c^2 \int \beta(y)dy = -2D_6^\bullet, \\ D_{32}^\bullet &:= \frac{\mu_2(K)}{2}c^2 \left[\int \alpha(y)g''(y)dy + \int \gamma''(y)f(y)(v(y) - \mu_v)dy\right], \\ D_{33}^\bullet &:= \frac{\mu_4(K)}{24}c^2 \left[\int \alpha(y)g^{(4)}(y)dy + \int \gamma^{(4)}(y)f(y)(v(y) - \mu_v)dy\right]. \end{aligned}$$

Proof. Let us now consider $Cov(\widehat{A}_1^\bullet, \widehat{A}_3^\bullet)$:

$$\begin{aligned} Cov(\widehat{A}_1^\bullet, \widehat{A}_3^\bullet) = Cov\left(\frac{1}{N}\sum_{i=1}^N \omega_i^\bullet, \frac{1}{N^2}\sum_{j=1}^N\sum_{k=1}^N \tau_{jk}^\bullet\right) = \frac{1}{N^3}\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N Cov(\omega_i^\bullet, \tau_{jk}^\bullet) \\ = \frac{N-1}{N^2}Cov(\omega_1^\bullet, \tau_{12}^\bullet) + \frac{N-1}{N^2}Cov(\omega_1^\bullet, \tau_{21}^\bullet) + \frac{1}{N^2}Cov(\omega_1^\bullet, \tau_{11}^\bullet), \end{aligned} \tag{A.162}$$

since $\omega_1^\bullet$ and $\tau_{23}^\bullet$ are independent and $\omega_1^\bullet$ and $\tau_{22}^\bullet$ are also independent.

Let us study now the terms in (A.162):

$$Cov(\omega_1^\bullet, \tau_{12}^\bullet) = E\left[Cov(\omega_1^\bullet, \tau_{12}^\bullet|Y_1)\right] + Cov(E(\omega_1^\bullet|Y_1), E(\tau_{12}^\bullet|Y_1)).$$

But $Cov(\omega_1^\bullet, \tau_{12}^\bullet|Y_1) = 0$ since $\omega_1^\bullet$ is a function of Y_1.

On the other hand, $E(\omega_1^\bullet|Y_1) = \omega_1^\bullet = c \cdot \gamma(Y_1)$, and

$$E(\tau_{12}^\bullet|Y_1) = E\left[\frac{v(Y_1)-\mu_v}{m(Y_1)}\left(K_b(Y_1-Y_2)-g(Y_1)\right)|Y_1\right]$$
$$= \frac{v(Y_1)-\mu_v}{m(Y_1)}\left(E\left[K_b(Y_1-Y_2)|Y_1\right]-g(Y_1)\right) = \frac{v(Y_1)-\mu_v}{m(Y_1)}\left((K_b*g)(Y_1)-g(Y_1)\right)$$

and

$$E\left(\omega_1^\bullet\right) = cE\left(\gamma(Y_1)\right) = c\int\gamma(y)g(y)dy = c\int f(y)(v(y)-\mu_v) = 0.$$

Thus

$$\begin{aligned}
Cov(\omega_1^\bullet,\tau_{12}^\bullet) &= Cov(E(\omega_1^\bullet|Y_1),E(\tau_{12}^\bullet|Y_1)) = E\left[E(\omega_1^\bullet|Y_1)E(\tau_{12}^\bullet|Y_1)\right]\\
&- E\left[E(\omega_1^\bullet|Y_1)\right]E\left[E(\tau_{12}^\bullet|Y_1)\right] = E\left[\omega_1^\bullet E(\tau_{12}^\bullet|Y_1)\right]-E(\omega_1^\bullet)E(\tau_{12}^\bullet)\\
&= cE\left[\gamma(Y_1)\frac{v(Y_1)-\mu_v}{m(Y_1)}\left((K_b*g)(Y_1)-g(Y_1)\right)\right]\\
&= c^2\int\alpha(y)\left((K_b*g)(y)-g(y)\right)dy\\
&= c^2\int\alpha(y)\left(\frac{\mu_2(K)}{2}b^2g''(y)+\frac{\mu_4(K)}{24}b^4g^{(4)}(y)+O(b^6)\right)dy\\
&= \frac{\mu_2(K)}{2}b^2c^2\int\alpha(y)g''(y)dy\\
&+ \frac{\mu_4(K)}{24}b^4c^2\int\alpha(y)g^{(4)}(y)dy+O(b^6). \qquad (A.163)
\end{aligned}$$

Now consider

$$\begin{aligned}
Cov(\omega_1^\bullet,\tau_{21}^\bullet) &= E\left[Cov(\omega_1^\bullet,\tau_{21}^\bullet|Y_1)\right]+Cov(E(\omega_1^\bullet|Y_1),E(\tau_{21}^\bullet|Y_1))\\
&= Cov(\omega_1^\bullet,E(\tau_{21}^\bullet|Y_1)) = E\left[\omega_1^\bullet E(\tau_{21}^\bullet|Y_1)\right]-E(\omega_1^\bullet)E\left[E(\tau_{21}^\bullet|Y_1)\right]\\
&= E\left[\omega_1^\bullet E(\tau_{21}^\bullet|Y_1)\right].
\end{aligned}$$

But

$$\begin{aligned}
E(\tau_{21}^\bullet|Y_1) &= E\left[\frac{v(Y_2)-\mu_v}{m(Y_2)}\left(K_b(Y_2-Y_1)-g(Y_2)\right)\middle|Y_1\right]\\
&= c\int\gamma(y)\left(K_b(y-Y_1)-g(y)\right)dy.
\end{aligned}$$

Thus

$$\begin{aligned}
Cov(\omega_1^\bullet,\tau_{21}^\bullet) &= c^2\int\gamma(z)\left(\int\gamma(y)\left(K_b(y-z)-g(y)\right)dy\right)g(z)dz\\
&= c^2\int\int\left(K_b(y-z)-g(y)\right)\gamma(y)f(z)(v(z)-\mu_v)dydz\\
&= c^2\int\int K_b(y-z)\gamma(y)f(z)(v(z)-\mu_v)dydz\\
&- c^2\int\int f(y)(v(y)-\mu_v)f(z)(v(z)-\mu_v)dydz. \qquad (A.164)
\end{aligned}$$

But

$$\int\int f(y)(v(y)-\mu_v)f(z)(v(z)-\mu_v)dydz = \left[\int f(y)(v(y)-\mu_v)dy\right]^2 = 0 \quad (A.165)$$

and

$$\begin{aligned}
&\int\int K_b(y-z)\gamma(y)f(z)(v(z)-\mu_v)dydz \\
= \; & \int\left(\int K(t)\gamma(z+bt)dt\right)f(z)(v(z)-\mu_v)dz \\
= \; & \int\left(\gamma(z)+\frac{\mu_2(K)}{2}b^2\gamma''(z)+\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(z)+O(b^6)\right)f(z)(v(z)-\mu_v)dz \\
= \; & \int\beta(z)dz+\frac{\mu_2(K)}{2}b^2\int\gamma''(z)f(z)(v(z)-\mu_v)dz \\
+ \; & \frac{\mu_4(K)}{24}b^4\int\gamma^{(4)}(z)f(z)(v(z)-\mu_v)dz+O(b^6). \qquad (A.166)
\end{aligned}$$

Using (A.165) and (A.166) in (A.164) we get:

$$\begin{aligned}
Cov(\omega_1^\bullet,\tau_{21}^\bullet) \; = \; & c^2\int\beta(y)dy+\frac{\mu_2(K)}{2}b^2c^2\int\gamma''(y)f(y)(v(y)-\mu_v)dy \\
+ \; & \frac{\mu_4(K)}{24}b^4c^2\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)dy+O(b^6). \qquad (A.167)
\end{aligned}$$

The last term in (A.161) is:

$$\begin{aligned}
&Cov(\omega_1^\bullet,\tau_{11}^\bullet) = E\left(\omega_1^\bullet\tau_{11}^\bullet\right)-E(\omega_1^\bullet)E(\tau_{11}^\bullet) = E\left(\omega_1^\bullet\tau_{11}^\bullet\right) \\
= \; & cE\left[\gamma(Y_1)\frac{v(Y_1)-\mu_v}{m(Y_1)}\left(K_b(Y_1-Y_1)-g(Y_1)\right)\right] \\
= \; & c^2\int\alpha(y)\left(K_b(0)-g(y)\right)dy = \frac{K(0)}{b}c^2\int\alpha(y)dy-c^2\int\beta(y)dy. \; (A.168)
\end{aligned}$$

Using (A.163), (A.167) and (A.168) in (A.162) gives (5.15). □

Lemma 5.5.9. *The covariance of $\widehat{A}_2^\bullet$ and $\widehat{A}_3^\bullet$ is*

$$\begin{aligned}
&Cov\left(\widehat{A}_2^\bullet,\widehat{A}_3^\bullet\right) = D_{34}^\bullet\frac{h^2}{N}+D_{35}^\bullet\frac{h^2b^2}{N}+D_{36}^\bullet\frac{h^4}{N}+D_{37}^\bullet\frac{h^4b^2}{N}+D_{38}^\bullet\frac{h^2b^4}{N} \\
+ \; & D_{39}^\bullet\frac{h^2}{N^2b}+D_{40}^\bullet\frac{h^2}{N^2}+D_{41}^\bullet\frac{h^4}{N^2b}+D_{42}^\bullet\frac{h^4}{N^2}+D_{43}^\bullet\frac{h^2b^2}{N^2}+O\left(\frac{h^6}{N}\right) \\
+ \; & O\left(\frac{h^2b^6}{N}\right)+O\left(\frac{h^4b^4}{N}\right)+O\left(\frac{h^6}{N^2b}\right)+O\left(\frac{h^2b^4}{N^2}\right)+O\left(\frac{h^4b^2}{N^2}\right), \; (5.16)
\end{aligned}$$

where

$$\begin{aligned}
D^{\bullet}_{34} &:= \frac{\mu_2(K)}{2}c^3\int \psi(y)m''(y)dy = D^{\bullet}_{27},\\
D^{\bullet}_{35} &:= \frac{\mu_2(K)^2}{4}c^3\left[\int \zeta(y)g''(y)m''(y)dy + \int m''(y)\gamma''(y)\Omega(y)dy\right.\\
&- \left. 2B^{\bullet}(\Omega'')\int \gamma(y)g''(y)dy\right],\\
D^{\bullet}_{36} &:= \frac{\mu_4(K)}{24}c^3\int \psi(y)m^{(4)}(y)dy = D^{\bullet}_{28},
\end{aligned}$$

$$\begin{aligned}
D^{\bullet}_{37} &:= \frac{\mu_2(K)\mu_4(K)}{48}c^3\left[\int \zeta(y)m^{(4)}(y)g''(y)dy + \int m^{(4)}(y)\gamma''(y)\Omega(y)dy\right.\\
&- \left. 2B^{\bullet}(\Omega^{(4)})\int \gamma(y)g''(y)dy\right],\\
D^{\bullet}_{38} &:= \frac{\mu_2(K)\mu_4(K)}{48}c^3\left[\int \zeta(y)(y)m''(y)g^{(4)}(y)dy + \int m''(y)\gamma^{(4)}(y)\Omega(y)dy\right.\\
&- \left. 2B^{\bullet}(\Omega'')\int \gamma(y)g^{(4)}(y)dy\right],\\
D^{\bullet}_{39} &:= c^3\frac{\mu_2(K)K(0)}{2}\left[\int \zeta(y)m''(y)dy - B^{\bullet}(\Omega'')\int \gamma(y)dy\right],\\
D^{\bullet}_{40} &:= -2D^{\bullet}_{34},\\
D^{\bullet}_{41} &:= \frac{\mu_4(K)K(0)}{24}c^3\left[\int \zeta(y)m^{(4)}(y)dy - B^{\bullet}(\Omega^{(4)})\int \gamma(y)dy\right],\\
D^{\bullet}_{42} &:= -2D^{\bullet}_{36},\\
D^{\bullet}_{43} &:= -D^{\bullet}_{35},
\end{aligned}$$

with

$$\zeta(y) := \frac{f(y)^3}{g(y)^4}(v(y)-\mu_v)^2.$$

Proof. Let us consider $Cov(\widehat{A}^{\bullet}_2, \widehat{A}^{\bullet}_3)$:

$$\begin{aligned}
Cov(\widehat{A}^{\bullet}_2, \widehat{A}^{\bullet}_3) &= Cov\left(\frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}\eta^{\bullet}_{ij}, \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\tau^{\bullet}_{kl}\right)\\
&= \frac{1}{N^3 n}\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{N}Cov(\eta^{\bullet}_{ij}, \tau^{\bullet}_{kl})\\
&= \frac{N-1}{N^2}Cov(\eta^{\bullet}_{11}, \tau^{\bullet}_{12}) + \frac{N-1}{N^2}Cov(\eta^{\bullet}_{11}, \tau^{\bullet}_{21}) + \frac{1}{N^2}Cov(\eta^{\bullet}_{11}, \tau^{\bullet}_{11}), \quad \text{(A.169)}
\end{aligned}$$

since $\eta^{\bullet}_{11}$ only depends on Y_1 and X_1 and $\tau^{\bullet}_{23}$ only depends on Y_2 and Y_3, then $\eta^{\bullet}_{11}$ and $\tau^{\bullet}_{23}$ are independent. Similiarly $\eta^{\bullet}_{11}$ and $\tau^{\bullet}_{22}$ are also independent. Consequently:

$$Cov(\eta^{\bullet}_{11}, \tau^{\bullet}_{23}) = 0 \quad \text{and} \quad Cov(\eta^{\bullet}_{11}, \tau^{\bullet}_{22}) = 0.$$

Let us now consider the three terms in (A.169):

$$\begin{aligned} Cov(\eta_{11}^{\bullet}, \tau_{12}^{\bullet}) &= E\left[Cov(\eta_{11}^{\bullet}, \tau_{12}^{\bullet}|Y_1)\right] + Cov(E(\eta_{11}^{\bullet}|Y_1), E(\tau_{12}^{\bullet}|Y_1)) \\ &= Cov(E(\eta_{11}^{\bullet}|Y_1), E(\tau_{12}^{\bullet}|Y_1)), \end{aligned}$$

since $Cov(\eta_{11}^{\bullet}, \tau_{12}^{\bullet}|Y_1) = 0$ because $\eta_{11}^{\bullet}$ and $\tau_{12}^{\bullet}$ are conditionally independent given Y_1 ($\eta_{11}^{\bullet}$ is only function of Y_1 and X_1, $\tau_{12}^{\bullet}$ is only function of Y_1 and Y_2 and X_1 and Y_2 are independent).

