



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Construcción de funciones de estimación múltiple protegidas en modelos con verosimilitud factorizada

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en
el área Ciencias Matemáticas

Fernanda Julieta Molina

Directores de tesis: Dra. Andrea Rotnitzky

Dra. Mariela Sued

Consejero de estudios: Dra. Mariela Sued.

Lugar de trabajo: Instituto de Cálculo, FCEyN, UBA.

Buenos Aires, Noviembre 2013.

Construcción de funciones de estimación múltiple protegidas en modelos con verosimilitud factorizada

Resumen

En los últimos años, los métodos de estimación doble protegidos dieron lugar al desarrollo de propuestas múltiple protegidas. Diversos autores han presentado estimadores consistentes en más escenarios que los contemplados por los estimadores doble protegidos. Estas propuestas de estimación, lejos de construirse bajo una teoría general, son el resultado de procedimientos ad-hoc creados para cada modelo considerado.

En este trabajo, desarrollamos un marco teórico que explica la existencia de estimadores múltiple protegidos en varias de las propuestas de estimación múltiple protegidas ya existentes, en modelos donde la verosimilitud se factoriza. Dicha teoría provee condiciones suficientes bajo las cuales es posible la construcción de ecuaciones de estimación que, bajo condiciones de regularidad, tienen soluciones que constituyen estimadores múltiple protegidos. Presentamos también condiciones suficientes bajo las cuales estimar de manera múltiple protegida ciertos parámetros de ruido, cuya estimación es necesaria para poder estimar de manera múltiple protegida al parámetro de interés.

Además, aplicamos los métodos desarrollados para estimar en forma cuádruple protegida la componente lineal bajo un modelo de regresión parcialmente lineal con datos faltantes, ya que dicho modelo se inscribe en el marco de la teoría desarrollada. Incluimos también, los resultados de un estudio de Monte Carlo realizado bajo el modelo mencionado.

Palabras Clave: Estimación Múltiple Protegida, Verosimilitud Factorizada, Estadística Semiparamétrica, Datos Faltantes, Modelos Parcialmente Lineales.

Constructing multiple robust estimating functions in factorized likelihood models

Abstract

In recent years, double robust procedures have led to the development of multiple robust approaches. Different authors have shown consistent estimators that confer even more protection than double robust estimators. These estimation proposals, far from being constructed under a general theory, are the result of ad-hoc procedures created for each model considered.

In this work, we develop a theoretical framework that explains the existence of multiple robust estimators in many of the multiple robust procedures existing in factorized likelihood models. This theory provides sufficient conditions for the construction of multiple robust estimating equations for the parameter of interest. We also provide sufficient conditions under which it is possible to construct multiple robust estimators for certain parameters needed for the estimation of the parameter of interest.

We apply this procedure to obtain a multiple robust estimator for the linear component under a partial linear regression model with missing outcomes, since this model fits in the framework of the developed theory. Also, we present the results of a Monte Carlo simulation study performed under the same model.

Key Words: Multiple Robust Estimation, Factorized Likelihood, Semiparametric Statistics, Missing Data, Partially Linear Regression Model With Missing Outcomes.

Agradecimientos

Quiero agradecer a mis compañeros doctorandos y a mis compañeros de trabajo en el Instituto de Cálculo, quienes me ayudaron en distintas cuestiones a lo largo de mis estudios de posgrado, en particular agradezco a L. Babino, L. Bali, J. Galofré, V. Pater-nostro, M. Prieto, D. Rodriguez, P. Spano, M.E. Szretter, M. Valdora, A. Ferrari, M. Mazzitelli, F. Statti.

También quiero agradecer a mis profesores en esta segunda etapa de mis estudios, especialmente a G. Boente y a V. Yohai por los cursos dictados en estos últimos años, los cuales fueron de gran utilidad para mi formación doctoral.

Agradezco también a A. Rotnitzky por su tiempo y por haberme propuesto temáticas de estudio muy interesantes.

Un agradecimiento especial a varios de los coordinadores del Departamento de Matemática del Ciclo Básico Común por haberme recibido nuevamente como docente, en especial a F. Gutierrez, J. C. Pedraza, B. Russo, A. Wykowski, L. D´alfonso, M. Mereb y a E. Werner.

Quiero agradecer también a mi familia y amigos por el apoyo y el cariño. Especial-mente a mi compañero, Emilio Noce, por su ayuda incondicional y a mi futuro hijo Nicanor por haberme dado el impulso final para terminar mis estudios de posgrado.

Finalmente quiero dar mi más sincero agradecimiento a la Doctora Mariela Sued por tantas cosas...por su compromiso, por brindarme su amistad, por su entusiasmo, su profesionalismo y sobre todo por darme la posibilidad de estudiar bajo su tutela lo cual es una de las cosas más lindas que le pueden a pasar a alguien que disfruta de la matemática.

Índice general

1. Introducción	2
1.1. Notación	9
2. Ejemplos-Estudio de Monte Carlo	12
2.1. Ejemplo 1: datos longitudinales.	12
2.1.1. Estimador de regresión.	14
2.1.2. Estimador IPW	15
2.1.3. Estimación doble protegida	16
2.1.4. Propuesta cuádruple protegida.	18
2.2. Ejemplo 2: modelo de regresión parcialmente lineal con datos faltantes.	20
2.2.1. Modelo de regresión parcialmente lineal.	20
2.2.2. Modelo de regresión parcialmente lineal con datos faltantes . .	27
2.2.3. Estimadores basados en una primera regresión.	30
2.2.4. Estimadores basados en pesos.	32
2.2.5. Propuesta cuádruple protegida.	35
2.3. Estudio de Monte Carlo	37
3. Supuestos y Definiciones	56
3.1. Definición del funcional de interés.	56
3.1.1. Supuestos.	56
3.1.2. Supuestos en los ejemplos.	57
3.2. Estimadores múltiple protegidos en modelos de intersección y unión. . .	59
3.3.1. Función de estimación insesgada múltiple protegida en modelos de unión intersección.	63

3.4.1.	Función de estimación insesgada múltiple protegida para los parámetros de ruido en modelos de unión intersección.	65
4.	Funciones de estimación múltiple protegidas	69
4.1.	Modelos convexos y condiciones.	69
4.2.	Resultados previos.	71
4.3.	Construcción de funciones de estimación insesgadas múltiple protegidas para $\beta(g)$ en modelos de unión e intersección.	79
5.	Funciones de estimación para los parámetros de ruido	87
5.1.	Construcción	88
5.2.	Reconstrucción de los procedimientos dados en los ejemplos.	94
5.2.1.	Ejemplo 1	94
5.2.2.	Ejemplo 2	96
6.	Problemas abiertos	99
6.1.	Aplicación del procedimiento múltiple protegido propuesto en este trabajo a otros modelos inferenciales.	99
6.2.	Mejorando el desempeño en términos de eficiencia de los estimadores múltiple protegidos construídos en este trabajo.	100
6.3.	Estimadores de tipo Plug-In.	102
6.4.	Relajando algunas hipótesis.	104
6.5.	Estimación múltiple protegida en modelos para medidas repetidas con non-ignorable non-response.	104
7.	Apéndice A: Espacio Tangente	105
7.1.	Modelos para la distribución de un elemento aleatorio.	105
7.1.1.	Modelos Paramétricos.	108
7.2.	Espacio Tangente.	109
7.2.1.	Submodelos unidimensionales - Curvas	109
7.3.1.	Conjunto tangente relativo - Tangente maximal	111
7.6.1.	Caracterización de la proyección bajo la Condición 1.	113
7.6.2.	Revisión de los Supuestos hechos a lo largo de la Sección 4.2.	117
8.	Apéndice B	122