On the other hand, using expressions (A.134) and (A.124), we have:

$$\begin{aligned} & Cov(\eta_{11}^{\bullet}, \tau_{12}^{\bullet}) = Cov\left(E(\eta_{11}^{\bullet}|Y_1), E(\tau_{12}^{\bullet}|Y_1)\right) \\ &= Cov\left(\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}(K_h * m)(Y_1) - m(Y_1), \frac{(K_b * g)(Y_1) - g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)\right) \\ &= E\left(\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left((K_h * m)(Y_1) - m(Y_1)\right)\frac{(K_b * g)(Y_1) - g(Y_1)}{m(Y_1)}(v(Y_1)-\mu_v)\right) \\ &- E(\eta_{12}^{\bullet})E(\tau_{11}^{\bullet}) \\ &= E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)^2}{m(Y_1)^3}\left[(K_h * m)(Y_1) - m(Y_1)\right]\left[(K_b * g)(Y_1) - g(Y_1)\right]\right] \\ &- E(\eta_{11}^{\bullet})E(\tau_{12}^{\bullet}). \end{aligned} \tag{A.170}$$

The first term in (A.170) is

$$\begin{aligned} & E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)^2}{m(Y_1)^3}\left[(K_h * m)(Y_1) - m(Y_1)\right]\left[(K_b * g)(Y_1) - g(Y_1)\right]\right] \\ &= \int \frac{g(y)(v(y)-\mu_v)^2}{m(y)^3}\left[(K_h * m)(y) - m(y)\right]\left[(K_b * g)(y) - g(y)\right]g(y)dy \\ &= \int \frac{g(y)^2(v(y)-\mu_v)^2}{m(y)^3}\left[\frac{\mu_2(K)}{2}h^2 m''(y) + \frac{\mu_4(K)}{24}h^4 m^{(4)}(y) + O(h^6)\right] \\ &\cdot \left[\frac{\mu_2(K)}{2}b^2 g''(y) + \frac{\mu_4(K)}{24}b^4 g^{(4)}(y) + O(b^6)\right]dy \\ &= \frac{\mu_2(K)^2}{4}h^2 b^2 c^3 \int \zeta(y) m''(y) g''(y) dy \\ &+ \frac{\mu_2(K)\mu_4(K)}{48}c^3\left[h^2 b^4 \int \zeta(y) m''(y) g^{(4)}(y) dy\right. \\ &+ \left. h^4 b^2 \int \zeta(y) m^{(4)}(y) g''(y) dy\right] + O(h^6 b^2) + O(h^4 b^4) + O(h^2 b^6), \end{aligned} \tag{A.171}$$

On the other hand, since

$$\begin{aligned}
E(\eta_{11}^{\bullet}) &= E\left[E(\eta_{11}^{\bullet}|Y_1)\right] = E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h*m)(Y_1)-m(Y_1)\right]\right] \\
&= \int \frac{g(y)^2(v(y)-\mu_v)}{m(y)^2}\left[(K_h*m)(y)-m(y)\right]dy \\
&= c^2\int \Omega(y)(K_h*m)(y)dy - c\int \gamma(y)g(y)dy \\
&= c^2\int \Omega(y)\left(\int K_h(y-z)m(z)dz\right)dy = c^2\int m(z)\left(\int K(t)\Omega(z+ht)dt\right)dz \\
&= c^2\int m(y)\Omega(y)dy + \frac{h^2}{2}\mu_2(K)c^2\int \Omega''(y)m(y)dy \\
&+ \frac{h^4}{24}\mu_4(K)c^2\int \Omega^{(4)}(y)m(y)dy + O(h^6) \\
&= \frac{h^2}{2}\mu_2(K)c^2\int \Omega''(y)m(y)dy + \frac{h^4}{24}\mu_4(K)c^2\int \Omega^{(4)}(y)m(y)dy + O(h^6)
\end{aligned}$$

and

$$E(\tau_{12}^{\bullet}) = \frac{\mu_2(K)}{2}b^2c\int \gamma(y)g''(y)dy + \frac{\mu_4(K)}{24}b^4c\int \gamma(y)g^{(4)}(y)dy + O(b^6),$$

and the second term in (A.170) is

$$\begin{aligned}
E(\eta_{11}^{\bullet})E(\tau_{12}^{\bullet}) &= \frac{\mu_2(K)^2}{4}h^2b^2c^3\int \Omega''(y)m(y)dy\int \gamma(y)g''(y)dy \\
&+ \frac{\mu_2(K)\mu_4(K)}{48}c^3\left[h^2b^4\int \Omega''(y)m(y)dy\int \gamma(y)g^{(4)}(y)dy\right. \qquad (A.172) \\
&+ \left. h^4b^2\int \Omega^{(4)}(y)m(y)dy\int \gamma(y)g''(y)dy\right] + O(h^6b^2) + O(h^4b^4) + O(h^2b^6).
\end{aligned}$$

Plugging (A.171) and (A.172) into (A.170) leads to:

$$\begin{aligned}
Cov(\eta_{11}^{\bullet},\tau_{12}^{\bullet}) &= \frac{\mu_2(K)^2}{4}h^2b^2c^3\left[\int \zeta(y)m''(y)g''(y)dy\right. \\
&- \left.\int \Omega''(y)m(y)dy\int \gamma(y)g''(y)dy\right] \\
&+ \frac{\mu_2(K)\mu_4(K)}{48}c^3\left[h^2b^4\left(\int \zeta(y)m''(y)g^{(4)}(y)dy\right.\right. \\
&- \left.\int \Omega''(y)m(y)dy\int \gamma(y)g^{(4)}(y)dy\right) \\
&+ h^4b^2\left(\int \zeta(y)m^{(4)}(y)g''(y)dy\right. \\
&- \left.\left.\int \Omega^{(4)}(y)m(y)dy\int \gamma(y)g''(y)dy\right)\right] \\
&+ O(h^6b^2) + O(h^4b^4) + O(h^2b^6). \qquad (A.173)
\end{aligned}$$

Let's deal now with the term $Cov(\eta_{11}^{\bullet}, \tau_{21}^{\bullet})$ in (A.169):

$$Cov(\eta_{11}^{\bullet}, \tau_{21}^{\bullet}) = E(\eta_{11}^{\bullet}\tau_{21}^{\bullet}) - E(\eta_{11}^{\bullet})E(\tau_{21}^{\bullet}) = E(\eta_{11}^{\bullet}\tau_{21}^{\bullet}) - E(\eta_{11}^{\bullet})E(\tau_{12}^{\bullet}),$$

since $E(\tau_{12}^{\bullet}) = E(\tau_{21}^{\bullet})$.

Thus, using that $\tau_{21}^{\bullet}$ is a function of only Y_1 and Y_2 we get:

$$\begin{aligned} Cov(\eta_{11}^{\bullet}, \tau_{21}^{\bullet}) &= E(\eta_{11}^{\bullet}\tau_{21}^{\bullet}) - E(\eta_{11}^{\bullet})E(\tau_{12}^{\bullet}) = E\left[E(\eta_{11}^{\bullet}\tau_{21}^{\bullet}|Y_1, Y_2)\right] - E(\eta_{11}^{\bullet})E(\tau_{12}^{\bullet}) \\ &= E\left[E(\eta_{11}^{\bullet}|Y_1)\tau_{21}^{\bullet}\right] - E(\eta_{11}^{\bullet})E(\tau_{12}^{\bullet}), \end{aligned} \tag{A.174}$$

since $\eta_{11}^{\bullet}$ is independent of Y_2.

Using once more the expression (A.124) for $E(\eta_{11}^{\bullet}|Y_1)$, we have:

$$\begin{aligned} & E\left[E(\eta_{11}^{\bullet}|Y_1)\tau_{21}^{\bullet}\right] \\ = \;& E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h * m)(Y_1) - m(Y_1)\right]\frac{v(Y_2)-\mu_v}{m(Y_2)}\left(K_b(Y_2-Y_1) - g(Y_2)\right)\right] \\ = \;& E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h * m)(Y_1) - m(Y_1)\right]\frac{(v(Y_2)-\mu_v)}{m(Y_2)}K_b(Y_2-Y_1)\right] \\ - \;& c \cdot E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h * m)(Y_1) - m(Y_1)\right]\gamma(Y_2)\right]. \end{aligned} \tag{A.175}$$

Let us consider the two terms in (A.175):

$$\begin{aligned} & E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h * m)(Y_1) - m(Y_1)\right]\frac{v(Y_2)-\mu_v}{m(Y_2)}K_b(Y_2-Y_1)\right] \\ = \;& \int\int \frac{g(y)(v(y)-\mu_v)}{m(y)^2}\left[(K_h * m)(y) - m(y)\right]\frac{v(z)-\mu_v}{m(z)}K_b(z-y)g(y)g(z)dydz \\ = \;& c^3\int((K_h * m)(y) - m(y))\Omega(y)\left(\int \gamma(z)K_b(z-y)dz\right)dy \\ = \;& c^3\int((K_h * m)(y) - m(y))\Omega(y)\left(\int \gamma(y+bt)K(t)dt\right)dy \\ = \;& c^3\int((K_h * m)(y) - m(y))\Omega(y)\left(\gamma(y) + \frac{\mu_2(K)}{2}b^2\gamma''(y)\right. \\ + \;& \left.\frac{\mu_4(K)}{24}b^4\gamma^{(4)}(y) + O(b^6)\right)dy \\ = \;& c^3\int\left(\frac{\mu_2(K)}{2}h^2m''(y) + \frac{\mu_4(K)}{24}h^4m^{(4)}(y) + O(h^6)\right)\Omega(y) \\ \cdot \;& \left(\gamma(y) + \frac{\mu_2(K)}{2}b^2\gamma''(y) + \frac{\mu_4(K)}{24}b^4\gamma^{(4)}(y) + O(b^6)\right)dy \end{aligned}$$

$$\begin{aligned} &= \frac{\mu_2(K)}{2} h^2 c^3 \int \psi(y) m''(y) dy + \frac{\mu_4(K)}{24} h^4 c^3 \int \psi(y) m^{(4)}(y) dy \\ &+ \frac{\mu_2(K)^2}{4} h^2 b^2 c^3 \int m''(y)\gamma''(y)\Omega(y) dy \\ &+ \frac{\mu_2(K)\mu_4(K)}{48} h^2 b^4 c^3 \int m''(y)\gamma^{(4)}(y)\Omega(y) dy \\ &+ \frac{\mu_2(K)\mu_4(K)}{48} h^4 b^2 c^3 \int m^{(4)}(y)\gamma''(y)\Omega(y) dy + O(h^6) + O(h^4b^4) + O(h^2b^6). \end{aligned} \tag{A.176}$$

The second term in (A.175) is:

$$\begin{aligned} &E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h * m)(Y_1) - m(Y_1)\right]\gamma(Y_2)\right] \\ &\quad = E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h * m)(Y_1) - m(Y_1)\right]\right] E\left[\gamma(Y_2)\right] = 0, \end{aligned}$$

since

$$E\left[\gamma(Y_2)\right] = \int \gamma(y) g(y) dy = \int f(y)(v(y)-\mu_v) dy = 0.$$

Plugging equation (A.176) and (A.172) into (A.174) gives:

$$\begin{aligned} Cov(\eta^\bullet_{11}, \tau^\bullet_{21}) &= \frac{\mu_2(K)}{2} h^2 c^3 \int \psi(y) m''(y) dy + \frac{\mu_4(K)}{24} h^4 c^3 \int \psi(y) m^{(4)}(y) dy \\ &+ \frac{\mu_2(K)^2}{4} h^2 b^2 c^3 \left[\int m''(y)\gamma''(y)\Omega(y) dy - \int \Omega''(y) m(y) dy \int \gamma(y) g''(y) dy\right] \\ &+ \frac{\mu_2(K)\mu_4(K)}{48} h^2 b^4 c^3 \left[\int m''(y)\gamma^{(4)}(y)\Omega(y) dy - \int \Omega''(y) m(y) dy \int \gamma(y) g^{(4)}(y) dy\right] \\ &+ \frac{\mu_2(K)\mu_4(K)}{48} h^4 b^2 c^3 \left[\int m^{(4)}(y)\gamma''(y)\Omega(y) dy - \int \Omega^{(4)}(y) m(y) dy \int \gamma(y) g''(y) dy\right] \\ &+ O(h^6) + O(h^4b^4) + O(h^2b^6). \end{aligned} \tag{A.177}$$

We finally deal with the last term in (A.169):

$$Cov(\eta^\bullet_{11}, \tau^\bullet_{11}) = E(\eta^\bullet_{11}\tau^\bullet_{11}) - E(\eta^\bullet_{11})E(\tau^\bullet_{11})$$

where the first term is

$$\begin{aligned} &E(\eta^\bullet_{11}\tau^\bullet_{11}) = E\left[E(\eta^\bullet_{11}|Y_1)\tau^\bullet_{11}\right] \\ &= E\left[\frac{g(Y_1)(v(Y_1)-\mu_v)}{m(Y_1)^2}\left[(K_h * m)(Y_1) - m(Y_1)\right]\frac{(v(Y_1)-\mu_v)}{m(Y_1)}(K_b(0) - g(Y_1))\right] \\ &= \int \frac{g(y)^2(v(y)-\mu_v)^2}{m(y)^3}\left[(K_h * m)(y) - m(y)\right](K_b(0) - g(y)) dy \\ &= \frac{K(0)}{b} c^3 \int \zeta(y)\left(\frac{\mu_2(K)}{2} h^2 m''(y) + \frac{\mu_4(K)}{24} h^4 m^{(4)}(y) + O(h^6)\right) dy \\ &- c^3 \int \gamma(y)\Omega(y)\left(\frac{\mu_2(K)}{2} h^2 m''(y) + \frac{\mu_4(K)}{24} h^4 m^{(4)}(y) + O(h^6)\right) dy \end{aligned}$$

$$
\begin{aligned}
&= \frac{\mu_2(K)K(0)}{2}\frac{h^2}{b}c^3\int\zeta(y)m''(y)dy - \frac{\mu_2(K)}{2}h^2c^3\int\psi(y)m''(y)dy \\
&+ \frac{\mu_4(K)K(0)}{24}\frac{h^4}{b}c^3\int\zeta(y)m^{(4)}(y)dy - \frac{\mu_4(K)}{24}h^4c^4\int\psi(y)m^{(4)}(y)dy + O\left(\frac{h^6}{b}\right),
\end{aligned}
$$

and the second one

$$
\begin{aligned}
E(\eta^\bullet_{11})E(\tau^\bullet_{11}) &= \left(\frac{h^2}{2}\mu_2(K)c^2\int\Omega''(y)m(y)dy\right. \\
&+ \left.\frac{h^4}{24}\mu_4(K)c^2\int\Omega^{(4)}(y)m(y)dy + O(h^6)\right)\left(\frac{K(0)}{b}c\int\gamma(y)dy\right) \\
&= \frac{\mu_2(K)K(0)}{2}\frac{h^2}{b}c^3\int\Omega''(y)m(y)dy\int\gamma(y)dy \\
&+ \frac{\mu_4(K)K(0)}{24}\frac{h^4}{b}c^3\int\Omega^{(4)}(y)m(y)dy\int\gamma(y)dy + O\left(\frac{h^6}{b}\right).
\end{aligned}
$$

Consequently:

$$
\begin{aligned}
Cov(\eta^\bullet_{11},\tau^\bullet_{11}) &= \frac{\mu_2(K)K(0)}{2}\frac{h^2}{b}c^3\left[\int\zeta(y)m''(y)dy - \int\Omega''(y)m(y)dy\int\gamma(y)dy\right] \\
&- \frac{\mu_2(K)}{2}h^2c^3\int\psi(y)m''(y)dy \\
&+ \frac{\mu_4(K)K(0)}{24}\frac{h^4}{b}c^3\left[\int\zeta(y)m^{(4)}(y)dy - \int\Omega^{(4)}(y)m(y)dy\int\gamma(y)dy\right] \\
&- \frac{\mu_4(K)}{24}h^4c^3\int\psi(y)m^{(4)}(y)dy + O\left(\frac{h^6}{b}\right). \qquad \text{(A.178)}
\end{aligned}
$$

We now plug (A.173), (A.177) and (A.178) into (A.169) to obtain (5.16). □

Lemma 5.5.10. *The variance of $\widehat{A}^\bullet$ is*

$$
\begin{aligned}
Var\left(\widehat{A}^\bullet\right) &\simeq D^\bullet_7\frac{1}{n} + D^\bullet_8\frac{1}{Nn} + 4D^\bullet_6\frac{1}{N} - 12D^\bullet_6\frac{1}{N^2} + D^\bullet_{26}\frac{1}{N^3} + D^\bullet_9\frac{1}{Nnh} + D^\bullet_{44}\frac{1}{N^2b} \\
&+ D^\bullet_{24}\frac{1}{N^3b^2} + D^\bullet_{25}\frac{1}{N^3b} - 2D^\bullet_{39}\frac{h^2}{N^2b} - 2D^\bullet_{41}\frac{h^4}{N^2b} + D^\bullet_{10}\frac{h^2}{n} + D^\bullet_{11}\frac{h^4}{n} - 4D^\bullet_{27}\frac{h^2}{N} \\
&+ 2D^\bullet_{19}\frac{b^2}{N} + D^\bullet_{46}\frac{h^4}{N} + D^\bullet_{45}\frac{b^4}{N} - 2D^\bullet_{35}\frac{h^2b^2}{N} - 2D^\bullet_{37}\frac{h^4b^2}{N} - 2D^\bullet_{38}\frac{h^2b^4}{N} + D^\bullet_{13}\frac{h}{Nn} \\
&+ D^\bullet_{14}\frac{h^2}{Nn} + D^\bullet_{15}\frac{h^3}{Nn} + D^\bullet_{23}\frac{b}{N^2} + 4D^\bullet_{27}\frac{h^2}{N^2} + 4D^\bullet_{28}\frac{h^4}{N^2} + 2D^\bullet_{35}\frac{h^2b^2}{N^2} + O\left(\frac{h^6}{n}\right) \\
&+ O\left(\frac{b^6}{N}\right) + O\left(\frac{h^4b^4}{N}\right) + O\left(\frac{h^4}{Nn}\right) + O\left(\frac{b^2}{N^2}\right) + O\left(\frac{h^6}{N^2b}\right),
\end{aligned}
$$

where

$$\begin{aligned}
D_{44}^{\bullet} &:= D_{21}^{\bullet}+2D_{30}^{\bullet}=2\mu_0(K^2)c^2\int\alpha(y)dy+4K(0)c^2\int\alpha(y)dy,\\
D_{45}^{\bullet} &:= D_{20}^{\bullet}+2D_{33}^{\bullet}=\frac{\mu_4(K)}{6}c^2\left[\int\alpha(y)g^{(4)}(y)dy+\int\gamma^{(4)}(y)f(y)(v(y)-\mu_v)\right]\\
&+ \frac{\mu_2(K)^2}{4}c^2\left[\int\delta(y)g''(y)^2dy-4\left(\int\gamma(y)g''(y)dy\right)^2\right.\\
&+ \left.\int\gamma''(y)^2g(y)dy+2\int\gamma(y)\gamma''(y)g''(y)dy\right].
\end{aligned}$$

$$\begin{aligned}
D_{46}^{\bullet} &:= D_{12}^{\bullet}-4D_{28}^{\bullet}=\frac{\mu_2(K)^2}{4}c^4\left[\int\xi(y)m''(y)^2dy-B^{\bullet}(\Omega'')^2\right]\\
&- \frac{\mu_4(K)}{6}c^3\int\psi(y)m^{(4)}(y)dy.
\end{aligned}$$

Proof. Consequence of Lemmas 5.5.2, 5.5.4, 5.5.5, 5.5.6, 5.5.7, 5.5.8 and 5.5.9. □

Theorem 5.2.1. *Under the classical conditions on the bandwiths and the sample sizes, i.e. $h\to 0$, $b\to 0$, $nh\to\infty$, $Nb\to\infty$ and $N/n\to\infty$, if Conditions A1, A11 and A12 are fulfilled, then the asymptotic mean squared error of $\hat{\hat{\mu}}_v$ is*

$$\begin{aligned}
AMSE\left(\hat{\hat{\mu}}_v^{2,h,b}\right) &= \left(C_1^{\bullet}b^2+\frac{C_2^{\bullet}}{Nb}+C_3^{\bullet}h^2\right)^2+\frac{C_4^{\bullet}}{n}+\frac{C_5^{\bullet}}{Nn}+\frac{C_{13}^{\bullet}}{N}+\frac{C_6^{\bullet}}{N^2}+\frac{C_7^{\bullet}}{Nnh}\\
&+ \frac{C_8^{\bullet}}{N^2b}+\frac{C_9^{\bullet}h^2}{n}+\frac{C_{14}^{\bullet}h^2}{N}+\frac{C_{10}^{\bullet}h^2}{N^2b}+\frac{C_{11}^{\bullet}h^4}{N}+\frac{C_{15}^{\bullet}b^2}{N}+\frac{C_{12}^{\bullet}h^2b^2}{N},
\end{aligned}$$

where the first three terms come from the squared bias and the rest of them from the variance of the estimator. The constants $C_1^{\bullet},\ldots,C_{15}^{\bullet}$ are defined in the sketch of the proofs (Subsection 5.5.1).

Proof. Consequence of Lemmas 5.5.3 and 5.5.10. □

A.2.2 Proof of Theorem 5.2.2

The proof of Theorem 5.2.2 follows parallel lines to that of Theorem 5.2.1.

Lemma 5.5.11. *The expectation and variance of $\widehat{A}^{*\bullet}$, being*

$$\widehat{A}^{*\bullet}:=\widehat{A}_1^{*\bullet}-\widehat{A}_2^{*\bullet}+\widehat{A}_3^{*\bullet},$$

are

$$E\left(\widehat{A}^{*\bullet}\right)=E\left(\widehat{A}_1^{*\bullet}\right)-E\left(\widehat{A}_2^{*\bullet}\right)+E\left(\widehat{A}_3^{*\bullet}\right),\tag{5.17}$$

$$\begin{aligned}
Var\left(\widehat{A}^{*\bullet}\right) &= Var\left(\widehat{A}_1^{*\bullet}\right)+Var\left(\widehat{A}_2^{*\bullet}\right)+Var\left(\widehat{A}_3^{*\bullet}\right)\\
&- 2Cov\left(\widehat{A}_1^{*\bullet},\widehat{A}_2^{*\bullet}\right)+2Cov\left(\widehat{A}_1^{*\bullet},\widehat{A}_3^{*\bullet}\right)-2Cov\left(\widehat{A}_2^{*\bullet},\widehat{A}_3^{*\bullet}\right),
\end{aligned}\tag{5.18}$$

where

$$\begin{aligned}
\widehat{A}_1^{*\bullet} &:= \frac{1}{N}\sum_{i=1}^{N}\frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)}(v(Y_i)-\mu_v),\\
\widehat{A}_2^{*\bullet} &:= \frac{1}{N}\sum_{i=1}^{N}\frac{(K_b*g)(Y_i)(\hat{m}_h(Y_i)-(K_h*m)(Y_i))}{(K_h*m)(Y_i)^2}(v(Y_i)-\mu_v),\\
\widehat{A}_3^{*\bullet} &:= \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i)-(K_b*g)(Y_i)}{(K_h*m)(Y_i)}(v(Y_i)-\mu_v).
\end{aligned}$$

Proof. We may write some expression for the ratio $\dfrac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}$:

$$\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} = \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\left(\frac{\hat{m}_h(Y_i)}{(K_h*m)(Y_i)}+1-\frac{\hat{m}_h(Y_i)}{(K_h*m)(Y_i)}\right)$$

which gives:

$$\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)} = \frac{\hat{g}_b(Y_i)}{(K_h*m)(Y_i)}+\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\frac{(K_h*m)(Y_i)-\hat{m}_h(Y_i)}{(K_h*m)(Y_i)}.$$

As a consequence,

$$\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}-\frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)} = \frac{\hat{g}_b(Y_i)-(K_b*g)(Y_i)}{(K_h*m)(Y_i)}+\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\frac{(K_h*m)(Y_i)-\hat{m}_h(Y_i)}{(K_h*m)(Y_i)}.$$

Applying similar techniques, once more, to the term $\dfrac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}$ in the right-hand side of the previous expression gives:

$$\begin{aligned}
\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}-\frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)} &= \frac{\hat{g}_b(Y_i)-(K_b*g)(Y_i)}{(K_h*m)(Y_i)}\\
&+ \frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)}\frac{(K_h*m)(Y_i)-\hat{m}_h(Y_i)}{(K_h*m)(Y_i)}\\
&+ \frac{\hat{g}_b(Y_i)-(K_b*g)(Y_i)}{(K_h*m)(Y_i)}\frac{(K_h*m)(Y_i)-\hat{m}_h(Y_i)}{(K_h*m)(Y_i)}\\
&+ \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\left(\frac{(K_h*m)(Y_i)-\hat{m}_h(Y_i)}{(K_h*m)(Y_i)}\right)^2
\end{aligned}$$

which can also be expressed as:

$$\begin{aligned}
\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}-\frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)} &= \frac{\hat{g}_b(Y_i)-(K_b*g)(Y_i)}{(K_h*m)(Y_i)}\\
&- \frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)}\frac{\hat{m}_h(Y_i)-(K_h*m)(Y_i)}{(K_h*m)(Y_i)}\\
&- \frac{(\hat{g}_b(Y_i)-(K_b*g)(Y_i))(\hat{m}_h(Y_i)-(K_h*m)(Y_i))}{(K_h*m)(Y_i)^2}\\
&+ \frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\left(\frac{\hat{m}_h(Y_i)-(K_h*m)(Y_i)}{(K_h*m)(Y_i)}\right)^2. \qquad \text{(A.179)}
\end{aligned}$$

Using (A.179) in the definition of $\widehat{A}^{\bullet}$ (expression (5.6)) gives:

$$\widehat{A}^{\bullet} = \widehat{A}_1^{*\bullet} - \widehat{A}_2^{*\bullet} + \widehat{A}_3^{*\bullet} - \widehat{A}_4^{*\bullet} + \widehat{A}_5^{*\bullet},$$

where

$$\widehat{A}_1^{*\bullet} := \frac{1}{N}\sum_{i=1}^{N}\frac{(K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v), \tag{A.180}$$

$$\widehat{A}_2^{*\bullet} := \frac{1}{N}\sum_{i=1}^{N}\frac{(K_b * g)(Y_i)(\hat{m}_h(Y_i) - (K_h * m)(Y_i))}{(K_h * m)(Y_i)^2}(v(Y_i) - \mu_v), \tag{A.181}$$

$$\widehat{A}_3^{*\bullet} := \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v), \tag{A.182}$$

$$\widehat{A}_4^{*\bullet} := \frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{g}_b(Y_i) - (K_b * g)(Y_i))(\hat{m}_h(Y_i) - (K_h * m)(Y_i))}{(K_h * m)(Y_i)^2}(v(Y_i) - \mu_v), \tag{A.183}$$

$$\widehat{A}_5^{*\bullet} := \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{g}_b(Y_i)}{\hat{m}_h(Y_i)}\left(\frac{\hat{m}_h(Y_i) - (K_h * m)(Y_i)}{(K_h * m)(Y_i)}\right)^2(v(Y_i) - \mu_v). \tag{A.184}$$

Since the terms $\widehat{A}_4^{*\bullet}$ and $\widehat{A}_5^{*\bullet}$ have some factors of quadratic nature within the sum (i.e. $(\hat{g}_b(Y_i) - (K_b * g)(Y_i))(\hat{m}_h(Y_i) - (K_h * m)(Y_i))$ and $(\hat{m}_h(Y_i) - (K_h * m)(Y_i))^2$) it is expected that applying results of the type by Mack & Silverman (1982) one could prove negligibility of the terms (A.183) and (A.184).

Thus we will consider

$$\widehat{A}^{*\bullet} := \widehat{A}_1^{*\bullet} - \widehat{A}_2^{*\bullet} + \widehat{A}_3^{*\bullet}.$$

Since we want to obtain the mean and variance of $\widehat{A}^{*\bullet}$, we proceed as shown in (5.17) and (5.18). □

We now consider the terms in the right-hand sides of (5.17).

Lemma 5.5.12. *The expectation of* $\widehat{A}^{*\bullet}$ *is*

$$E(\widehat{A}^{*\bullet}) \;=\; D_1^{*\bullet} + D_2^{*\bullet}\frac{1}{N}. \tag{5.19}$$

where

$$\begin{aligned} D_1^{*\bullet} &:= \int \gamma^{*\bullet}(y)g(y)dy, \\ D_2^{*\bullet} &:= -D_1^{*\bullet} + \frac{K(0)}{b}\int \frac{(v(y) - \mu_v)}{(K_h * m)(y)}g(y)dy, \end{aligned}$$

with

$$\gamma^{*\bullet}(y) \quad := \quad \frac{(K_b * g)(y)}{(K_h * m)(y)}(v(y) - \mu_v).$$

Proof.

$$\begin{aligned}
E\left(\widehat{A}_1^{*\bullet}\right) &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{(K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{(K_b * g)(Y_1)}{(K_h * m)(Y_1)}(v(Y_1) - \mu_v)\right] \\
&= E\left[\frac{(K_b * g)(Y_1)}{(K_h * m)(Y_1)}(v(Y_1) - \mu_v)\right] = \int \frac{(K_b * g)(y)}{(K_h * m)(y)}(v(y) - \mu_v)g(y)dy \\
&= \int \gamma^{*\bullet} g(y)dy = D_1^{*\bullet},. \qquad \text{(A.185)}
\end{aligned}$$

Since the random variables

$$\eta_i^{*\bullet} := \Psi\left(X_1, \ldots, X_n, Y_i\right) := \frac{(K_b * g)(Y_i)(\hat{m}_h(Y_i) - (K_h * m)(Y_i))}{(K_h * m)(Y_i)^2}(v(Y_i) - \mu_v),$$

for $i = 1, 2, \ldots, N$, are identical distributed (although not independent) then

$$E\left(\widehat{A}_2^{*\bullet}\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(\eta_i^{*\bullet}\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(\eta_1^{*\bullet}\right) = E\left(\eta_1^{*\bullet}\right).$$

But

$$\begin{aligned}
E\left(\eta_1^{*\bullet}\right) &= E\left[E\left(\eta_1^{*\bullet}|Y_1\right)\right] \\
&= E\left[E\left(\frac{(K_b * g)(Y_1)(\hat{m}_h(Y_1) - (K_h * m)(Y_1))}{(K_h * m)(Y_1)^2}(v(Y_1) - \mu_v)\middle| Y_1\right)\right] \\
&= E\left[\frac{(K_b * g)(Y_1)(E\left[\hat{m}_h(Y_1)|Y_1\right] - (K_h * m)(Y_1))}{(K_h * m)(Y_1)^2}(v(Y_1) - \mu_v)\right] \\
&= E\left[\frac{(K_b * g)(Y_1)((K_h * m)(Y_1) - (K_h * m)(Y_1))}{(K_h * m)(Y_1)^2}(v(Y_1) - \mu_v)\right] = 0.
\end{aligned}$$

As a consequence,

$$E\left(\widehat{A}_2^{*\bullet}\right) = E\left(\eta_1^{*\bullet}\right) = 0. \qquad \text{(A.186)}$$

In view of (A.182),

$$\begin{aligned}
E(\widehat{A}_3^{*\bullet}) &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v)\right] \\
&= E\left[\frac{\hat{g}_b(Y_1) - (K_b * g)(Y_1)}{(K_h * m)(Y_1)}(v(Y_1) - \mu_v)\right] \\
&= E\left[E\left(\frac{\hat{g}_b(Y_1) - (K_b * g)(Y_1)}{(K_h * m)(Y_1)}(v(Y_1) - \mu_v)\middle| Y_1\right)\right] \\
&= E\left[\frac{E(\hat{g}_b(Y_1)|Y_1) - (K_b * g)(Y_1)}{(K_h * m)(Y_1)}(v(Y_1) - \mu_v)\right]
\end{aligned}$$

But since

$$\begin{aligned}\hat{g}_b(Y_1) &= \frac{1}{N}\sum_{i=1}^{N} K_b(Y_1 - Y_i) = \frac{1}{N}\left(K_b(0) + \sum_{i=2}^{N} K_b(Y_1 - Y_i)\right) \\ &= \frac{K(0)}{Nb} + \frac{N-1}{N}\frac{1}{N-1}\sum_{i=2}^{N} K_b(Y_1 - Y_i) = \frac{K(0)}{Nb} + \frac{N-1}{N}\hat{g}_b^{(-1)}(Y_1),\end{aligned}$$

we get

$$E\left[\hat{g}_b(Y_1)|Y_1\right] = \frac{K(0)}{Nb} + \frac{N-1}{N}E\left[\hat{g}_b^{(-1)}(Y_1)|Y_1\right] = \frac{K(0)}{Nb} + \frac{N-1}{N}(K_b * g)(Y_1).$$

Using this expression above we have:

$$\begin{aligned}E\left(\widehat{A}_3^{*\bullet}\right) &= E\left[\frac{\frac{K(0)}{Nb} + \frac{N-1}{N}(K_b * g)(Y_1) - (K_b * g)(Y_1)}{(K_h * m)(Y_1)}(v(Y_1) - \mu_v)\right] \\ &= E\left[\frac{\frac{K(0)}{Nb} - \frac{1}{N}(K_b * g)(Y_1)}{(K_h * m)(Y_1)}(v(Y_1) - \mu_v)\right] \\ &= \int \frac{\frac{K(0)}{Nb} - \frac{1}{N}(K_b * g)(y)}{(K_h * m)(y)}(v(y) - \mu_v)g(y)dy \\ &= \frac{K(0)}{Nb}\int \frac{(v(y) - \mu_v)}{(K_h * m)(y)}g(y)dy - \frac{1}{N}D_1^{*\bullet} = \frac{D_2^{*\bullet}}{N}. \qquad \text{(A.187)}\end{aligned}$$