8.1.	Demostraciones correspondientes al Ejemplo 1.	122
8.1.1.	Demostración de la Proposición 2.1.1.	122
8.1.2.	Demostración del Lema 2.1.1	123
8.1.3.	Demostración del Lema 2.1.3	123
8.2.	Demostraciones correspondientes al Ejemplo 2.	125
8.2.1.	Demostración de la Proposición 2.2.1.	125
8.2.2.	Demostración del Lema 2.2.1.	126
8.3.	Demostraciones correspondientes al Capítulo 4.	127
8.3.1.	Demostración del Lema 4.1.1	127
8.3.2.	Demostración del Teorema 4.2.1.	128
8.3.3.	Demostración del Corolario 4.2.1.	131
8.3.4.	Demostración del la Proposición 4.3.1.	131
8.3.5.	Demostración de la Proposición 4.3.2.	132
8.4.	Demostraciones correspondientes al Capítulo 5.	133
8.4.1.	Demostración del la Proposición 5.1.1.	134
8.4.2.	Demostración del Lema 5.1.1.	134
8.4.3.	Demostración del Lema 5.1.2.	135
8.4.4.	Demostración del la Proposición 5.1.2.	136

Capítulo 1

Introducción

En el ámbito de la bioestadística, frecuentemente se quiere hacer inferencia sobre cierto parámetro β finito dimensional asociado a la distribución F que genera los datos. Muchas veces, la estimación del parámetro de interés requiere estimaciones adicionales de funcionales que no son relevantes en sí mismos (llamados funcionales nuisance o de ruido). A modo de ejemplo, puede ser necesario tener que estimar una esperanza condicional o algún otro funcional infinito dimensional para poder estimar a β . Por lo general, la alta dimensión de las observaciones dificulta la utilización de procedimientos no paramétricos para estimar los funcionales de ruido, debido a la *maldición de la dimensión*. Ante esta dificultad, muchos autores optan por reducir la complejidad del problema asumiendo un submodelo paramétrico para los funcionales de ruido, de forma tal que resulten estimables. Para muchos parámetros de interés β , ocurre que existen distintas maneras viables de reducción de la complejidad del problema. Cada posible manera implica asumir un modelo (en la práctica, generalmente paramétrico) para un cierto funcional de ruido. Luego, cada posible reducción da lugar a diferentes estimadores, cada uno de ellos consistente si la distribución que genera los datos pertenece al submodelo asumido. Resulta entonces crucial decidir cuál es la reducción asumida a la hora de exhibir estimadores consistentes.

En los últimos años, cuando el escenario es propicio, algunos autores optaron por dejar de lado la discusión acerca de cuál de las posibles reducciones es verdadera, procurando estimadores que resulten consistentes si alguno de los modelos asumidos es correcto dando origen a los así llamados estimadores doble o múltiple protegidos. Estos estimadores surgieron con el propósito de resolver problemas relacionados a la inferencia causal y a la inferencia con datos faltantes. A menudo en estas áreas, el parámetro de interés está definido bajo un modelo base no-paramétrico o semi-paramétrico.

Este modelo base incluye las suposiciones que el investigador está dispuesto a asumir sobre el problema científico en consideración, mientras que el parámetro de interés finito dimensional, generalmente cuantifica la respuesta al problema concreto de interés. Sin

embargo, en la inferencia causal así como en el análisis de datos faltantes, es necesario hacer suposiciones adicionales más allá de las suposiciones codificadas en el modelo base. Estas suposiciones no son de interés científico pero, debido a la maldición de la dimensionalidad, resultan imprescindibles para poder estimar el parámetro de interés. En muchos casos, existe más de una vía para imponer suposiciones adicionales, y cada una de ellas implica un submodelo diferente del modelo base. Por ejemplo, como veremos más adelante en esta misma introducción, en datos faltantes, una vía es imponer suposiciones que caracterizan la asociación entre la variable respuesta (faltante en algunos individuos de la muestra) y las variables observadas. Una segunda vía es imponer suposiciones que caracterizan la dependencia de la probabilidad de no-respuesta en las variables observadas. Cualquiera de estas dos suposiciones sería evitable si uno contara con una muestra sin datos faltantes. Como estas suposiciones no se relacionan con el problema científico de interés, es deseable relajarlas en la medida de lo posible, de modo de llegar a obtener una inferencia bajo suposiciones parecidas a las que uno hubiera impuesto en el caso de no tener variables faltantes. La búsqueda de procedimientos que relajen estas suposiciones adicionales ha sido un área central dentro de la investigación sobre metodologías para datos faltantes e inferencia causal y en particular ha dado origen a los, así llamados, métodos doble protegidos. Estos métodos relajan la dependencia en estas suposiciones adicionales porque ofrecen la oportunidad de evitar comprometerse con una de las dos vías de modelado adicional: los mismos proveen estimadores que son consistentes siempre que uno de los submodelos sea cierto pero, sin requerir que el investigador sepa de antemano cual de ellos es cierto.

En los últimos años han surgido los así llamados estimadores múltiple protegidos que relajan aún más las suposiciones impuestas por los estimadores doble protegidos, ya que son consistentes siempre que uno de $m > 2$ submodelos sea verdadero sin requerir de antemano saber cuál de ellos es correcto. Diferentes autores han presentado procedimientos ad hoc para obtener estimadores múltiple protegidos en distintos escenarios. Una contribución fundamental de esta tesis es la elaboración de una teoría matemática rigurosa que explica la existencia de estimadores múltiple protegidos en muchos de los casos ya considerados y que provee condiciones suficientes bajo las cuales es posible la construcción de ecuaciones de estimación cuyas soluciones permiten definir estimadores múltiple protegidos. Una distinción fundamental entre los estimadores doble protegidos y los múltiple protegidos es la estimación de los parámetros de ruido bajo cada uno de los submodelos. En estimación doble protegida queda claro que es posible estimar los parámetros de ruido sin necesidad de imponer suposiciones adicionales. En estimación múltiple protegida esto no es así. Otra contribución fundamental de esta tesis es la derivación de condiciones suficientes bajo las cuales los parámetros de ruido también pueden estimarse de manera múltiple protegida, sin requerir suposiciones adicionales más allá de las requeridas para estimar el parámetro de interés. En particular, la teoría desarrollada en esta tesis abarca también la construcción de ecuaciones de estimación

múltiple protegidas para los parámetros de ruido. Ecuaciones de esta naturaleza han sido propuestas de manera ad hoc en algunos trabajos publicados recientemente pero no existía hasta el momento una teoría que explicase su construcción sistemática.

Vamos ahora a presentar un primer ejemplo de los procedimientos descritos, en el marco de la inferencia estadística con datos faltantes. Imaginemos que se dispone de una muestra aleatoria en la que cierta variable L es observada en cada unidad de la muestra, mientras que Y , la variable respuesta de interés, sólo es observada en algunos individuos. Denotamos por R a la variable binaria que indica cuando la respuesta es observada, es decir: $R = 1$ si Y es observado y $R = 0$ caso contrario. A partir de esta información, se desea estimar $\beta_0 = E(Y)$. Para que este parámetro quede determinado a partir de la distribución de los datos observados (identificabilidad, Rubin (1976)), se asume que las respuestas son perdidas al azar (missing at random, MAR). Esta hipótesis establece que Y es condicionalmente independiente de R , dado L . En tal caso, tenemos que $Y|L \sim Y|L, R = 1$, de donde concluimos que $E(Y) = E(E(Y|L, R = 1))$, quedando demostrada la identificabilidad. A partir de esta última fórmula, una propuesta de estimación para β_0 es la siguiente: considerar un modelo paramétrico para $E(Y|L, R = 1)$ dado por

$$E(Y|L, R = 1) = b(L, \tau_0), \quad (1.1)$$

donde b es una función suave conocida salvo por el parámetro $\tau_0 \in \mathbb{R}^\ell$ y estimar τ_0 con $\hat{\tau}$, el estimador “por defecto”, es decir el que se utilizaría comúnmente en ausencia de datos faltantes para estimar a τ_0 bajo el modelo, pero utilizando los elementos de la muestra para los cuales las respuestas son observadas. Para fijar ideas, si consideremos a modo de ejemplo $b(L, \tau_0) = \tau_0^t L$, entonces, $\hat{\tau}$ es el estimador de cuadrados mínimos calculado en la submuestra sin datos faltantes. Finalmente, definimos $\hat{\beta}_{reg}$ siendo

$$\hat{\beta}_{reg} = \mathbb{P}_n(b(L, \hat{\tau})),$$

donde \mathbb{P}_n denota el operador de esperanza empírica, dado por $\mathbb{P}_n(h(O)) = 1/n \sum_{i=1}^n h(O_i)$. En este caso la reducción elegida es suponer como cierto el modelo (1.1). Bajo esta reducción, bajo condiciones de regularidad típicas y asumiendo el supuesto MAR, el estimator $\hat{\beta}_{reg}$ es consistente para β_0 .

Otra propuesta de estimación para $\beta_0 = E(Y)$ se basa en el siguiente hecho: bajo la suposición MAR, tenemos que $P(R = 1|Y, L) = P(R = 1|L)$, luego, β_0 satisface

$$E\left(\frac{R}{P(R = 1|L)}(Y - \beta_0)\right) = 0. \quad (1.2)$$

Esto sugiere estimar a β_0 proponiendo un modelo paramétrico para $P(R = 1|L)$ dado por

$$P(R = 1|L) = \pi(L, \alpha_0). \quad (1.3)$$

Definimos entonces el estimador $\widehat{\beta}_{IPW}$ como la solución en β de la ecuación

$$\mathbb{P}_n \left(\frac{R}{\pi(L, \widehat{\alpha})} (Y - \beta) \right) = 0,$$

con $\widehat{\alpha}$ el estimador de máxima verosimilitud de α_0 bajo el modelo (1.3).

Ahora, la reducción elegida consiste en suponer como cierto el modelo (1.3). Bajo esta reducción y el supuesto MAR, el estimator $\widehat{\beta}_{IPW}$ es consistente para β_0 .

Por el momento presentamos dos estimadores de β_0 : $\widehat{\beta}_{reg}$ y $\widehat{\beta}_{IPW}$ cumpliendo

- ★ $\widehat{\beta}_{reg}$ es consistente para β_0 si el modelo (1.1) es verdadero.
- ★ $\widehat{\beta}_{IPW}$ es consistente para β_0 si el modelo (1.3) es verdadero.

En otras palabras: $\widehat{\beta}_{reg}$ no es consistente si (1.1) es falso y $\widehat{\beta}_{IPW}$ no es consistente si (1.3) es falso. En este contexto tanto τ_0 como α_0 son parámetros de ruido del problema y las reducciones posibles resultan ser los modelos dados por (1.3) y (1.1).

Sorprendentemente, Scharfstein, Rotnitzky y Robins (1999) presentaron un tercer procedimiento de estimación llamado doble protegido que confiere mayor protección que las dos propuestas presentadas al momento en el siguiente sentido: si el modelo dado en (1.1) es verdadero o si el modelo (1.3) es verdadero, este tercer estimador resulta consistente. Es decir, el nuevo estimador resulta consistente si se cumple el modelo dado en (1.1) o bien se cumple el modelo dado en (1.3), sin necesidad de especificar cual de los dos modelos se verifica. Dicho estimador, denotado en esta tesis por $\widehat{\beta}_{DP}$, se obtiene como la solución en β de la ecuación

$$\mathbb{P}_n \left(\frac{R}{\pi(L, \widehat{\alpha})} (Y - \beta) - \left(\frac{R}{\pi(L, \widehat{\alpha})} - 1 \right) (b(L, \widehat{\tau}) - \beta) \right) = 0, \quad (1.4)$$

donde $\widehat{\alpha}$ es el estimador de máxima verosimilitud de α_0 bajo el modelo (1.3) y $\widehat{\tau}$ es un estimador de τ_0 construido bajo el modelo (1.1). Desde el artículo de Scharfstein et. al. (1999) se han propuesto diferentes estimadores doble protegidos de la media poblacional bajo el modelo recién presentado así como también se propusieron estimadores de parámetros específicos bajo otros modelos estadísticos que se asumen frecuentemente en la inferencia causal y en el análisis de datos faltantes. La siguiente lista incluye algunas referencias en el área.

1. Robins J.(2000). En este artículo, Robins construye estimadores doble protegidos bajo un modelo semiparamétrico que involucra variables longitudinales con datos faltantes y bajo ciertos modelos causales. Además muestra que los estimadores encontrados pueden representarse como estimadores secuenciales de regresión.