From (A.185), (A.186) and (A.187), we obtain (5.19). □

We now consider the terms in the right-hand side of (5.18):

Lemma 5.5.13. *The variance of $\widehat{A}_1^{*\bullet}$ is*

$$Var\left(\widehat{A}_1^{*\bullet}\right) = \frac{D_3^{*\bullet}}{N}. \qquad (5.20)$$

with

$$D_3^{*\bullet} := \int \alpha^{*\bullet}(y)g(y)dy - D_1^{*\bullet 2},$$

where

$$\alpha^{*\bullet}(y) := \frac{(K_b * g)(y)^2}{(K_h * m)(y)^2}(v(y) - \mu_v)^2.$$

Proof.

$$
\begin{aligned}
Var\left(\widehat{A}_1^{*\bullet}\right) &= \frac{1}{N^2}\sum_{i=1}^{N} Var\left[\frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)}(v(Y_i)-\mu_v)\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var\left[\frac{(K_b*g)(Y_1)}{(K_h*m)(Y_1)}(v(Y_1)-\mu_v)\right] \\
&= \frac{1}{N} Var\left[\frac{(K_b*g)(Y_1)}{(K_h*m)(Y_1)}(v(Y_1)-\mu_v)\right] \\
&= \frac{1}{N}\left\{E\left[\frac{(K_b*g)(Y_1)^2}{(K_h*m)(Y_1)^2}(v(Y_1)-\mu_v)^2\right]\right. \\
&- \left.\left(E\left[\frac{(K_b*g)(Y_1)}{(K_h*m)(Y_1)}(v(Y_1)-\mu_v)\right]\right)^2\right\} \\
&= \frac{1}{N}\left\{\int \frac{(K_b*g)(y)^2}{(K_h*m)(y)^2}(v(y)-\mu_v)^2 g(y)dy\right. \\
&- \left.\left(\int \frac{(K_b*g)(y)}{(K_h*m)(y)}(v(y)-\mu_v)g(y)dy\right)^2\right\}.
\end{aligned}
$$

The definitions of $D_1^{*\bullet}$ and $D_3^{*\bullet}$ lead to (5.20). □

Lemma 5.5.14. *The variance of $\widehat{A}_2^{*\bullet}$ is*

$$
Var\left(\widehat{A}_2^{*\bullet}\right) = D_4^{*\bullet}\frac{1}{n} + D_5^{*\bullet}\frac{1}{Nn}. \tag{5.21}
$$

where

$$
\begin{aligned}
D_4^{*\bullet} &:= \int\left(\int \varphi^{*\bullet}(y)K_h(y-z)g(y)dy - D_1^{*\bullet}\right)^2 m(z)dz, \\
D_5^{*\bullet} &:= \int \varphi^{*\bullet}(y)^2((K_h)^2*m)(y)g(y)dy - \int \alpha^{*\bullet}(y)g(y)dy - D_4^{*\bullet},
\end{aligned}
$$

where

$$
\varphi^{*\bullet}(y) := \frac{(K_b*g)(y)}{(K_h*m)(y)^2}(v(y)-\mu_v).
$$

Proof. In order to compute the variance of $\widehat{A}_2^{*\bullet}$, let us rewrite the terms $\eta_i^{*\bullet}$ as follows:

$$
\begin{aligned}
\eta_i^{*\bullet} &= \frac{(K_b*g)(Y_i)(v(Y_i)-\mu_v)}{(K_h*m)(Y_i)^2}\left(\hat{m}_h(Y_i)-(K_h*m)(Y_i)\right) \\
&= \varphi^{*\bullet}(Y_i)\left(\frac{1}{n}\sum_{j=1}^{n}K_h(Y_i-X_j)-(K_h*m)(Y_i)\right) = \frac{1}{n}\sum_{j=1}^{n}\eta_{ij}^{*\bullet},
\end{aligned}
$$

with

$$\eta_{ij}^{*\bullet} := \varphi^{*\bullet}(Y_i)\left[K_h(Y_i - X_j) - (K_h * m)(Y_i)\right], \tag{A.188}$$

for $i = 1, \ldots, N; j = 1, \ldots, n$.

Now, using (A.188), $\widehat{A}_2^{*\bullet}$ can be written as

$$\widehat{A}_2^{*\bullet} = \frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}\eta_{ij}^{*\bullet}. \tag{A.189}$$

Thus,

$$Var(\widehat{A}_2^{*\bullet}) = \frac{1}{N^2n^2}\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{n} Cov(\eta_{ij}^{*\bullet}, \eta_{kl}^{*\bullet}). \tag{A.190}$$

Collecting all the equal terms in (A.190) gives:

$$\begin{aligned} Var(\widehat{A}_2^{*\bullet}) &= \frac{1}{N^2n^2}\left[Nn(n-1)Cov(\eta_{11}^{*\bullet}, \eta_{12}^{*\bullet})\right. \\ &+ \left. nN(N-1)Cov(\eta_{11}^{*\bullet}, \eta_{21}^{*\bullet}) + NnVar(\eta_{11}^{*\bullet})\right]. \end{aligned} \tag{A.191}$$

We now work these covariance terms out:

$$Cov(\eta_{11}^{*\bullet}, \eta_{12}^{*\bullet}) = Cov(E\left(\eta_{11}^{*\bullet}|Y_1\right), E\left(\eta_{12}^{*\bullet}|Y_1\right)) + E\left[Cov(\eta_{11}^{*\bullet}, \eta_{12}^{*\bullet}|Y_1)\right].$$

But $Cov(\eta_{11}^{*\bullet}, \eta_{12}^{*\bullet}|Y_1) = 0$, since

$$\eta_{11}^{*\bullet} = \varphi^{*\bullet}(Y_1)\left[K_h(Y_1 - X_1) - (K_h * m)(Y_1)\right]$$

and

$$\eta_{12}^{*\bullet} = \varphi^{*\bullet}(Y_1)\left[K_h(Y_1 - X_2) - (K_h * m)(Y_1)\right]$$

are conditionally independent on Y_1 (because X_1 and X_2 are independent).

On the other hand,

$$\begin{aligned} E\left(\eta_{11}^{*\bullet}|Y_1\right) &= E\left(\eta_{12}^{*\bullet}|Y_1\right) = \varphi^{*\bullet}(Y_1)\left[E\left[K_h(Y_1 - X_1)|Y_1\right] - (K_h * m)(Y_1)\right] \\ &= \varphi^{*\bullet}(Y_1)\left[E\left[K_h(X_1 - Y_1)|Y_1\right] - (K_h * m)(Y_1)\right] \\ &= \varphi^{*\bullet}(Y_1)\left[(K_h * m)(Y_1) - (K_h * m)(Y_1)\right] = 0. \end{aligned}$$

So,

$$Cov(\eta_{11}^{*\bullet}, \eta_{12}^{*\bullet}) = Cov(E\left(\eta_{11}^{*\bullet}|Y_1\right), E\left(\eta_{12}^{*\bullet}|Y_1\right)) = Var(E\left(\eta_{11}^{*\bullet}|Y_1\right)) = 0. \tag{A.192}$$

Now we deal with the term $Cov(\eta_{11}^{*\bullet},\eta_{21}^{*\bullet})$ in (A.191):

$$
\begin{aligned}
&Cov(\eta_{11}^{*\bullet},\eta_{21}^{*\bullet}) = Cov(E\left(\eta_{11}^{*\bullet}|X_1\right), E\left(\eta_{21}^{*\bullet}|X_1\right)) + E\left[Cov(\eta_{11}^{*\bullet},\eta_{21}^{*\bullet}|X_1)\right]\\
= &\ Cov\left(E\left(\eta_{11}^{*\bullet}|X_1\right), E\left(\eta_{21}^{*\bullet}|X_1\right)\right) = E\left(E\left(\eta_{11}^{*\bullet}|X_1\right)E\left(\eta_{21}^{*\bullet}|X_1\right)\right)\\
- &\ E\left(E\left(\eta_{11}^{*\bullet}|X_1\right)\right)E\left(E\left(\eta_{21}^{*\bullet}|X_1\right)\right) = E\left(E\left(\eta_{11}^{*\bullet}|X_1\right)E\left(\eta_{21}^{*\bullet}|X_1\right)\right)\\
- &\ E\left(\eta_{11}^{*\bullet}\right)E\left(\eta_{21}^{*\bullet}\right) = E\left(E\left(\eta_{11}^{*\bullet}|X_1\right)E\left(\eta_{21}^{*\bullet}|X_1\right)\right),
\end{aligned}
$$

since $Cov(\eta_{11}^{*\bullet},\eta_{21}^{*\bullet}|X_1) = 0$ because

$$\eta_{11}^{*\bullet} = \varphi^{*\bullet}(Y_1)\left[K_h(Y_1 - X_1) - (K_h * m)(Y_1)\right]$$

and

$$\eta_{21}^{*\bullet} = \varphi^{*\bullet}(Y_2)\left[K_h(Y_2 - X_1) - (K_h * m)(Y_2)\right]$$

are conditionally independent given X_1 (since Y_1 and Y_2 are independent) and

$$E\left(\eta_{21}^{*\bullet}\right) = E\left(E\left(\eta_{21}^{*\bullet}|Y_2\right)\right) = 0.$$

But

$$E\left(\eta_{11}^{*\bullet}|X_1\right) = E\left(\eta_{21}^{*\bullet}|X_1\right) = \int \varphi^{*\bullet}(y)\left[K_h(y - X_1) - (K_h * m)(y)\right]g(y)dy$$

then

$$
\begin{aligned}
&Cov(\eta_{11}^{*\bullet},\eta_{21}^{*\bullet}) = E\left(\left(\int \varphi^{*\bullet}(y)\left[K_h(y - X_1) - (K_h * m)(y)\right]g(y)dy\right)^2\right)\\
= &\int\left(\int \varphi^{*\bullet}(y)\left[K_h(y - z) - (K_h * m)(y)\right]g(y)dy\right)^2 m(z)dz\\
= &\int\left(\int \varphi^{*\bullet}(y)K_h(y - z)g(y)dy - \int \gamma^{*\bullet}(y)g(y)dy\right)^2 m(z)dz. \qquad \text{(A.193)}
\end{aligned}
$$

We now examine the term $Var(\eta_{11}^{*\bullet})$ in (A.191):

$$Var(\eta_{11}^{*\bullet}) = E\left(\eta_{11}^{*\bullet 2}\right) + E(\eta_{11}^{*\bullet})^2 = E\left(\eta_{11}^{*\bullet 2}\right) = E\left[E(\eta_{11}^{*\bullet 2}|Y_1)\right],$$

since $E(\eta_{11}^{*\bullet}) = E\left[E\left(\eta_{11}^{*\bullet}|Y_1\right)\right] = 0$.

So,

$$
\begin{aligned}
Var(\eta_{11}^{*\bullet}) &= E\left[E\left(\eta_{11}^{*\bullet 2}|Y_1\right)\right]\\
&= E\left[E\left(\left(\varphi^{*\bullet}(Y_1)\left[K_h(Y_1 - X_1) - (K_h * m)(Y_1)\right]\right)^2|Y_1\right)\right]\\
&= \int \varphi^{*\bullet}(y)^2\left[((K_h)^2 * m)(y) - (K_h * m)(y)^2\right]g(y)dy\\
&= \int \varphi^{*\bullet}(y)^2((K_h)^2 * m)(y)g(y)dy - \int \alpha^{*\bullet}(y)g(y)dy. \qquad \text{(A.194)}
\end{aligned}
$$

Now, using (A.192), (A.193) and (A.194) in (A.191) gives:

$$
\begin{aligned}
Var(\widehat{A}_2^{*\bullet}) =& \frac{N-1}{Nn}\left[\int\left(\int\varphi^{*\bullet}(y)K_h(y-z)g(y)dy-\int\gamma^{*\bullet}(y)g(y)dy\right)^2 m(z)dz\right]\\
+ & \frac{1}{Nn}\left[\int\varphi^{*\bullet}(y)^2((K_h)^2*m)(y)g(y)dy-\int\alpha^{*\bullet}(y)g(y)dy\right]\\
= & \frac{1}{n}\left[\int\left(\int\varphi^{*\bullet}(y)K_h(y-z)g(y)dy-\int\gamma^{*\bullet}(y)g(y)dy\right)^2 m(z)dz\right]\\
+ & \frac{1}{Nn}\left[\int\varphi^{*\bullet}(y)^2((K_h)^2*m)(y)g(y)dy-\int\alpha^{*\bullet}(y)g(y)dy\right.\\
& \left.-\int\left(\int\varphi^{*\bullet}(y)K_h(y-z)g(y)dy-\int\gamma^{*\bullet}(y)g(y)dy\right)^2 m(z)dz\right].
\end{aligned}
$$

Using the definitions of $D_4^{*\bullet}$ and $D_5^{*\bullet}$, we obtain (5.21). □

Lemma 5.5.15. *The variance of $\widehat{A}_3^{*\bullet}$ is*

$$
Var\left(\widehat{A}_3^{*\bullet}\right) = D_6^{*\bullet}\frac{1}{N}+D_7^{*\bullet}\frac{1}{N^2}+D_8^{*\bullet}\frac{1}{N^3}, \tag{5.22}
$$

where

$$
\begin{aligned}
D_6^{*\bullet} &:= \int\left[\int\frac{v(y)-\mu_v}{(K_h*m)(y)}K_b(y-z)g(y)dy-\int\gamma^{*\bullet}(y)g(y)dy\right]^2 g(z)dz\\
&= \int\left[\int\frac{v(y)-\mu_v}{(K_h*m)(y)}K_b(y-z)g(y)dy-D_1^{*\bullet}\right]^2 g(z)dz,
\end{aligned}
$$

$$
\begin{aligned}
D_7^{*\bullet} &:= \frac{2K(0)}{b}\left[\int\int\frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz\right.\\
&- \left.\int\gamma^{*\bullet}(y)g(y)dy\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right]\\
&- 4\int\int\gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz+3\left(\int\gamma^{*\bullet}(y)g(y)dy\right)^2\\
&- \int\alpha^{*\bullet}(y)g(y)dy+\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2((K_b)^2*g)(y)g(y)dy\\
&+ \int\int\frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)^2g(y)g(z)dydz\\
&- 3\int\left[\int\frac{v(y)-\mu_v}{(K_h*m)(y)}K_b(y-z)g(y)dy-\int\gamma^{*\bullet}(y)g(y)dy\right]^2 g(z)dz
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2K(0)}{b}\Bigg[\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz \\
&- \int \gamma^{*\bullet}(y)g(y)dy\int \frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\Bigg] \\
&- 4\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz + 3D_1^{*\bullet 2} \\
&- \int \alpha^{*\bullet}(y)g(y)dy + \int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2((K_b)^2*g)(y)g(y)dy \\
&+ \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)^2g(y)g(z)dydz - 3D_6^{*\bullet},
\end{aligned}
$$

$$
\begin{aligned}
D_8^{*\bullet} &:= \frac{K(0)^2}{b^2}\left[\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2g(y)dy - \left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)^2\right] \\
&+ \frac{2K(0)}{b}\Bigg[-\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz \\
&+ 2\int \gamma^{*\bullet}(y)g(y)dy\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy - \int\rho^{*\bullet}(y)g(y)dy\Bigg] \\
&+ 4\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz - 4\left(\int\gamma^{*\bullet}(y)g(y)dy\right)^2 \\
&+ 2\int\alpha^{*\bullet}(y)g(y)dy - \int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2((K_b)^2*g)(y)g(y)dy
\end{aligned}
$$

$$
\begin{aligned}
&- \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)^2g(y)g(z)dydz \\
&+ 2\int\left[\int\frac{v(y)-\mu_v}{(K_h*m)(y)}K_b(y-z)g(y)dy - \int\gamma^{*\bullet}(y)g(y)dy\right]^2g(z)dz \\
&= \frac{K(0)^2}{b^2}\left[\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2g(y)dy - \left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)^2\right] \\
&+ \frac{2K(0)}{b}\Bigg[-\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz \\
&+ 2\int \gamma^{*\bullet}(y)g(y)dy\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy - \int\rho^{*\bullet}(y)g(y)dy\Bigg] \\
&+ 4\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz - 4D_1^{*\bullet 2} \\
&+ 2\int\alpha^{*\bullet}(y)g(y)dy - \int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2((K_b)^2*g)(y)g(y)dy \\
&- \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)^2g(y)g(z)dydz + 2D_6^{*\bullet},
\end{aligned}
$$

with

$$\rho^{*\bullet}(y) := \frac{(K_b * g)(y)}{(K_h * m)(y)^2}(v(y) - \mu_v)^2.$$

Proof. To deal with the variance of $\widehat{A}_3^{*\bullet}$ we first consider

$$\tau_i^{*\bullet} := \frac{\hat{g}_b(Y_i) - (K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v), \quad i = 1, \ldots, N$$

and write

$$\hat{g}_b(Y_i) = \frac{1}{N}\sum_{j=1}^{N} K_b(Y_i - Y_j).$$

As a consequence,

$$\tau_i^{*\bullet} = \frac{\displaystyle\frac{1}{N}\sum_{j=1}^{N} K_b(Y_i - Y_j) - (K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v) = \frac{1}{N}\sum_{j=1}^{N}\tau_{ij}^{*\bullet},$$

where

$$\tau_{ij}^{*\bullet} := \frac{K_b(Y_i - Y_j) - (K_b * g)(Y_i)}{(K_h * m)(Y_i)}(v(Y_i) - \mu_v),$$

for $i, j = 1, \ldots, N$. Then

$$\widehat{A}_3^{*\bullet} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij}^{*\bullet}. \tag{A.195}$$