2. Robins J, Rotnitzky A, van der Laan (2000). Los autores extienden los resultados presentados en Scharfstein et. al. (1999) presentando un procedimiento de estimación doble protegido en modelos donde la verosimilitud es de la forma $L(\kappa, \gamma) = L_1(\kappa)L_2(\gamma)$ con κ y γ variando en forma independiente, para parámetros μ que solo dependen de κ .
3. Murphy S, van der Laan M, Robins J. (2001). Motivados por el estudio de un ensayo clínico llamado Fast Track Prevention Program, en este artículo los autores construyen estimadores doble protegidos para la media de cierta variable respuesta en modelos marginales estructurales para tratamientos dinámicos. También repasan conceptos relacionados con la teoría causal basados en variables contrafactuales.
4. van der Laan M, Robins J. (2003). En este libro los autores presentan métodos de estimación doble protegidos en distintos escenarios, dando en detalle la teoría matemática detrás de cada metodología presentada.
5. Davidian M, Lunceford J. (2004). Bajo ciertas hipótesis, los autores hacen un repaso de distintos métodos de estimación para el efecto causal de un tratamiento sobre una cierta variable respuesta. Entre otros, revisan métodos de estimación basados en estratificación, en pesos y el método doble protegido. Además reportan un extenso estudio de simulación donde comparan el desempeño de los estimadores expuestos.
6. Bang H, Robins J. (2005). Los autores revisan algoritmos y métodos de estimación para construir estimadores doble protegidos para la media de una cierta variable respuesta en modelos no longitudinales con datos faltantes y en modelos causales inferenciales no longitudinales. Luego, extienden dichos métodos al caso longitudinal, es decir desarrollan procedimientos para estimar la media de una cierta variable respuesta en modelos longitudinales con datos faltantes y en modelos causales inferenciales longitudinales, específicamente en modelos marginales estructurales para la media. Finalmente aplican el método propuesto a datos provenientes de un ensayo clínico cardiovascular.
7. Carpenter J, Kenward, M, Vansteelandt, S. (2006). Este trabajo presenta una revisión de la metodología de estimación doble protegida en escenarios con datos faltantes y luego contrasta estos métodos con ciertas técnicas de estimación usadas también en el contexto de datos faltantes llamadas de múltiple imputación. Dicho contraste se hace desde un punto de vista teórico y práctico.
8. Tan, Z. (2006). Bajo ciertas hipótesis, Tan presenta estimadores doble protegidos en modelos causales usando variables instrumentales en el contexto de selección de tratamientos basado en covariables. En paralelo, en el mismo caso, también

presenta estimadores basados en métodos de ajuste de regresión y en métodos de pesos (IPW-method).

9. Kang J, Schafer J. (2007). En este trabajo, Kang y Schafer hacen foco en el problema de estimar la esperanza de una variable aleatoria en presencia de datos faltantes bajo el supuesto de ignorabilidad fuerte. En ese marco muestran que los estimadores doble protegidos pueden construirse de diferentes maneras. Además comparan el desempeño de este tipo de estimadores con otros que no son doble protegidos mediante un estudio de simulación.
10. Orellana L, Rotnitzky A, Robins J. (2010). En este trabajo, en el contexto de datos longitudinales, dada una familia de regímenes dinámicos $G = \{g_x : x \in T\}$, se consideran modelos (paramétricos y semiparamétricos) marginales estructurales para la media condicional de cierta variable contrafactual de interés bajo el régimen g_x . Los autores construyen estimadores doble protegidos para parámetros asociados a dichos modelos así como también para el índice de un régimen que es óptimo en cierto sentido.
11. Carpenter J, Kenward, M, Vansteelandt S. (2010). En este artículo, los autores revisan métodos de estimación ipw y métodos de estimación doble protegidos para la media de una variable respuesta en escenarios con datos incompletos. Luego dan lineamientos de como aplicar estos procedimientos en una gran variedad de casos, como por ejemplo en modelos de regresión con respuesta incompleta o covariables incompletas enfatizando en el empleo de softwares de uso común en la comunidad científica.
12. Tchetgen Tchetgen E, Robins J, Rotnitzky A. (2010). En este trabajo los autores estiman en forma doble protegida los parámetros de un modelo semiparamétrico para funciones odd ratio. Además los estimadores presentados tienen propiedades de eficiencia. Una de las ventajas de esta propuesta de estimación es que a diferencia de otros métodos de estimación en estos modelos, el método desarrollado por los autores es de fácil implementación.
13. Rotnitzky A, Lei Q, Sued M, Robins J. (2012). Los autores proponen construir estimadores doble protegidos para los parámetros de un modelo de regresión con respuesta faltante bajo el supuesto de MAR y para los parámetros de modelos marginales estructurales para la media. Los estimadores propuestos resultan más eficientes que otros estimadores doble protegidos existentes para los mismos parámetros. Además logran que dichos estimadores estén en el rango donde toma valores el parámetro que estiman.

En los últimos años, diversos autores brindaron procedimientos de estimación que extienden los métodos doble protegidos, proponiendo procedimientos que llamaremos

múltiple protegidos, pues dan como resultado estimadores consistentes si alguno de ciertos m submodelos asumidos es verdadero, con $m > 2$. A modo de ejemplo damos algunas referencias:

1. Vansteelandt S, Robins J, Rotnitzky A. (2008). En este artículo los autores proponen una nueva clase de modelos para hacer inferencia sobre la media de un vector de variables respuesta, con respuestas repetidas, cuando dicho vector no es completamente observado y el mecanismo de pérdida no es monótono. Los autores estiman en forma doble y múltiple protegida a los parámetros de los modelos introducidos.
2. Vansteelandt S, VanderWeele T, Tchetgen Tchetgen E, Robins J. (2008). Con el objetivo de decidir si hay un efecto de la interacción entre dos variables sobre una determinada variable respuesta, los autores consideran ciertos modelos para la esperanza condicional que postulan la interacción estadística en términos de un parámetro finito dimensional, pero que además involucran otros parámetros adicionales. En ese contexto, los autores consideran procedimientos de estimación múltiple protegidos.
3. Tchetgen Tchetgen (2009) En este trabajo, el autor, en el contexto de datos faltantes longitudinales monótonos, propone diferentes estimadores para la media de una cierta variable respuesta faltante en una submuestra. Entre las distintas propuestas, presenta un estimador cuádruple protegido para dicho funcional.

Dos de los principales objetivos de este trabajo procuran presentar una teoría que

- (1) explique la existencia de estimadores múltiple protegidos en las propuestas desarrolladas en algunos de los trabajos mencionados anteriormente,
- (2) desarrolle condiciones suficientes bajo las cuales es posible la construcción de ecuaciones de estimación que, bajo condiciones de regularidad, tienen soluciones que constituyen estimadores múltiple protegidos.

Con ese fin, para acercar al lector a esta problemática, en el Capítulo 2 de esta tesis presentamos dos problemas inferenciales describiendo en cada uno diferentes estimadores, incluyendo procedimientos múltiple protegidos. El primer ejemplo es una pequeña modificación del caso estudiado por Tchetgen Tchetgen (2009), mientras que en el segundo ejemplo, se procura estimar la componente paramétrica bajo un modelo de regresión parcialmente lineal, ante la presencia de respuestas faltantes. La propuesta de estimación cuádruple protegida bajo este escenario es uno de los aportes de este trabajo, y para su formulación nos hemos inspirado en el problema estudiado en Goetgeluk et al. (2009), quienes desarrollan procedimientos doble protegidos para estimar parámetros asociados a un modelo estructural anidado para efectos directos.

En el Capítulo 2, presentamos también los resultados de un estudio de Monte Carlo realizado bajo el modelo presentado en el ejemplo 2. El mismo permite comparar el desempeño de los estimadores considerados, observándose, tal como sugiere la teoría, que el procedimiento múltiple protegido propuesto es el único que resulta consistente en todos los escenarios contemplados.

Una vez presentados estos dos ejemplos, en el Capítulo 3 establecemos un marco general donde desarrollar la teoría, comenzando por definir el funcional de interés. En ese mismo capítulo introducimos la noción de función de estimación múltiple protegida, generalizando la definición de función de estimación doble protegida propuesta en Robins, Rotnitzky (2001). En el Capítulo 4, construimos funciones de estimación múltiple protegidas asumiendo la convexidad de ciertos modelos, condición que ha sido largamente explorada en el área desde el trabajo de Bickel (1982). La construcción sugerida permite reconstruir en los ejemplos las funciones de estimación que dan lugar al estimador múltiple protegido presentado en cada uno de ellos. Sin embargo, el éxito del procedimiento depende de la posibilidad de poder estimar también los parámetros de ruido en forma múltiple protegida. Este último punto ha sido poco explorado en la literatura, más allá del tratamiento que recibió en los diferentes casos particulares estudiados. El tercer logro que queremos destacar, presentado en el Capítulo 5, consiste en brindar un procedimiento de estimación múltiple protegido para los parámetros de ruido asociados a los submodelos de reducción. En el Capítulo 6 discutimos problemas abiertos que surgieron a partir de esta investigación.

La teoría desarrollada en este trabajo está íntimamente relacionada con algunos conceptos pertenecientes a la estadística semiparamétrica. Es por ello que decidimos incorporar en el Apéndice A las nociones básicas relativas a dicha disciplina.

Finalmente, con el objeto de hacer fluida la lectura de esta tesis, en el Apéndice B presentamos las demostraciones de todos los resultados enunciadas a lo largo de los diferentes capítulos de la misma. Algunas de ellas son «*elementary, tedious*» (Newey (1990)). Es por ello que decidimos intercalar algunas demostraciones de carácter heurístico, para poder acompañar el espíritu de la temática a lo largo de los distintos capítulos.

1.1. Notación

Incluimos en esta sección una lista con algunas convenciones adoptadas a lo largo de este manuscrito. Generalmente, aunque sea un poco redundante, la mayoría de ellas son también presentadas en el lugar temporal natural en el que aparecen durante el desarrollo del trabajo. De todas formas, esperamos que el lector pueda encontrar aquí rápidamente algunas definiciones, en caso de que sea necesario.

1. Si $U = (U_1, \dots, U_n)$, usaremos \bar{U}_k para denotar las primeras k coordenadas de U : $\bar{U}_k \equiv (U_1, \dots, U_k)$.
2. Utilizaremos $E_p(\cdot)$ para denotar esperanzas respecto de p .
3. Si $Z_i|\bar{Z}_{i-1}$ tiene densidad p_i , utilizaremos $E_{p_i}(h(Z_i)|\bar{Z}_{i-1})$.
4. Dada una muestra $(Z_i)_{i \geq 1}$, la media empírica de la función h está dada por $\mathbb{P}_n(h(Z)) \equiv \frac{1}{n} \sum_{i=1}^n h(Z_i)$.
5. i.i.d. denota una sucesión de variables o vectores aleatorios independientes idénticamente distribuidos.
6. Salvo que se aclare lo contrario pondremos h para denotar a (h_1, \dots, h_K) y $p = gh$ para $p = gh_1 h_2 \dots h_K$.
7. Si K es un natural, pondremos $[K]$ para denotar al conjunto $\{1, 2, \dots, K\}$.

8. Dados los conjuntos $\mathcal{B}_1, \dots, \mathcal{B}_j$, denotamos con $\otimes_{i=1}^j \mathcal{B}_i$ al producto cartesiano de $\mathcal{B}_1, \dots, \mathcal{B}_j$, es decir :

$$\otimes_{i=1}^j \mathcal{B}_i = \mathcal{B}_1 \times \dots \times \mathcal{B}_j$$

9. Un subconjunto en un subíndice denota una arreglo de elementos indexados por los elementos del subconjunto, es decir: si $\gamma_1, \dots, \gamma_K$ son K escalares y $c \subseteq [K]$, entonces $\gamma_c = \left(\gamma_j \right)_{j \in c}$. Más aún, si $c = [K]$ eliminamos los subíndices, es decir, pondremos $\gamma \equiv \gamma_{[K]}$.

10. \bar{A}_0 y \bar{a}_0 indican los vectores vacío.

11. Dado un conjunto A , para $a \subseteq A$ denotamos \bar{a} al complemento de a , es decir $\bar{a} = A - a$.

12. Para cada $B \equiv b(Z) \in L_2(p)$ y para cada Λ subespacio cerrado de $L_2(p)$, denotamos con $\Pi_p(B|\Lambda)$ a la $L_2(p)$ – proyección de B en Λ , i.e. el único elemento de Λ que satisface $E_p(\{B - \Pi_p(B|\Lambda)\} B^\dagger) = 0$ para todo $B^\dagger \in \Lambda$.

13. Dado el modelo

$$\mathcal{P} \equiv \left\{ p = gh_1 h_2 \dots h_K : g \in \mathcal{U}^{np}, h_k \in \mathcal{V}_k^{np}, p(z) \geq 0 \text{ para todo } z \text{ y } \int p d\mu = 1 \right\},$$

fijado un subconjunto $c \subseteq [K]$, funciones $g \in \mathcal{U}^{np}$ y $h_c \in \otimes_{j \in c} \mathcal{V}_j^{np}$, definimos el submodelo de \mathcal{P}

$$\mathcal{P}_{g, h_c} \equiv \{p^* = gh^* \in \mathcal{P} \text{ tales que } h_k^* = h_k \text{ para } k \in c\}.$$

14. Denotamos con $\Lambda_{\bar{c}}(p)$ al espacio tangente maximal del submodelo \mathcal{P}_{g, h_c} en el punto $p = gh \in \mathcal{P}_{g, h_c}$. Es decir, $\Lambda_{\bar{c}}(p)$ es la clausura (en $L_2(p)$) del subespacio generado por todos los scores asociados a modelos unidimensionales regulares paramétricos de \mathcal{P}_{g, h_c} que pasan por p .
15. Pondremos $\Lambda_{\bar{c}}(h_{\bar{c}})$ en lugar de $\Lambda_{\bar{c}}(p)$. Más aún, pondremos $\Lambda_j(h_j)$ en lugar de $\Lambda_{\{j\}}(h_j)$.
16. Dado un espacio vectorial V y $V_1 \dots V_r$ subespacios de V definimos

$$\oplus_{j=1}^r V_j \equiv \{w \in V : w = \sum_{j=1}^r v_j : v_j \in V_j\}.$$

Capítulo 2

Ejemplos-Estudio de Monte Carlo

2.1. Ejemplo 1: datos longitudinales.

En el contexto de un ensayo clínico, podemos pensar que L_j^f denota cierta variable que se quiere medir en el instante t_j , $j = 1, 2$, mientras que utilizaremos Y^f para representar a la variable respuesta que se intenta medir a tiempo t_3 , donde $t_1 < t_2 < t_3$. Supongamos que algunas coordenadas del vector (L_1^f, L_2^f, Y^f) son faltantes. Más precisamente, asumimos que se observan n copias independientes e idénticamente distribuidas (en adelante i.i.d) del vector $Z = (L_1, R_1, L_2, R_2, Y)$ donde R_i para $i = 1, 2$ son variables binarias tomando valores en $\{0, 1\}$, $L_1 = L_1^f, L_2 = R_1 L_2^f + (1 - R_1) *$, mientras que $Y = R_2 Y^f + (1 - R_2) *$. Es decir

L_1^f es observada e igual a L_1 en toda la muestra,

L_2^f es observada e igual a L_2 solo en la submuestra donde $R_1 = 1$,

Y^f es observada e igual a Y solo en la submuestra donde $R_2 = 1$.