To compute the variance of $\widehat{A}_3^{*\bullet}$ we consider

$$\begin{aligned} Var(\widehat{A}_3^{*\bullet}) &= Cov(\widehat{A}_3^{*\bullet}, \widehat{A}_3^{*\bullet}) = Cov\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij}^{*\bullet}, \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\tau_{kl}^{*\bullet}\right) \\ &= \frac{1}{N^4}\sum_{i,j,k,l=1}^{N} Cov(\tau_{ij}^{*\bullet}, \tau_{kl}^{*\bullet}). \end{aligned}$$

Thus, the variance of $\widehat{A}_3^{\bullet}$ can be written as:

$$\begin{aligned} Var(\widehat{A}_3^{*\bullet}) = {} & \frac{(N-1)(N-2)}{N^3}Cov(\tau_{12}^{*\bullet}, \tau_{13}^{*\bullet}) + \frac{2(N-1)(N-2)}{N^3}Cov(\tau_{12}^{*\bullet}, \tau_{31}^{*\bullet}) \\ + {} & \frac{(N-1)(N-2)}{N^3}Cov(\tau_{12}^{*\bullet}, \tau_{32}^{*\bullet}) + \frac{N-1}{N^3}Var(\tau_{12}^{*\bullet}) + \frac{N-1}{N^3}Cov(\tau_{12}^{*\bullet}, \tau_{21}^{*\bullet}) \\ + {} & \frac{2(N-1)}{N^3}Cov(\tau_{12}^{*\bullet}, \tau_{11}^{*\bullet}) + \frac{2(N-1)}{N^3}Cov(\tau_{12}^{*\bullet}, \tau_{22}^{*\bullet}) + \frac{1}{N^3}Var(\tau_{11}^{*\bullet}). \end{aligned} \tag{A.196}$$

Let's deal with every one of these terms; but first, in order to save space, let us write $\tau_{ij}^{*\bullet}$ in a more compact form:

$$\tau_{ij}^{*\bullet} = \frac{v(Y_i) - \mu_v}{(K_h * m)(Y_i)}(K_b(Y_i - Y_j) - (K_b * g)(Y_i)), \quad i, j = 1, \ldots, N.$$

Let us now study the terms in (A.196):

$$Cov(\tau_{12}^{*\bullet},\tau_{13}^{*\bullet}) = E\left[Cov(\tau_{12}^{*\bullet},\tau_{13}^{*\bullet}|Y_1)\right] + Cov\left(E\left(\tau_{12}^{*\bullet}|Y_1\right), E\left(\tau_{13}^{*\bullet}|Y_1\right)\right) = Var\left(E\left(\tau_{12}^{*\bullet}|Y_1\right)\right)$$

because

$$\tau_{12}^{*\bullet} = \frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}(K_b(Y_1-Y_2)-(K_b*g)(Y_1))$$

and

$$\tau_{13}^{*\bullet} = \frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}(K_b(Y_1-Y_3)-(K_b*g)(Y_1))$$

are conditionally independent given Y_1 (so $Cov(\tau_{12}^{*\bullet},\tau_{13}^{*\bullet}|Y_1)=0$). But

$$\begin{aligned} E\left(\tau_{12}^{*\bullet}|Y_1\right) &= E\left(\tau_{13}^{*\bullet}|Y_1\right) = \frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}\left[E\left[K_b(Y_1-Y_2)|Y_1\right]-(K_b*g)(Y_1)\right] \\ &= \frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}\left[(K_b*g)(Y_1)-(K_b*g)(Y_1)\right]=0, \end{aligned}$$

then

$$Cov(\tau_{12}^{*\bullet},\tau_{13}^{*\bullet}) = 0. \tag{A.197}$$

Now consider

$$\begin{aligned} &Cov(\tau_{12}^{*\bullet},\tau_{31}^{*\bullet}) = E\left[Cov(\tau_{12}^{*\bullet},\tau_{31}^{*\bullet}|Y_1)\right] + Cov\left(E\left(\tau_{12}^{*\bullet}|Y_1\right), E\left(\tau_{31}^{*\bullet}|Y_1\right)\right) \\ &= Cov\left(E\left(\tau_{12}^{*\bullet}|Y_1\right), E\left(\tau_{31}^{*\bullet}|Y_1\right)\right) \\ &= E\left[E\left(\tau_{12}^{*\bullet}|Y_1\right)E\left(\tau_{31}^{*\bullet}|Y_1\right)\right] - E\left[E\left(\tau_{12}^{*\bullet}|Y_1\right)\right]E\left[E\left(\tau_{31}^{*\bullet}|Y_1\right)\right] = 0. \end{aligned} \tag{A.198}$$

since $E\left(\tau_{12}^{*\bullet}|Y_1\right)=0$.

We now deal with the covariance:

$$\begin{aligned} &Cov(\tau_{12}^{*\bullet},\tau_{32}^{*\bullet}) = E\left[Cov(\tau_{12}^{*\bullet},\tau_{32}^{*\bullet}|Y_2)\right] + Cov\left(E\left(\tau_{12}^{*\bullet}|Y_2\right), E\left(\tau_{32}^{*\bullet}|Y_2\right)\right) \\ &= Cov\left(E\left(\tau_{12}^{*\bullet}|Y_2\right), E\left(\tau_{32}^{*\bullet}|Y_2\right)\right) = E\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)E\left(\tau_{32}^{*\bullet}|Y_2\right)\right] \\ &- E\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)\right]E\left[E\left(\tau_{32}^{*\bullet}|Y_2\right)\right] = E\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)E\left(\tau_{32}^{*\bullet}|Y_2\right)\right] \\ &- E(\tau_{12}^{*\bullet})E(\tau_{32}^{*\bullet}) = E\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)E\left(\tau_{32}^{*\bullet}|Y_2\right)\right] = E\left[\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)\right]^2\right] \end{aligned}$$

since $E(\tau_{12}^{*\bullet}) = E\left[E\left(\tau_{12}^{*\bullet}|Y_1\right)\right]=0$.

But

$$\begin{aligned} E\left(\tau_{12}^{*\bullet}|Y_2\right) &= E\left[\frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}(K_b(Y_1-Y_2)-(K_b*g)(Y_1))\right] \\ &= \int \frac{v(y)-\mu_v}{(K_h * m)(y)}\left[K_b(y-Y_2)-(K_b*g)(y)\right]g(y)dy \end{aligned}$$

and consequently

$$
\begin{aligned}
Cov(\tau_{12}^{*\bullet}, \tau_{32}^{*\bullet}) &= E\left[\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)\right]^2\right] \\
&= E\left[\left(\int \frac{v(y)-\mu_v}{(K_h * m)(y)} \left[K_b(y-Y_2) - (K_b * g)(y)\right] g(y)dy\right)^2\right] \\
&= \int \left[\int \frac{v(y)-\mu_v}{(K_h * m)(y)} \left(K_b(y-z) - (K_b * g)(y)\right) g(y)dy\right]^2 g(z)dz \\
&= \int \left[\int \frac{v(y)-\mu_v}{(K_h * m)(y)} K_b(y-z)g(y)dy - \int \gamma^{*\bullet}(y)g(y)dy\right]^2 g(z)dz. \quad \text{(A.199)}
\end{aligned}
$$

Let's deal now with $Var(\tau_{12}^{*\bullet})$:

$$
\begin{aligned}
Var(\tau_{12}^{*\bullet}) &= E(\tau_{12}^{*\bullet 2}) - [E(\tau_{12}^{*\bullet})]^2 = E(\tau_{12}^{*\bullet 2}) = E\left[E(\tau_{12}^{*\bullet 2}|Y_1)\right] \\
&= E\left[E\left(\left[\frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}(K_b(Y_1-Y_2) - (K_b * g)(Y_1))\right]^2 \middle| Y_1\right)\right] \\
&= E\left[\left(\frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}\right)^2 \left(E\left[K_b(Y_1-Y_2)^2|Y_1\right]\right.\right. \\
&- \left.\left. 2(K_b * g)(Y_1)E\left[K_b(Y_1-Y_2)|Y_1\right] + (K_b * g)(Y_1)^2\right)\right] \\
&= E\left[\left(\frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}\right)^2 \left(((K_b)^2 * g)(Y_1) - 2(K_b * g)(Y_1)^2 + (K_b * g)(Y_1)^2\right)\right] \\
&= E\left[\left(\frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}\right)^2 \left(((K_b)^2 * g)(Y_1) - (K_b * g)(Y_1)^2\right)\right] \\
&= \int \left(\frac{v(y)-\mu_v}{(K_h * m)(y)}\right)^2 \left(((K_b)^2 * g)(y) - (K_b * g)(y)^2\right) g(y)dy \\
&= \int \left(\frac{v(y)-\mu_v}{(K_h * m)(y)}\right)^2 ((K_b)^2 * g)(y)g(y)dy - \int \alpha^{*\bullet}(y)g(y)dy. \quad \text{(A.200)}
\end{aligned}
$$

Let us now consider $Cov(\tau_{12}^{*\bullet}, \tau_{21}^{*\bullet})$:

$$
\begin{aligned}
Cov(\tau_{12}^{*\bullet}, \tau_{21}^{*\bullet}) &= E(\tau_{12}^{*\bullet}\tau_{21}^{*\bullet}) - E(\tau_{12}^{*\bullet})E(\tau_{21}^{*\bullet}) = E(\tau_{12}^{*\bullet}\tau_{21}^{*\bullet}) - [E(\tau_{12}^{*\bullet})]^2 \\
&= E(\tau_{12}^{*\bullet}\tau_{21}^{*\bullet}) = E\left[\frac{v(Y_1)-\mu_v}{(K_h * m)(Y_1)}(K_b(Y_1-Y_2)\right. \\
&- \left.(K_b * g)(Y_1))\frac{v(Y_2)-\mu_v}{(K_h * m)(Y_2)}(K_b(Y_2-Y_1) - (K_b * g)(Y_2))\right] \\
&= \int\int \frac{v(y)-\mu_v}{(K_h * m)(y)}(K_b(y-z) \\
&- (K_b * g)(y))\frac{v(z)-\mu_v}{(K_h * m)(z)}\left(K_b(z-y) - (K_b * g)(z)\right) g(y)g(z)dydz
\end{aligned}
$$

$$\begin{aligned}
&= \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)^2 g(y)g(z)dydz \\
&- \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)(K_b*g)(z)g(y)g(z)dydz \\
&- \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)(K_b*g)(y)g(y)g(z)dydz \\
&+ \int\int \frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b*g)(y)\frac{v(z)-\mu_v}{(K_h*m)(z)}(K_b*g)(z)g(y)g(z)dydz \\
&= \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)^2 g(y)g(z)dydz \\
&+ \left(\int \frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b*g)(y)g(y)dy\right)^2 \\
&- 2\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)(K_b*g)(y)g(y)g(z)dydz \\
&= \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)} K_b(y-z)^2 g(y)g(z)dydz \\
&- 2\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)} K_b(y-z)g(y)g(z)dydz \\
&+ \left(\int \gamma^{*\bullet}(y)g(y)dy\right)^2. \qquad \text{(A.201)}
\end{aligned}$$

Let's deal now with the term:

$$\begin{aligned}
Cov(\tau_{12}^{*\bullet},\tau_{11}^{*\bullet}) &= Cov\left(E\left(\tau_{12}^{*\bullet}|Y_1\right),E\left(\tau_{11}^{*\bullet}|Y_1\right)\right)+E\left[Cov(\tau_{12}^{*\bullet},\tau_{11}^{*\bullet}|Y_1)\right] \\
&= Cov\left(E\left(\tau_{12}^{*\bullet}|Y_1\right),E\left(\tau_{11}^{*\bullet}|Y_1\right)\right)=0, \qquad \text{(A.202)}
\end{aligned}$$

since $E\left(\tau_{12}^{*\bullet}|Y_1\right)=0$.

We now compute the covariance $Cov(\tau_{12}^{*\bullet},\tau_{22}^{*\bullet})$:

$$\begin{aligned}
&Cov(\tau_{12}^{*\bullet},\tau_{22}^{*\bullet}) = Cov\left(E\left(\tau_{12}^{*\bullet}|Y_2\right),E\left(\tau_{22}^{*\bullet}|Y_2\right)\right)+E\left[Cov(\tau_{12}^{*\bullet},\tau_{22}^{*\bullet}|Y_2)\right] \\
&= Cov\left(E\left(\tau_{12}^{*\bullet}|Y_2\right),E\left(\tau_{22}^{*\bullet}|Y_2\right)\right)=E\left(E\left(\tau_{12}^{*\bullet}|Y_2\right)\tau_{22}^{*\bullet}\right)-E\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)\right]E\left(\tau_{22}^{*\bullet}\right) \\
&= E\left(E\left(\frac{v(Y_1)-\mu_v}{(K_h*m)(Y_1)}(K_b(Y_1-Y_2)-(K_b*g)(Y_1))\middle|Y_2\right)\tau_{22}^{*\bullet}\right),
\end{aligned}$$

since $Cov(\tau_{12}^{*\bullet},\tau_{22}^{*\bullet}|Y_2)=0$ (because $\tau_{22}^{*\bullet}$ is a measurable function of Y_2) and since

$$E\left[E\left(\tau_{12}^{*\bullet}|Y_2\right)\right]=E(\tau_{12}^{*\bullet})=E\left[E\left(\tau_{12}^{*\bullet}|Y_1\right)\right]=0.$$

On the other hand,

$$
\begin{aligned}
& E\left[\frac{v(Y_1)-\mu_v}{(K_h*m)(Y_1)}(K_b(Y_1-Y_2)-(K_b*g)(Y_1))\middle| Y_2\right] \\
= & \int \frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b(y-Y_2)-(K_b*g)(y))g(y)dy
\end{aligned}
$$

Using the previous expressions and $\tau_{22}^{*\bullet} = \frac{v(Y_2)-\mu_v}{(K_h*m)(Y_2)}\left(\frac{K(0)}{b}-(K_b*g)(Y_2)\right)$, we have:

$$
\begin{aligned}
Cov(\tau_{12}^{*\bullet},\tau_{22}^{*\bullet}) = & E\left[\left(\int \frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b(y-Y_2)-(K_b*g)(y))g(y)dy\right)\right. \\
\cdot & \left.\frac{v(Y_2)-\mu_v}{(K_h*m)(Y_2)}\left(\frac{K(0)}{b}-(K_b*g)(Y_2)\right)\right] \\
= & \int\int \frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b(y-z)-(K_b*g)(y))\frac{v(z)-\mu_v}{(K_h*m)(z)} \\
\cdot & \left(\frac{K(0)}{b}-(K_b*g)(z)\right)g(y)g(z)dydz \\
= & \frac{K(0)}{b}\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz \\
- & \int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)(K_b*g)(z)g(y)g(z)dydz \\
- & \frac{K(0)}{b}\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}(K_b*g)(y)g(y)g(z)dydz \\
+ & \left(\int \frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b*g)(y)g(y)dy\right)^2 \\
= & \frac{K(0)}{b}\int\int \frac{(v(y)-\mu_v)(v(z)-\mu_v)}{(K_h*m)(y)(K_h*m)(z)}K_b(y-z)g(y)g(z)dydz \\
+ & \left(\int \gamma^{*\bullet}(y)g(y)dy\right)^2 - \int\int \frac{v(y)-\mu_v}{(K_h*m)(y)}\gamma^{*\bullet}(z)K_b(y-z)g(y)g(z)dydz \\
- & \frac{K(0)}{b}\int\int \gamma^{*\bullet}(y)\frac{v(z)-\mu_v}{(K_h*m)(z)}g(y)g(z)dydz. \qquad \text{(A.203)}
\end{aligned}
$$

Finally, we study the term $Var(\tau_{11}^{*\bullet})$:

$$Var(\tau_{11}^{*\bullet}) = E(\tau_{11}^{*\bullet 2}) - E(\tau_{11}^{*\bullet})^2,$$

where the first term $E(\tau_{11}^{*\bullet 2})$ is:

$$E(\tau_{11}^{*\bullet 2}) = E\left[\left(\frac{v(Y_1)-\mu_v}{(K_h*m)(Y_1)}\right)^2\left(\frac{K(0)}{b}-(K_b*g)(Y_1)\right)^2\right]$$
$$= \frac{K(0)^2}{b^2}E\left[\left(\frac{v(Y_1)-\mu_v}{(K_h*m)(Y_1)}\right)^2\right]-2\frac{K(0)}{b}E\left[\left(\frac{v(Y_1)-\mu_v}{(K_h*m)(Y_1)}\right)^2(K_b*g)(Y_1)\right]$$
$$+ \ E\left[\left(\frac{v(Y_1)-\mu_v}{(K_h*m)(Y_1)}\right)^2(K_b*g)(Y_1)^2\right]$$
$$= \frac{K(0)^2}{b^2}\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 g(y)dy$$
$$- \ \frac{2K(0)}{b}\int \rho^{*\bullet}(y)g(y)dy + \int \alpha^{*\bullet}(y)g(y)dy. \qquad (A.204)$$

On the other hand,

$$\begin{aligned}
E(\tau_{11}^{*\bullet})^2 &= \left[E\left(\frac{v(Y_1)-\mu_v}{(K_h*m)(Y_1)}\left(\frac{K(0)}{b}-(K_b*g)(Y_1)\right)\right)\right]^2\\
&= \left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}\left(\frac{K(0)}{b}-(K_b*g)(y)\right)g(y)dy\right)^2\\
&= \frac{K(0)^2}{b^2}\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)^2\\
&- 2\frac{K(0)}{b}\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)\left(\int\gamma^{*\bullet}(y)g(y)dy\right)\\
&+ \left(\int\gamma^{*\bullet}(y)g(y)dy\right)^2. \qquad (A.205)
\end{aligned}$$

Using (A.204) and (A.205) we obtain

$$\begin{aligned}
Var(\tau_{11}^{*\bullet}) &= \frac{K(0)^2}{b^2}\left[\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 g(y)dy-\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)^2\right]\\
&- \frac{2K(0)}{b}\left[\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2(K_b*g)(y)g(y)dy\right.\\
&- \left.\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b*g)(y)g(y)dy\right)\right]\\
&+ \left[\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2(K_b*g)(y)^2g(y)dy\right.\\
&- \left.\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}(K_b*g)(y)g(y)dy\right)^2\right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{K(0)^2}{b^2}\left[\int\left(\frac{v(y)-\mu_v}{(K_h*m)(y)}\right)^2 g(y)dy-\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)^2\right]\\
&- \frac{2K(0)}{b}\left[\int\rho^{*\bullet}(y)g(y)dy-\left(\int\frac{v(y)-\mu_v}{(K_h*m)(y)}g(y)dy\right)\left(\int\gamma^{*\bullet}(y)g(y)dy\right)\right]\\
&+ \int\alpha^{*\bullet}(y)g(y)dy-\left(\int\gamma^{*\bullet}(y)g(y)dy\right)^2. \qquad (A.206)
\end{aligned}$$

Expressions (A.197), (A.198), (A.199), (A.200), (A.201), (A.202), (A.203) and (A.206) can be used in (A.196) to obtain (5.22). □

We will proceed now with the covariance terms in (5.18).

Let us define

$$\omega_i^{*\bullet} := \frac{(K_b*g)(Y_i)}{(K_h*m)(Y_i)}(v(Y_i)-\mu_v).$$

Using this definition and the expressions for $\widehat{A}_1^{*\bullet}$ in (A.180), $\widehat{A}_2^{*\bullet}$ in (A.181) and in (A.189) and for $\widehat{A}_3^{*\bullet}$ in (A.182) and in (A.195), we have

$$\begin{aligned}
\widehat{A}_1^{*\bullet} &= \frac{1}{N}\sum_{i=1}^{N}\omega_i^{*\bullet},\\
\widehat{A}_2^{*\bullet} &= \frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}\eta_{ij}^{*\bullet}, \quad \text{with} \quad \eta_{ij}^{*\bullet}=\varphi^{*\bullet}(Y_i)\left[K_h(Y_i-X_j)-(K_h*m)(Y_i)\right],
\end{aligned}$$

and

$$\widehat{A}_3^{*\bullet} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij}^{*\bullet}, \quad \text{with} \quad \tau_{ij}^{*\bullet}=\frac{v(Y_i)-\mu_v}{(K_h*m)(Y_i)}(K_b(Y_i-Y_j)-(K_b*g)(Y_i))$$

Lemma 5.5.16. *The covariance of* $\widehat{A}_1^{*\bullet}$ *and* $\widehat{A}_2^{*\bullet}$ *is*

$$Cov\left(\widehat{A}_1^{*\bullet},\widehat{A}_2^{*\bullet}\right) = 0.$$

Proof. Let us now consider $Cov(\widehat{A}_1^{*\bullet},\widehat{A}_2^{*\bullet})$:

$$\begin{aligned}
Cov(\widehat{A}_1^{*\bullet},\widehat{A}_2^{*\bullet}) &= Cov\left(\frac{1}{N}\sum_{i=1}^{N}\omega_i^{*\bullet},\frac{1}{Nn}\sum_{j=1}^{N}\sum_{k=1}^{n}\eta_{jk}^{*\bullet}\right)\\
&= \frac{1}{N^2n}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{n}Cov(\omega_i^{*\bullet},\eta_{jk}^{*\bullet})\\
&= \frac{1}{N^2n}NnCov(\omega_1^{*\bullet},\eta_{11}^{*\bullet})=\frac{1}{N}Cov(\omega_1^{*\bullet},\eta_{11}^{*\bullet}),
\end{aligned}$$

since $Cov(\omega_i^{*\bullet}, \eta_{jk}^{*\bullet}) = 0$ for $i \neq j$ because $\omega_i^{*\bullet}$ and $\eta_{jk}^{*\bullet}$ are independent for $i \neq j$.

But

$$\begin{aligned} Cov(\omega_1^{*\bullet}, \eta_{11}^{*\bullet}) &= Cov\left(E\left(\omega_1^{*\bullet}|Y_1\right), E\left(\eta_{11}^{*\bullet}|Y_1\right)\right) + E\left[Cov(\omega_1^{*\bullet}, \eta_{11}^{*\bullet}|Y_1)\right] \\ &= Cov\left(E\left(\omega_1^{*\bullet}|Y_1\right), E\left(\eta_{11}^{*\bullet}|Y_1\right)\right) = 0, \end{aligned}$$

since $E\left(\eta_{11}^{*\bullet}|Y_1\right) = 0$. Consequently:

$$Cov(\widehat{A}_1^{*\bullet}, \widehat{A}_2^{*\bullet}) = 0.$$

□

Lemma 5.5.17. *The covariance of $\widehat{A}_1^{*\bullet}$ and $\widehat{A}_3^{*\bullet}$ is*

$$Cov\left(\widehat{A}_1^{*\bullet}, \widehat{A}_3^{*\bullet}\right) = D_9^{*\bullet}\frac{1}{N} + D_{10}^{*\bullet}\frac{1}{N^2}, \tag{5.23}$$

where

$$\begin{aligned} D_9^{*\bullet} &:= \int\int \gamma^{*\bullet}(z)\frac{v(y)-\mu_v}{(K_h * m)(y)}K_b(y-z)g(y)g(z)dydz - D_1^{*\bullet 2}, \\ D_{10}^{*\bullet} &:= \frac{K(0)}{b}\left[\int \rho^{*\bullet}(y)g(y)dy - \left(\int \frac{v(y)-\mu_v}{(K_h * m)(y)}g(y)dy\right)D_1^{*\bullet}\right] - D_3^{*\bullet} - D_9^{*\bullet}. \end{aligned}$$

Proof. Let us first consider $Cov(\widehat{A}_1^{*\bullet}, \widehat{A}_3^{*\bullet})$:

$$\begin{aligned} Cov(\widehat{A}_1^{*\bullet}, \widehat{A}_3^{*\bullet}) &= Cov\left(\frac{1}{N}\sum_{i=1}^{N}\omega_i^{*\bullet}, \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\tau_{jk}^{*\bullet}\right) = \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}Cov(\omega_i^{*\bullet}, \tau_{jk}^{*\bullet}) \\ &= \frac{1}{N^3}\left[N(N-1)(N-2)Cov(\omega_1^{*\bullet}, \tau_{23}^{*\bullet}) + N(N-1)Cov(\omega_1^{*\bullet}, \tau_{12}^{*\bullet})\right. \\ &+ \left. N(N-1)Cov(\omega_1^{*\bullet}, \tau_{21}^{*\bullet}) + N(N-1)Cov(\omega_1^{*\bullet}, \tau_{22}^{*\bullet}) + NCov(\omega_1^{*\bullet}, \tau_{11}^{*\bullet})\right] \\ &= \frac{N-1}{N^2}Cov(\omega_1^{*\bullet}, \tau_{12}^{*\bullet}) + \frac{N-1}{N^2}Cov(\omega_1^{*\bullet}, \tau_{21}^{*\bullet}) + \frac{1}{N^2}Cov(\omega_1^{*\bullet}, \tau_{11}^{*\bullet}), \end{aligned} \tag{A.207}$$

since $\omega_1^{*\bullet}$ and $\tau_{23}^{*\bullet}$ are independent and $\omega_1^{*\bullet}$ and $\tau_{22}^{*\bullet}$ are also independent.

Let us study now the terms in (A.207):

$$\begin{aligned} Cov(\omega_1^{*\bullet}, \tau_{12}^{*\bullet}) &= E\left[Cov(\omega_1^{*\bullet}, \tau_{12}^{*\bullet}|Y_1)\right] + Cov(E(\omega_1^{*\bullet}|Y_1), E(\tau_{12}^{*\bullet}|Y_1)) \\ &= Cov(E(\omega_1^{*\bullet}|Y_1), E(\tau_{12}^{*\bullet}|Y_1)) = 0, \end{aligned} \tag{A.208}$$

since $E(\tau_{12}^{*\bullet}|Y_1) = 0$.

Now consider

$$\begin{aligned}
Cov(\omega_1^{*\bullet}, \tau_{21}^{*\bullet}) &= E\left[Cov(\omega_1^{*\bullet}, \tau_{21}^{*\bullet}|Y_1)\right] + Cov(E(\omega_1^{*\bullet}|Y_1), E(\tau_{21}^{*\bullet}|Y_1)) \\
&= Cov(\omega_1^{*\bullet}, E(\tau_{21}^{*\bullet}|Y_1)) = E\left[\omega_1^{*\bullet} E(\tau_{21}^{*\bullet}|Y_1)\right] - E(\omega_1^{*\bullet}) E\left[E(\tau_{21}^{*\bullet}|Y_1)\right] \\
&= E\left[\omega_1^{*\bullet} E(\tau_{21}^{*\bullet}|Y_1)\right].
\end{aligned}$$

But

$$\begin{aligned}
E(\tau_{21}^{*\bullet}|Y_1) &= E\left[\frac{v(Y_2) - \mu_v}{(K_h * m)(Y_2)} \left(K_b(Y_2 - Y_1) - (K_b * g)(Y_2)\right)|Y_1\right] \\
&= \int \frac{v(y) - \mu_v}{(K_h * m)(y)} \left(K_b(y - Y_1) - (K_b * g)(y)\right) g(y) dy.
\end{aligned}$$

Thus

$$\begin{aligned}
&Cov(\omega_1^{*\bullet}, \tau_{21}^{*\bullet}) = \int \frac{(K_b * g)(z)}{(K_h * m)(z)} (v(z) - \mu_v) \\
&\cdot \left(\int \frac{v(y) - \mu_v}{(K_h * m)(y)} \left(K_b(y - z) - (K_b * g)(y)\right) g(y) dy\right) g(z) dz \\
&= \int\int \frac{(K_b * g)(z)}{(K_h * m)(z)} (v(z) - \mu_v) \frac{v(y) - \mu_v}{(K_h * m)(y)} \left(K_b(y - z) - (K_b * g)(y)\right) g(y) g(z) dy dz \\
&= \int\int \gamma^{*\bullet}(z) \frac{v(y) - \mu_v}{(K_h * m)(y)} K_b(y - z) g(y) g(z) dy dz - \left(\int \gamma^{*\bullet}(y) g(y) dy\right)^2. \qquad \text{(A.209)}
\end{aligned}$$

The last term in (A.207) is:

$$\begin{aligned}
&Cov(\omega_1^{*\bullet}, \tau_{11}^{*\bullet}) = E\left(\omega_1^{*\bullet} \tau_{11}^{*\bullet}\right) - E(\omega_1^{*\bullet}) E(\tau_{11}^{*\bullet}) \\
&= E\left[\gamma^{*\bullet}(Y_1) \frac{v(Y_1) - \mu_v}{(K_h * m)(Y_1)} \left(K_b(0) - (K_b * g)(Y_1)\right)\right] \\
&- E\left[\gamma^{*\bullet}(Y_1)\right] E\left[\frac{v(Y_1) - \mu_v}{(K_h * m)(Y_1)} \left(K_b(0) - (K_b * g)(Y_1)\right)\right] \\
&= \int \gamma^{*\bullet}(y) \frac{v(y) - \mu_v}{(K_h * m)(y)} \left(K_b(0) - (K_b * g)(y)\right) g(y) dy \\
&- \left(\int \gamma^{*\bullet}(y) g(y) dy\right) \left(\int \frac{v(y) - \mu_v}{(K_h * m)(y)} \left(K_b(0) - (K_b * g)(y)\right) g(y) dy\right) \\
&= \frac{K(0)}{b} \left[\int \rho^{*\bullet}(y) g(y) dy - \left(\int \frac{v(y) - \mu_v}{(K_h * m)(y)} g(y) dy\right) D_1^{*\bullet}\right] \\
&- \left[\int \alpha^{*\bullet}(y) g(y) dy - D_1^{*\bullet 2}\right] \\
&= \frac{K(0)}{b} \left[\int \rho^{*\bullet}(y) g(y) dy - \left(\int \frac{v(y) - \mu_v}{(K_h * m)(y)} g(y) dy\right) D_1^{*\bullet}\right] \\
&- \int \alpha^{*\bullet}(y) g(y) dy + D_1^{*\bullet 2}. \qquad \text{(A.210)}
\end{aligned}$$