Notemos que este escenario generaliza el ejemplo presentado en la introducción, pasando de dos a tres tiempos. Al igual que en ese ejemplo inicial, en este contexto, el objeto de interés es $E(Y^f)$. Bajo el supuesto de que los datos faltantes son perdidos al azar (MAR), es posible identificar $E(Y^f)$ (Rubin, 1976). El supuesto de MAR en este ejemplo consiste en asumir que

$$P(R_1 = 1 | L_1, Y^f) = P(R_1 = 1 | L_1) \text{ y} \quad (2.1)$$

$$P(R_2 = 1 | R_1 = 1, L_1, L_2, Y^f) = P(R_2 = 1 | R_1 = 1, L_1, L_2). \quad (2.2)$$

Hecho este supuesto, si definimos

$$\beta(p) \equiv E_p(E_p(E_p(Y|R_2 = 1, R_1 = 1L_1, L_2)|R_1 = 1, L_1)),$$

donde p es la densidad de Z , tenemos el siguiente resultado cuya demostración presentamos en el Apéndice B

Proposición 2.1.1. *Bajo (2.1) y (2.2) resulta que $\beta(p)$ es igual a la esperanza de la variable Y^f :*

$$E(Y^f) = E_p(E_p(E_p(Y|R_2 = 1, R_1 = 1L_1, L_2)|R_1 = 1, L_1)).$$

En vista del resultado anterior, podemos pensar que estamos ante un problema donde los datos observados son n copias independientes e idénticamente distribuidas de un vector $Z = (L_1, R_1, L_2, R_2, Y)$ con densidad desconocida p con respecto a una medida σ -finita μ , donde R_1 y R_2 son variables binarias tomando valores en $\{0, 1\}$ y estamos interesados en estimar el funcional

$$\beta(p) \equiv E_p(E_p(E_p(Y|R_2 = 1, R_1 = 1L_1, L_2)|R_1 = 1, L_1)). \quad (2.3)$$

Daremos ahora una posible factorización de la densidad p de forma tal que el funcional $\beta(\cdot)$ dependerá solo de uno de los factores involucrados en la misma. Para ello, consideremos $z = (l_1, r_1, l_2, r_2, y)$. Si definimos

$$\begin{aligned} g(z) &\equiv p_{L_1}(l_1)p_{L_2|R_1, L_1}(l_2|r_1, l_1)p_{Y|R_2, \bar{L}_2}(y|r_2, \bar{l}_2), \\ h_1(z) &\equiv p_{R_1|L_1}(r_1|l_1), \\ h_2(z) &\equiv p_{R_2|R_1, \bar{L}_2}(r_2|r_1, \bar{l}_2), \end{aligned} \quad (2.4)$$

resulta que podemos factorizar a p de la siguiente manera

$$p = gh_1h_2. \quad (2.5)$$

En base a la factorización (2.5), mirando cada una de las esperanzas involucradas en la definición del funcional $\beta(\cdot)$ dada en (2.3), observamos que $\beta(p)$ depende de p solo mediante g , con lo cual $\beta(\cdot)$ es en realidad un funcional que depende de g . Debido a este hecho, haciendo un pequeño abuso de notación, pondremos $\beta(g)$ en lugar de $\beta(p)$. Luego, es de esperar que modelando algunos funcionales que dependen exclusivamente de g arribemos a una posible estimación del parámetro de interés. Esa es la línea de trabajo de la primer propuesta de estimación desarrollada en la próxima sección.

2.1.1. Estimador de regresión.

Considerando la factorización de p dada en (2.5), tenemos que cada una de las esperanzas (y esperanzas condicionales) involucradas en la definición del funcional de interés presentada en (2.3), dependen solo de g . Definimos entonces

$$\nu_1(g) \equiv E_g(E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot) | R_1 = 1, L_1 = \cdot), \quad (2.6)$$

$$\nu_2(g) \equiv E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot). \quad (2.7)$$

Un primer estimador se obtiene combinando modelos paramétricos para dichos funcionales con medias empíricas. Explícitamente, vamos a asumir que

$$\nu_1(g) \in \mathcal{R}_{sub,1} = \{b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\} \text{ y } \nu_2(g) \in \mathcal{R}_{sub,2} = \{b_2(\cdot, \cdot; \tau_2) : \tau_2 \in \Upsilon_2\}, \quad (2.8)$$

para ciertas funciones $b_1(L_1; \cdot)$ y $b_2(\bar{L}_2; \cdot)$ conocidas, indexadas por parámetros finito dimensionales pertenecientes a conjuntos $\Upsilon_j \subset \mathbb{R}^{r_j}$, para $j = 1, 2$. Recordamos que como anunciamos en la sección (1.1), dado $U = (U_1, \dots, U_n)$, usaremos \bar{U}_k para denotar las primeras k coordenadas de U : $\bar{U}_k \equiv (U_1, \dots, U_k)$.

A partir de hacer este supuesto tenemos el siguiente procedimiento de estimación:

Paso 1- Calculamos el estimador $\hat{\tau}_{2,reg}$ regresando a Y en L_1 y L_2 , entre los individuos con $R_2 = 1$ y $R_1 = 1$, por ejemplo: $\hat{\tau}_{2,reg}$ resuelve la ecuación en τ_2 dada por

$$\mathbb{P}_n \left(R_1 R_2 \frac{\partial b_2(\bar{L}_2, \tau_2)}{\partial \tau_2} (Y - b_2(\bar{L}_2, \tau_2)) \right) = 0. \quad (2.9)$$

Paso 2- Calculamos $\hat{\tau}_{1,reg}$ regresando $b_2(\bar{L}_2; \hat{\tau}_{2,reg})$ en L_1 , entre los individuos con $R_1 = 1$, por ejemplo: $\hat{\tau}_{1,reg}$ resuelve la ecuación en τ_1 dada por

$$\mathbb{P}_n \left(R_1 \frac{\partial b_1(L_1, \tau_1)}{\partial \tau_1} (b_2(\bar{L}_2; \hat{\tau}_{2,reg}) - b_1(L_1, \tau_1)) \right) = 0. \quad (2.10)$$

Paso 3- Finalmente estimamos a $\beta(g)$ con $\hat{\beta}_{reg} \equiv \mathbb{P}_n(b_1(L_1; \hat{\tau}_{1,reg}))$.

Bajo condiciones de regularidad, el estimador $\hat{\beta}_{reg}$ es consistente para $\beta(g)$, siempre que los modelos propuestos en (2.8) sean correctos. Es decir, si la distribución $p = gh_1h_2$ que generó los datos pertenece a $\mathcal{G}_1 \cap \mathcal{G}_2$, siendo

$$\mathcal{G}_k = \{p \in \mathcal{M} : \nu_k(g) \in \mathcal{R}_{sub,k}\}. \quad (2.11)$$

Notemos que si $p = gh_1h_2$ pertenece a \mathcal{G}_k para $k = 1$ o $k = 2$, resulta que el funcional $\nu_k(g)$ pertenece al conjunto $\mathcal{R}_{sub,k}$. En este contexto llamaremos $\tau_k(g)$ al único elemento

de Υ_k para el cual se verifica que $\nu_k(g) = b_k(\cdot; \tau_k(g))$.

En el procedimiento de estimación que acabamos de describir modelamos funcionales relacionados exclusivamente con g . En la próxima sección vamos a presentar otra forma de estimar al funcional de interés basada en modelar h_1 y h_2 .

2.1.2. Estimador IPW

Para la próxima propuesta de estimación, necesitamos el siguiente resultado de Robins, Rotnitzky y Zhao (1995) cuya demostración de todos modos incluimos en el Apendice B.

Lema 2.1.1. *Dada una función $H(\bar{L}_2, Y)$, se verifica*

$$E_p \left(\frac{R_1 R_2 H(\bar{L}_2, Y)}{E_p(R_1|L_1) E_p(R_2|\bar{L}_2, R_1)} \right) = E_p \left(E_p \left(E_p \left(H(\bar{L}_2, Y) | R_1 = 1, R_2 = 1, \bar{L}_2 \right) | L_1, R_1 = 1 \right) \right).$$

En particular, tomando $H(\bar{L}_2, Y) = Y - \beta$, del Lema 2.1.1 tenemos que el funcional $\beta(g)$ se puede pensar como la solución en β de la ecuación dada por

$$E_p \left(\frac{R_2 R_1}{E_p(R_1|L_1) E_p(R_2|R_1, \bar{L}_2)} (Y - \beta) \right) = 0, \quad (2.12)$$

siendo $p = gh_1h_2$. Notemos que $E_p(R_1|L_1)$ y $E_p(R_2|R_1, \bar{L}_2)$ dependen exclusivamente de h_1 y h_2 , respectivamente. Por esta razón, en adelante, utilizaremos E_{h_i} para enfatizar que las esperanzas dependen de p sólo mediante h_i , tal como mencionamos en la lista de notación.

Es decir, definiendo

$$M(\beta, h_1, h_2) \equiv \frac{R_2 R_1}{E_{h_1}(R_1|L_1) E_{h_2}(R_2|R_1, \bar{L}_2)} (Y - \beta), \quad (2.13)$$

tenemos que $\beta(g)$ es el único número que verifica $E_p \{M(\beta(g), h_1, h_2)\} = 0$. Esta caracterización del funcional de interés sugiere una propuesta de estimación que consiste en asumir

$$h_1 \in \mathcal{V}_{sub,1} = \{h_{1,\alpha_1} = p_{R_1|L_1}(\cdot|\cdot; \alpha_1) : \alpha_1 \in \Xi_1\} \text{ y } h_2 \in \mathcal{V}_{sub,2} = \{h_{2,\alpha_2} = p_{R_2|R_1, \bar{L}_2}(\cdot|\cdot; \alpha_2) : \alpha_2 \in \Xi_2\}, \quad (2.14)$$

con $\Xi_j \subset R^{\ell_j}$, $j = 1, 2$, mientras que para cada α_1, α_2 , $p_{R_1|L_1}(\cdot|\cdot; \alpha_1)$ y $p_{R_2|R_1, \bar{L}_2}(\cdot|\cdot; \alpha_2)$ son densidades condicionales conocidas.

En adelante, utilizaremos la siguiente convención notacional:

$$E(R_1|L_1; \alpha_1) =: E_{h_1, \alpha_1}(R_1|L_1) \quad (2.15)$$

$$E(R_2|R_1, \bar{L}_2; \alpha_2) =: E_{h_2, \alpha_2}(R_2|R_1, \bar{L}_2). \quad (2.16)$$

Para poder definir al nuevo estimador bajo el presente contexto, consideremos $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2)$ el estimador de máxima verosimilitud de $\alpha = (\alpha_1, \alpha_2)$ bajo los modelos dados en (2.14). Sean $\hat{h}_1 = p_{R_1|L_1}(\cdot|\hat{\alpha}_1)$, $\hat{h}_2 = p_{R_2|\bar{L}_2, R_1}(\cdot|\hat{\alpha}_2)$. Definimos el estimador $\hat{\beta}_{IPW}$ siendo la solución de

$$\mathbb{P}_n \left(M \left(\beta, \hat{h}_1, \hat{h}_2 \right) \right) = 0.$$

Bajo condiciones de regularidad, $\hat{\beta}_{IPW}$ es consistente para $\beta(g)$ siempre que sean verdaderos los modelos dados en (2.14). Es decir $\hat{\beta}_{IPW}$ es consistente para $\beta(g)$ si $p = gh_1h_2 \in \mathcal{H}_1 \cap \mathcal{H}_2$ donde

$$\mathcal{H}_k = \{p \in \mathcal{M} : h_k \in \mathcal{V}_{sub,k}\}. \quad (2.17)$$

Al igual que en la sección anterior, si $p = gh_1h_2$ pertenece a \mathcal{H}_k para $k = 1$ o $k = 2$, resulta que h_k pertenece al conjunto $\mathcal{V}_{sub,k}$. En este contexto llamaremos $\alpha_k(h_k)$ al elemento de Ξ_k que determina a h_k : $\alpha_k(h_k)$ es tal que $h_k = h_{k \alpha_k(h_k)}$.

Para garantizar la consistencia de los estimadores $\hat{\beta}_{reg}$ y $\hat{\beta}_{IPW}$ se requiere que los respectivos modelos asumidos sean verdaderos. En otras palabras: $\hat{\beta}_{reg}$ no es consistente si (2.8) es falso y $\hat{\beta}_{IPW}$ no es consistente si (2.14) es falso. Sin embargo, en un contexto análogo al que estamos estudiando, van der Laan et. al. (2003) presentaron un estimador llamado doble protegido, cuya adaptación al presente problema confiere mayor protección que las dos propuestas presentadas al momento. El procedimiento doble protegido resulta consistente si las condiciones requeridas para la consistencia de $\hat{\beta}_{reg}$ ($p \in \mathcal{G}_1 \cap \mathcal{G}_2$) o las requeridas para la consistencia de $\hat{\beta}_{IPW}$ ($p \in \mathcal{H}_1 \cap \mathcal{H}_2$) se satisfacen, sin necesidad de especificar cuales. Es decir, el nuevo estimador resulta consistente si $p \in \{\mathcal{G}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{H}_1 \cap \mathcal{H}_2\}$.