Using (A.208), (A.209) and (A.210) in (A.207) gives (5.23). □

Lemma 5.5.18. *The covariance of* $\widehat{A}_2^{*\bullet}$ *and* $\widehat{A}_3^{*\bullet}$ *is*

$$Cov\left(\widehat{A}_2^{*\bullet}, \widehat{A}_3^{*\bullet}\right) \;=\; 0.$$

Proof. Finally, we consider $Cov(\widehat{A}_2^{*\bullet}, \widehat{A}_3^{*\bullet})$:

$$\begin{aligned}
Cov(\widehat{A}_2^{*\bullet}, \widehat{A}_3^{*\bullet}) &= Cov\left(\frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}\eta_{ij}^{*\bullet}, \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\tau_{kl}^{*\bullet}\right)\\
&= \frac{1}{N^3 n}\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{N} Cov(\eta_{ij}^{*\bullet}, \tau_{kl}^{*\bullet}) = \frac{1}{N^3 n}\left[nN(N-1)(N-2)Cov(\eta_{11}^{*\bullet}, \tau_{23}^{*\bullet})\right.\\
&+ nN(N-1)Cov(\eta_{11}^{*\bullet}, \tau_{12}^{*\bullet}) + nN(N-1)Cov(\eta_{11}^{*\bullet}, \tau_{21}^{*\bullet}) + nN(N-1)Cov(\eta_{11}^{*\bullet}, \tau_{22}^{*\bullet})\\
&+ \left. nNCov(\eta_{11}^{*\bullet}, \tau_{11}^{*\bullet})\right]\\
&= \frac{N-1}{N^2}Cov(\eta_{11}^{*\bullet}, \tau_{12}^{*\bullet}) + \frac{N-1}{N^2}Cov(\eta_{11}^{*\bullet}, \tau_{21}^{*\bullet}) + \frac{1}{N^2}Cov(\eta_{11}^{*\bullet}, \tau_{11}^{*\bullet}), \qquad \text{(A.211)}
\end{aligned}$$

since $\eta_{11}^{*\bullet}$ only depends on Y_1 and X_1 and $\tau_{23}^{*\bullet}$ only depends on Y_2 and Y_3, then $\eta_{11}^{*\bullet}$ and $\tau_{23}^{*\bullet}$ are independent. Similiarly $\eta_{11}^{*\bullet}$ and $\tau_{22}^{*\bullet}$ are also independent. Consequently:

$$Cov(\eta_{11}^{*\bullet}, \tau_{23}^{*\bullet}) = 0 \quad \text{and} \quad Cov(\eta_{11}^{*\bullet}, \tau_{22}^{*\bullet}) = 0.$$

Let us now consider the other three terms in (A.211):

$$\begin{aligned}
Cov(\eta_{11}^{*\bullet}, \tau_{12}^{*\bullet}) &= E\left[Cov(\eta_{11}^{*\bullet}, \tau_{12}^{*\bullet}|Y_1)\right] + Cov(E(\eta_{11}^{*\bullet}|Y_1), E(\tau_{12}^{*\bullet}|Y_1))\\
&= Cov(E(\eta_{11}^{*\bullet}|Y_1), E(\tau_{12}^{*\bullet}|Y_1)) = 0, \qquad \text{(A.212)}
\end{aligned}$$

since $E(\eta_{11}^{*\bullet}|Y_1) = 0$ and $E(\tau_{12}^{*\bullet}|Y_1) = 0$.

Let's deal now with the term $Cov(\eta_{11}^{*\bullet}, \tau_{21}^{*\bullet})$ in (A.211):

$$\begin{aligned}
Cov(\eta_{11}^{*\bullet}, \tau_{21}^{*\bullet}) &= E\left[Cov(\eta_{11}^{*\bullet}, \tau_{21}^{*\bullet}|Y_1)\right] + Cov(E(\eta_{11}^{*\bullet}|Y_1), E(\tau_{21}^{*\bullet}|Y_1))\\
&= Cov(E(\eta_{11}^{*\bullet}|Y_1), E(\tau_{21}^{*\bullet}|Y_1)) = 0, \qquad \text{(A.213)}
\end{aligned}$$

since $E(\eta_{11}^{*\bullet}|Y_1) = 0$.

We finally deal with the last term in (A.211):

$$\begin{aligned}
Cov(\eta_{11}^{*\bullet}, \tau_{11}^{*\bullet}) &= E\left[Cov(\eta_{11}^{*\bullet}, \tau_{11}^{*\bullet}|Y_1)\right] + Cov(E(\eta_{11}^{*\bullet}|Y_1), E(\tau_{11}^{*\bullet}|Y_1))\\
&= Cov(E(\eta_{11}^{*\bullet}|Y_1), E(\tau_{11}^{*\bullet}|Y_1)) = 0, \qquad \text{(A.214)}
\end{aligned}$$

since $E(\eta_{11}^{*\bullet}|Y_1) = 0$.

As a consequence of (A.212), (A.213) and (A.214) in (A.211), we obtain:

$$Cov(\widehat{A}_2^{*\bullet}, \widehat{A}_3^{*\bullet}) = 0.$$

□

Lemma 5.5.19. *The variance of $\widehat{A}^{\bullet}$ is*

$$Var\left(\widehat{A}^{*\bullet}\right) = D_4^{*\bullet}\frac{1}{n} + D_{11}^{*\bullet}\frac{1}{N} + D_5^{*\bullet}\frac{1}{Nn} + D_{12}^{*\bullet}\frac{1}{N^2} + D_8^{*\bullet}\frac{1}{N^3},$$

where

$$\begin{aligned} D_{11}^{*\bullet} &:= D_3^{*\bullet} + D_6^{*\bullet} + 2D_9^{*\bullet}, \\ D_{12}^{*\bullet} &:= D_7^{*\bullet} + 2D_{10}^{*\bullet}. \end{aligned}$$

Proof. Consequence of Lemmas 5.5.11, 5.5.13, 5.5.14, 5.5.15, 5.5.16, 5.5.17 and 5.5.18. □

Theorem 5.2.2. *Let us assume $h \to h_0 > 0$, $b \to b_0 > 0$, $n \to \infty$, $N/n \to \infty$, and Conditions A1 and A13-A15. The asymptotic mean squared error for the estimator $\hat{\hat{\mu}}_v$ in (5.4) is given by*

$$AMSE\left(\hat{\hat{\mu}}_v^{2,h,b}\right) = C_1^{*\bullet} + \frac{C_2^{*\bullet}}{n} + \frac{C_3^{*\bullet}}{N} + \frac{C_4^{*\bullet}}{Nn} + \frac{C_5^{*\bullet}}{N^2} + \frac{C_6^{*\bullet}}{N^3},$$

where the first two constants are

$$\begin{aligned} C_1^{*\bullet} &= \left(\frac{1}{c}\int \frac{K_b * g(y)}{K_h * m(y)}(v(y) - \mu_v)g(y)dy\right)^2, \\ C_2^{*\bullet} &= \frac{1}{c^2}\int \left(\int \frac{K_b * g(y)^2(v(y) - \mu_v)^2}{K_h * m(y)^4}K_h(y - z)g(y)dy - C_1^{*\bullet 1/2}\right)^2 m(z)dz \end{aligned}$$

and $C_3^{*\bullet}$, $C_4^{*\bullet}$, $C_5^{*\bullet}$ and $C_6^{*\bullet}$ are defined in the sketch of the proofs (Subsection 5.5.2)

Proof. Consequence of Lemmas 5.5.12 and 5.5.19. □

Appendix B

Resumo en galego

Este traballo pretende resumir os estudos desenvolvidos ao longo do período de realización da Tese de Doutoramento. Céntrase principalmente en propor diferentes métodos para detectar e corrixir o nesgo nun contexto de datos de gran volume. De xeito específico, a metodoloxía proposta emprégase para realizar catro aplicacións con diferentes conxuntos de datos reais: na primeira aplicación emprégase unha base de datos que contén o tempo de demora na chegada dos voos comerciais das aeroliñas estadounidenses, o segundo conxunto de datos recolle información sobre a calidade do aire na cidade da Coruña, o terceiro conta con información sobre campañas de retención de clientes da empresa de telecomunicacións Vodafone ES e na última aplicación, empréganse dous conxuntos de datos sobre a COVID-19.

Capítulo 1: Introdución

O primeiro capítulo da tese está dedicado a introducir ao lector no contexto no que foi desenvolta esta tese: os datos nesgados de gran volume (B3D, polas súas siglas en inglés). A Sección 1.1 comeza cunha presentación de diferentes exemplos motivadores atopados na literatura, como é o caso dos datos recollidos polo *StreetBump* ou o dos chíos orixinados por mor do furacán Sandy. Na Sección 1.2, considéranse outros traballos que tratan o problema do nesgo na mostraxe, mais nun contexto diferente. Para corrixir o nesgo presente nunha mostra de gran tamaño nesgada, propóñense na Subsección 1.2.1 dous posibles escenarios de actuación diferentes, nos que se precisa de certa información adicional. Dado que o estudo realizado se leva a cabo nun contexto non paramétrico, a estimación da densidade non paramétrica convértese nunha ferramenta moi importante ao longo da tese, motivo polo cal se inclúe na Subsección 1.2.2 unha revisión sobre este tema.

Capítulo 2: Estimación da media para datos nesgados de gran volume

No Capítulo 2 formalízase matematicamente o contexto de datos nesgados de gran volume proposto, denotando por w a función peso responsable do nesgo, a cal describe a relación entre as densidades involucradas, xa que se define como o cociente entre a densidade nesgada g e a densidade f da poboación verdadeira.

Neste capítulo introdúcese o problema concreto da estimación da media no contexto de datos nesgados de gran volume. Como a mostra nesgada de gran tamaño non é suficiente para levar a cabo dita estimación, propóñense dous escenarios diferentes nos que se conta con certa información adicional que permite corrixir o nesgo. No escenario 1, ademais da mostra de gran tamaño e nesgada, suponse que se observa unha mostra aleatoria simple (SRS, polas súas siglas en inglés) de tamaño pequeno procedente da verdadeira poboación; mentres que no escenario 2, suponse que se observa outra mostra adicional que é dúas veces nesgada. No primeiro escenario propóñense dous posibles estimadores para a media, considerando o caso improbable de que a función peso sexa coñecida. De xeito análogo, proponse un estimador da media no escenario 2 para o caso non realista no que a función peso é coñecida, sendo esta función neste escenario o cociente entre a densidade dúas veces nesgada e a densidade nesgada. O caso máis realista, no que a función peso é descoñecida, abórdase nos Capítulos 4 e 5 para cada escenario respectivamente, nos que tamén se trata o problema xeral da estimación da media dunha transformación dunha variable aleatoria continua.

Capítulo 3: Contrastes de nesgo

No Capítulo 3, proponse un procedemento para contrastar a presencia de nesgo ao traballar con bases de datos de gran tamaño. Este procedemento consiste en utilizar métodos de contraste de igualdade de distribucións para dúas mostras. Os métodos empregados son tests xa existentes: o test de Kolmogorov-

Smirnov, a proba U de Mann-Whitney e o criterio de Cramer-von Mises. Pese a que son métodos amplamente coñecidos, cómpre ter en conta a característica distintiva do noso contexto: o cociente entre os tamaños mostrais non tende a unha constante, senón que tende a infinito, xa que o tamaño da mostra nesgada tende a infinito moito máis rápido que o tamaño da mostra aleatoria simple, polo que se inclúen algúns resultados teóricos sobre o comportamento dos estatísticos nesta situación.

Neste capítulo emprégase tamén un método para contrastar a igualdade de medias, necesario cando se aborda o problema concreto da estimación da media, xa que o feito de que as mostras proveñan de distribucións diferentes non implica, necesariamente, que teñan medias diferentes. No caso desta análise, emprégase a adaptación de Welch da proba t de Student.

Ademais, neste capítulo propóñense diferentes índices que permiten medir a cantidade de nesgo en cada un dos escenarios. Todos os índices propostos son invariantes ante transformacións de localización e escala, o que ocasiona que non dependen das unidades nas que se mide a variable obxecto de estudo.

Por último, lévase a cabo un estudo de simulación. O modelo proposto para este estudo depende de dous parámetro que permiten controlar e regular o grao de nesgo, considerando así diferentes combinacións, dando lugar a situacións de moito nesgo, nesgo medio, nesgo case inapreciable e ausencia total de nesgo. Agás o test empregado para a implementación do criterio de Cramer-von Mises, cuxo mal funcionamento se debe, aparentemente, a un erro do paquete existente en R empregado, os restantes métodos de contraste de igualdade de distribucións así como o método de igualdade de medias permiten diferenciar correctamente as situacións de máis nesgo das de menos. Os diferentes índices empregados tamén permiten obter conclusións similares.

Capítulo 4: Estimación non paramétrica no escenario 1

O contido do Capítulo 4 foi publicado no artigo da revista *TEST* (Borrajo e Cao, 2021). Nel considérase como parámetro xeral a estimar a media dunha transformación dunha variable aleatoria continua, o que inclúe como casos particulares a estimación dos momentos (o que a súa vez inclúe a media), da función de distribución, ou da función característica, entre outros. En particular, neste capítulo abórdase o problema da estimación non paramétrica deste parámetro xeral no escenario 1 e tendo en conta o caso realista no que se supón que a función peso é descoñecida, polo que debe ser estimada. Para iso, ademais da mostra nesgada, emprégase a mostra aleatoria simple de tamaño pequeno da poboación verdadeira, observada neste escenario.

Neste capítulo proponse un novo estimador non paramétrico que incorpora a estimación de tipo núcleo da densidade e que depende de dous parámetros de suavizado, obtendo as súas propiedades asintóticas baixo condicións límite axeitadas sobre os dous tamaños mostrais e baixo as condicións asintóticas estándar sobre os dous parámetros de suavizado involucrados. Amósanse fórmulas explícitas para o caso particular da estimación da media nas que se observa que o estimador proposto non é quen de superar ás medias empíricas clásicas obtidas coas dúas mostras implicadas. Os resultados obtidos no estudo de simulación contradín este feito, pois amosan como o estimador da media mellora o comportamento dos estimadores clásicos dispoñibles para escollas axeitadas dos dous parámetros de suavizado implicados. Analizando os valores óptimos obtidos para estes parámetros, obsérvase que semellan contradicir a condición clásica na que se supón que deben tender a cero, feito que motivou o estudo das propiedades asintóticas do estimador proposto baixo as condicións non estándar de que os dous parámetros de suavizado tenden a valores constantes positivos. Trátase dunha condición moi sorprendente, xa que baixo este suposto o estimador da densidade non é consistente. Obtense a expresión do erro cadrático medio asintótico do estimador do parámetro xeral baixo estas condicións non estándar. Analizando o comportamento teórico do estimador para o caso particular da estimación da media, obsérvase como o estimador proposto neste caso si pode ter un mellor comportamento que os estimadores clásicos para unha escolla axeitada dos parámetros de suavizado, o que xustifica o bo comportamento observado no estudo de simulación.

Neste escenario amósase a importancia de escoller de forma axeitada o parámetro de suavizado da mostra nesgada, pois en caso contrario o comportamento do estimador proposto pode empeorar significativamente. Unha vez escollido ese parámetro de suavizado, o estimador ten un bo comportamento para un rango moi amplo de valores do parámetro de suavizado da mostra de pequeno tamaño procedente da poboación xenuina.

Finalmente, proponse un método para a escolla automática dos parámetros de suavizado. Trátase dun algoritmo bootstrap que aproxima o erro cadrático medio do estimador proposto, cuxa minimización proporciona os selectores automáticos de ditos parámetros. O bo funcionamento deste algoritmo, posteriormente empregado nas aplicacións a datos reais, queda probado mediante un estudo de simulación realizado cos modelos propostos no Capítulo 3.

No apéndice A.1 recóllense as demostracións dos resultados teóricos presentados neste capítulo.

Capítulo 5: Estimación non paramétrica no escenario 2

O Capítulo 5 segue liñas paralelas ás do Capítulo 4 pero abordando o problema da estimación do parámetro xeral no escenario 2, baixo o suposto de que a función peso é descoñecida. Neste caso, para estimar a función peso, ademais da mostra nesgada de gran tamaño, suponse que se observa unha mostra dobremente nesgada de tamaño pequeno, o que permite definir un estimador do parámetro xeral neste escenario, o cal dependerá, de novo, de dous prámetros de suavizado. O comportamento asintótico do estimador non paramétrico proposto neste escenario analízase, en primeiro lugar, baixo as condicións asintóticas estándar para os dous parámetros de suavizado. Non obstante, os resultados obtidos no estudo de simulación semellan contradicir, de novo, a suposición de que os parámetros de suavizado tenden a cero, mostrando o bo comportamento do estimador proposto en condicións non estándar, o que leva a estudar as propiedades asintóticas do estimador proposto baixo a suposición de que os parámetros de suavizado tenden a constantes positivas.

Neste escenario, ao contrario do que ocorría no escenario 1, obsérvase a importancia de escoller de forma axeitada o parámetro de suavizado da mostra dobremente nesgada, sendo o comportamento do estimador proposto moi estable respecto ao parámetro de suavizado da mostra de gran tamaño nesgada unha soa vez.

Finalmente, proponse neste escenario un novo algoritmo bootstrap para aproximar o erro cadrático medio do estimador proposto cando non se observa unha mostra aleatoria simple da poboación verdadeira, o cal permite obter selectores automáticos para os dous parámetros de suavizado. Para implementar este novo algoritmo emprégase a relación entre ambos escenarios, o cal permite expresar a densidade da poboación verdadeira como función das densidades das mostras observadas no escenario 2. O bo funcionamento deste algoritmo, posteriormente empregado nas aplicación cos datos reais da COVID-19, queda probado mediante un estudo de simulación con modelos totalmente análogos aos propostos no Capítulo 3.

No apéndice A.2 recóllense as demostracións dos resultados teóricos presentados neste capítulo.

Capítulo 6: Aplicacións a datos reais

No Capítulo 6 os métodos propostos nos capítulos 4 e 5 aplícanse a varios conxuntos de datos reais. En particular, coas catro bases de datos realízanse aplicacións dentro do contexto proposto no escenario 1, mentres que o contexto definido no escenario 2 dáse na última aplicación a datos reais, a correspondente aos datos da COVID-19.

En primeiro lugar, na Sección 6.1, considérase un conxunto de datos que contén información sobre a chegada e saída de todos os voos comerciais das compañías aéreas estadounidenses dende outubro de 1987 a maio de 2018. Trátase dun gran conxunto de datos con case 180 millóns de rexistros. Neste caso, temos interese en estimar a media e a desviación típica do tempo de demora (en minutos) na chegada dos voos dos Estados Unidos en 2017. Para iso, empregamos o estimador proposto considerando como mostra nesgada de gran tamaño a mostra de datos correspondente aos tempos de demora dos voos no ano 2016 (máis de cinco millóns de rexistros) e supomos que só os tempos de demora na chegada dos voos do 11 de xaneiro de 2017 está dispoñible, o que constitúe algo cercano a unha mostra aleatoria simple (algo menos de catorce mil rexistros). Para a estimación da desviación típica, abonda con considerar o caso particular do momento de segunda orde no estimador xeral para obter a varianza e logo considerar a súa raíz cadrada. Como moi boas estimacións da media e varianza reais, tomamos as correspondentes aos tempos de demora dos voos de todo o ano 2017, información que non sería coñecida ata finais de ese ano. Os resultados obtidos amosan como o estimador proposto pode funcionar moi ben para unha escolla axeitada dos parámetros de suavizado, tanto na estimación da media como da desviación típica.

En segundo lugar, na Sección 6.2 abórdase o problema da contaminación atmosférica en cidades intelixentes (*smart cities*). O contido desta aplicación foi publicado no artigo da revista *Electronics* (Borrajo e Cao, 2020). Nel considérase unha base de datos con información sobre a calidade do aire na cidade da Coruña, a cal contén rexistros por hora da temperatura e dos niveis de ozono e de dióxido de nitróxeno no aire ao longo dos últimos 15 anos. Neste caso, temos interese en aplicar os métodos propostos no escenario 1 para estimar a media e a función de distribución destes dous contaminantes cando a temperatura é maior ou igual a 30°C, xa que se pensa que as altas temperaturas poden propiciar un aumento do nivel destes contaminantes no aire. Como mostra de gran tamaño nesgada considéranse todos os rexistros dos últimos 15 anos e como mostra aleatoria simple os datos correspondentes cando a temperatura foi maior ou igual a 30°C nese período de tempo. Obsérvase como as dúas densidades son moi similares para o dióxido de nitróxeno, mentres que se diferencian bastante para o ozono, o cal era de esperar xa que estudos previos realizados por outros autores mostran como as altas temperaturas propician un aumento do nivel de ozono, mentres que o efecto sobre os óxidos de nitróxeno é incerto. Unha vez obtidas as estimacións empregando os nosos métodos, obsérvase como no caso do dióxido de nitróxeno (pouco nesgo) o comportamento dos nosos estimadores non difire moito das estimacións obtidas cos estimadores clásicos, mentres que no caso do ozono, a estimación proposta, aínda que próxima ao comportamento do estimador obtido coa mostra aleatoria simple, distánciase máis das estimacións clásicas debido a maior presenza de nesgo.

Na Sección 6.3, considérase un conxunto de datos reais da compañía de telecomunicacións Vodafone ES con información sobre campañas de retención de clientes. O contido desta aplicación foi publicado en Borrajo e Cao (2021). O conxunto de datos consta de case 2,5 millóns de rexistros e 176 variables con información sobre os clientes da empresa. Neste caso, constrúese unha nova variable, o *index*, a partir das 14 variables que mellor reflicten a tendencia do cliente a abandonar a empresa. Temos interese en estimar a media e a función de distribución acumulada dese *index*, empregando para iso a información do grupo obxectivo (mostra nesgada) e do grupo de control universal (mostra aleatoria simple) das campañas de retención. Cabe destacar que esta segunda mostra tería un tamaño de 1466 clientes, o cal resulta ser bastante grande no caso concreto que nos ocupa. Neste caso, dado que nos interesa estimar tanto a media como a función de distribución, e co obxectivo de obter estimacións coherentes de ambas medidas e evitar así un maior custe computacional, decantámonos por empregar unha nova versión do algoritmo bootstrap para a distancia de Kolmogorov-Smirnov do estimador da función de distribución, cuxos resultados proporcionan os valores dos parámetros de suavizado óptimos que dan lugar a un bo estimador da función de distribución e, en consecuencia, da media. Ambas estimacións apenas amosan diferenzas respecto dos estimadores clásicos obtidos a partir da mostra pequena, o cal era de esperar, xa que o tamaño mostral é bastante elevado. Para amosar o efecto que ten o tamaño da mostra pequena no noso estimador, repetimos o procedemetno pero cunha submostra de 100 clientes do grupo de control universal, obtendo novas estimacións máis diferenciadas das clásicas e próximas as obtidas coa mostra orixinal, cuxo bo comportamento xa fora observado.

Finalmente, na Sección 6.4 faise unha aplicación con dúas bases de datos reais sobre COVID-19 coas que se pretende facer estimacións sobre a idade de contaxio por coronavirus en ambos escenarios. No escenario 2 consideramos a base de datos correspondente aos casos identificados pola Rede Nacional de Vixilancia Epidemiolóxica (RENAVE) a través do Sistema de Vixilancia de España (SiViEs) e os datos proporcionados polas comunidades autónomas. Ademais da variable idade, esta base de datos ofrece información sobre se as persoas identificadas precisaron ser hospitalizadas. Estes datos están claramente nesgados debido a que, como xa sabemos, hai un número elevado de casos asintomáticos entre a poboación que non son detectados polo sistema de saúde. A principal característica que nos permite detectar a alguén infectado con coronavirus en comparación cun asintomático é un aumento da gravidade dos seus síntomas, a mesma causa que leva á hospitalización dun paciente xa identificado, polo que é razoable pensar que a relación entre os casos identificados e os casos totais, incluídos os asintomáticos (que corresponderían ao escenario 1) é similar á relación entre pacientes hospitalizados e casos identificados (escenario 2). Tomamos como mostra nesgada de gran tamaño as idades dos casos identificados polo SiViES ata o 11 de maio de 2020 e como mostra dobremente nesgada a submostra correspondente ás idades dos pacientes hospitalizados ata ese momento. Dado que neste caso a mostra aleatoria simple non estaría dispoñible, procedemos segundo o algoritmo bootstrap do escenario 2. Os resultados obtidos amosan unha estimación de 53,84 anos para a idade dos contaxiados pola COVID-19, fronte á estimación de 62 anos obtida coa mostra nesgada e a de 67,61 anos obtida coas mostra dobremente nesgada. É preciso mencionar que no caso particular dos datos españois realizouse un Estudo Nacional Epidemiolóxico da infección por SARS-CoV2 (o ENECOVID), que á súa vez nos permitiu realizar unha aplicación no escenario 1, xa que o deseño deste estudo proporciona unha mostra de tamaño pequeno moi próxima a unha mostra aleatoria simple da poboación que nos interesa. Nesta análise obtivemos unha media desta mostra de 48,57 anos. Supoñendo que a verdadeira media é próxima a este valor, observamos como a estimación obtida no escenario 2 semella funcionar ben, corrixindo en certa medida o nesgo das medias de ambas mostras involucradas: a nesgada e a dobremente nesgada.

Capítulo 7: Conclusión e traballo futuro

No Capítulo 7 amósanse as principais conclusións derivadas da realización desta tese e fanse algúns comentarios sobre as liñas de investigación futuras. Considérase a posibilidade de estender a metodoloxía proposta ao estudo de variables categóricas e a dimensión múltiple. Outra idea é aplicar os métodos propostos á estimación da matriz de varianzas-covarianzas e á estimación da matriz de correlacións, o que nos permitiría realizar análise de compoñentes principais e análise lineal discriminante. Ademais, recompilarase todo o software que se foi elaborando nun formato axeitado para a creación dun paquete de R que permita aplicar as técnicas propostas a conxuntos de datos reais.

Bibliography

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 25.

Ameer, S., Shah, M. A., Khan, A., Song, H., Maple, C., Islam, S. U., & Asghar, M. N. (2019). Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access*, *7*, 128325–128338.

Anderson, T. W. (1962). On the distribution of the two-sample Cramer-von Mises criterion. *The Annals of Mathematical Statistics*, *33*(3), 1148–1159.

Ayuntamiento de A Coruña. (2020). *Coruña sostenible.* `http://coruna.es/infoambiental/`.

Borrajo, L., & Cao, R. (2020). Big-but-biased data analytics for air quality. *Electronics*, *9*(9), 1551.

Borrajo, L., & Cao, R. (2021). Nonparametric estimation for big-but-biased data. *TEST*, to appear.

Calissano, A., Vantini, S., & Arnaboldi, M. (2018). An elephant in the room: Twitter sampling methodology. *MOX-Report*, *16/2018*.

Cao, R. (2015). Inferencia estadística con datos de gran volumen. *Gaceta de la Real Sociedad Matematica Española*, *18*(2), 393–417.

Cao, R., & Borrajo, L. (2018). Nonparametric mean estimation for big-but-biased data. In E. Gil, E. Gil, J. Gil, & M. Á. Gil (Eds.), *The mathematics of the uncertain* (pp. 55–65). Cham: Springer International Publishing.

Cardelino, C., & Chameides, W. (1990). Natural hydrocarbons, urbanization, and urban ozone. *Journal of Geophysical Research: Atmospheres*, *95*(D9), 13971–13979.

Carvalho, L. (2015). An improved evaluation of Kolmogorov's distribution. *Journal of Statistical Software*, *65*(3), 1–7.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., ... Scholl, H. J. (2012). Understanding smart cities: An integrative framework. In *2012 45th Hawaii international conference on system sciences* (pp. 2289–2297).

Cramer, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, *1928*(1), 13–74.

Crawford, K. (2013). The hidden biases in big data. *Harvard Business Review*. Blog, 1 April, `https://hbr.org/2013/04/the-hidden-biases-in-big-data`.

Cristóbal, J. A., & Alcalá, J. T. (2001). An overview of nonparametric contributions to the problem of functional estimation from biased data. *TEST*, *10*(2), 309–332.

Delyon, B., & Portier, F. (2016). Integral approximation by kernel smoothing. *Bernoulli*, *22*(4), 2177–2208.

Deville, J., & Särndal, C. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*(418), 376–382.

Devroye, L. (1986). *Non-uniform random variate generation.* Springer.

Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, *20*(3), 393–403.

Feller, W. (1948). On the Kolmogorov-Smirnov limit theorems for empirical distributions. *The Annals of Mathematical Statistics*, *19*(2), 177–189.

Fisher, R. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, *6*(1), 13–25.

Genton, M. G., Kim, M., & Ma, Y. (2012). Semiparametric location estimation under non-random sampling. *Stat*, *1*(1), 1–11.

Gill, R. D., Vardi, Y., & Wellner, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *The Annals of Statistics*, *16*(3), 1069–1112.

Hargittai, E. (2015). Is bigger always better? potential biases of big data derived from social network sites. *The Annals of the American Academy of Political and Social Science*, *659*(1), 63–76.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748–758.

Jhun, I., Fann, N., Zanobetti, A., & Hubbell, B. (2014). Effect modification of ozone-related mortality risks by temperature in 97 US cities. *Environment International*, *73*, 128–134.

Kasuya, E. (2001). Mann-Whitney U test when variances are unequal. *Animal Behaviour*, *61*(6), 1247–1249.

Kök, İ., Şimşek, M. U., & Özdemir, S. (2017). A deep learning model for air quality prediction in smart cities. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 1983–1990).

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, *4*, 83–91.

Kott, P. S. (2016). Calibration weighting in survey sampling. *Wiley Interdisciplinary Reviews: Computational Statistics*, *8*(1), 39–53.

Li, Q., & Racine, J. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis*, *86*(2), 266–292.

Lim, C., Kim, K.-J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities*, *82*, 86–99.

Lloyd, C. J., & Jones, M. (2000). Nonparametric density estimation from biased data with unknown biasing function. *Journal of the American Statistical Association*, *95*(451), 865–876.

Ma, Y., Genton, M. G., & Tsiatis, A. A. (2005). Locally efficient semiparametric estimators for generalized skew-elliptical distributions. *Journal of the American Statistical Association*, *100*(471), 980–989.

Ma, Y., Kim, M., & Genton, M. G. (2013). Semiparametric efficient and robust estimation of an unknown symmetric population under arbitrary sample selection bias. *Journal of the American Statistical Association*, *108*(503), 1090–1104.

Mack, Y. P., & Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *61*(3), 405–415.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50–60.

Martínez-España, R., Bueno-Crespo, A., Timón-Pérez, I. M., Soto, J. A., Ortega, A. M., & Cecilia, J. M. (2018). Air-pollution prediction in smart cities through machine learning methods: A case of study in Murcia, Spain. *Journal of Universal Computer Science*, *24*(3), 261–276.

Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68–78.

Meleux, F., Solmon, F., & Giorgi, F. (2007). Increase in summer European ozone amounts due to climate change. *Atmospheric Environment*, *41*(35), 7577–7587.

Montanari, G. E., & Ranalli, M. G. (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association*, *100*(472), 1429–1442.

Osman, A. M. S. (2019). A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, *91*, 620–633.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, *33*(3), 1065–1076.

Patil, G. P., & Rao, C. R. (1978). Weighted distributions and size biased sampling with applications to wildlife populations and human families. *Biometrics*, *34*, 179–189.

Ramos, F., Trilles, S., Muñoz, A., & Huerta, J. (2018). Promoting pollution-free routes in smart cities using air quality sensor networks. *Sensors*, *18*(8), 2507.

Rosenblatt, M. (1952). Limit theorems associated with variants of the von Mises statistic. *The Annals of Mathematical Statistics*, *23*(4), 617–623.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, *27*, 832–837.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* New York: Chapman and Hall.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Moscow University*, *2*(2), 3–14.

Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, *19*(2), 279–281.

Vardi, Y. (1985). Empirical distributions in selection bias models. *The Annals of Statistics*, *13*(1), 178–203.

von Mises, R. (1928). Statistik und wahrheit. *Julius Springer*, *20*.

Welch, B. L. (1947). The generalization of student's' problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28–35.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, *38*(3/4), 330–336.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83.

World Health Organization. (2020). *Air pollution.* `http://www.who.int/airpollution/en/`.

Zheng, Y., Liu, F., & Hsieh, H.-P. (2013). U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1436–1444).