2.1.3. Estimación doble protegida

El estimador doble protegido, denotado por $\hat{\beta}_{DP}$, resuelve la ecuación de estimación $\mathbb{P}_n \{U(\beta, \hat{\tau}_{1,reg}, \hat{\tau}_{2,reg}, \hat{\alpha}_1, \hat{\alpha}_2)\} = 0$, donde

$$\begin{aligned} U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) \equiv & \frac{R_2 R_1}{E(R_1|L_1; \alpha_1) E(R_2|R_1, \bar{L}_2; \alpha_2)} \{Y - b_2(\bar{L}_2; \tau_2)\} \\ & + \frac{R_1}{E(R_1|L_1; \alpha_1)} \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} + \{b_1(L_1; \tau_1) - \beta\}, \end{aligned} \quad (2.18)$$

mientras que $\widehat{\tau}_{1,reg}, \widehat{\tau}_{2,reg}, \widehat{\alpha}_1, \widehat{\alpha}_2$ son los estimadores construidos en las secciones anteriores. Es decir, consideremos la función $\widehat{\beta}(\tau_1, \tau_2, \alpha_1, \alpha_2)$, definida como la solución en β de la ecuación $\mathbb{P}_n \{U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2)\} = 0$:

$$\widehat{\beta}(\tau_1, \tau_2, \alpha_1, \alpha_2) \text{ verifica que } \mathbb{P}_n \left(U \left(\widehat{\beta}(\tau_1, \tau_2, \alpha_1, \alpha_2), \tau_1, \tau_2, \alpha_1, \alpha_2 \right) \right) = 0. \quad (2.19)$$

Tenemos entonces que

$$\widehat{\beta}_{DP} = \widehat{\beta}(\widehat{\tau}_{1,reg}, \widehat{\tau}_{2,reg}, \widehat{\alpha}_1, \widehat{\alpha}_2). \quad (2.20)$$

El estimador $\widehat{\beta}_{DP}$ tiene la siguiente propiedad: es consistente para $\beta(g)$ si los modelos propuestos en (2.8) o los modelos presentados (2.14) son correctos, sin que sea necesario saber cual de los dos escenarios es el verdadero. Luego, $\widehat{\beta}_{DP}$ es consistente para $\beta(g)$ bajo un conjunto más grande de distribuciones posibles para la ley que genera los datos que los estimadores $\widehat{\beta}_{reg}$ y $\widehat{\beta}_{IPW}$. La consistencia de $\widehat{\beta}_{DP}$ se desprende del siguiente resultado, cuya demostración omitimos, siendo que veremos en la próxima sección un resultado más general.

Lema 2.1.2. *Sea $p = gh_1h_2$. Consideremos los modelos \mathcal{G}_k y \mathcal{H}_k , definidos en (2.11) y (2.17), respectivamente. Tenemos que $\beta(g)$ es el único valor satisfaciendo*

$$E_p(U(\beta(g), \tau_1, \tau_2, \alpha_1, \alpha_2)) = 0, \quad (2.21)$$

en cualquiera de los siguientes casos:

1. $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, $\tau_1 = \tau_1(g)$ y $\tau_2 = \tau_2(g)$, cualesquiera sean los valores para α_1 y α_2 .
2. $p \in \mathcal{H}_1 \cap \mathcal{H}_2$, $\alpha_1 = \alpha_1(h_1)$ y $\alpha_2 = \alpha_2(h_2)$, cualesquiera sean los valores para τ_1 y τ_2 .

Bajo condiciones de regularidad, $\widehat{\beta}_{DP}$ definido en (2.20), converge a la única solución en β de la ecuación

$$E_p(U(\beta, \tau_1^\ell, \tau_2^\ell, \alpha_1^\ell, \alpha_2^\ell)) = 0, \quad (2.22)$$

asumiendo que $\widehat{\tau}_{i,reg}$ converge a τ_i^ℓ y que $\widehat{\alpha}_i$ converge a α_i^ℓ , para $i = 1, 2$. Cuando los modelos propuestos en (2.8) son correctos, tenemos $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, y resulta ser que $\tau_i^\ell = \tau_i(g)$. Luego, apelando al caso tratado en el item 1 del Lema 2.1.2, deducimos que $\widehat{\beta}_{DP}$ converge a $\beta(g)$. Por otra parte, cuando los modelos dados en (2.17) son verdaderos, $p \in \mathcal{H}_1 \cap \mathcal{H}_2$ y tenemos que $\alpha_i^\ell = \alpha_i(h_i)$, para $i = 1, 2$. Haciendo uso del item 2 del Lema 2.1.2, concluimos que $\widehat{\beta}_{DP}$ converge a $\beta(g)$.

En suma, $\widehat{\beta}_{DP}$ resulta ser consistente si $p = gh_1h_2 \in \{\mathcal{G}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{H}_1 \cap \mathcal{H}_2\}$, para \mathcal{G}_k y \mathcal{H}_k definidos en (2.11) y (2.17), respectivamente. En realidad se puede estudiar la

doble protección de un estimador en términos mucho más generales, pero la idea de este capítulo es intentar familiarizar al lector mediante ejemplos con el concepto de doble y múltiple protección (que presentaremos en la próxima sección).

A modo de repaso, por el momento presentamos 3 estimadores:

- ★ $\widehat{\beta}_{reg}$: consistente bajo (2.8): $p \in \mathcal{G}_1 \cap \mathcal{G}_2$.
- ★ $\widehat{\beta}_{IPW}$: consistente bajo (2.14): $p \in \mathcal{H}_1 \cap \mathcal{H}_2$.
- ★ $\widehat{\beta}_{DP}$: consistente bajo (2.8) o (2.14): $p \in \{\mathcal{G}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{H}_1 \cap \mathcal{H}_2\}$.

En la próxima sección introducimos un cuarto estimador llamado $\widehat{\beta}_{MP}$ el cual, resulta consistente en más escenarios que los contemplados por estos tres estimadores.

2.1.4. Propuesta cuádruple protegida.

En el contexto de datos faltantes longitudinales que presentamos al principio de esta sección, pero asumiendo que los mismos tienen una estructura de pérdida monótona ($R_1 = 0$ implica que $R_2 = 0$), Tchetgen Tchetgen (2009) propone diferentes estimadores para la media de la variable Y^f .

Adaptando el estimador múltiple protegido propuesto por Tchatgen Tchetgen (2009) a nuestro contexto, se puede ver que haciendo una pequeña modificación en $\widehat{\beta}_{DP}$ puede obtenerse un cuarto estimador que confiere aún mas protección que la alcanzada por el estimador doble protegido presentado en la sección anterior. Denotemos por $\widehat{\beta}_{MP}$ a dicho estimador. Este se obtiene como $\widehat{\beta}_{DP}$, reemplazando $\widehat{\tau}_{1,reg}$ por un estimador $\widehat{\tau}_{1,MP}$ que es solución de la ecuación de estimación $\mathbb{P}_n(T(\widehat{\alpha}_2, \tau_1, \widehat{\tau}_{2,reg})) = 0$, siendo $\widehat{\alpha}_2$ y $\widehat{\tau}_{2,reg}$ los estimadores presentados en las secciones anteriores, mientras que

$$T(\alpha_2, \tau_1, \tau_2) \equiv R_1 d(L_1, \tau_1, \tau_2, \alpha_2) \left(\frac{R_2 \{Y - b_2(\bar{L}_2; \tau_2)\}}{E(R_2 | R_1, \bar{L}_2; \alpha_2)} + \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} \right), \quad (2.23)$$

con $d(L_1, \tau_1, \tau_2, \alpha_2)$ una función vectorial arbitraria de la misma dimensión que τ_1 . En otras palabras, $\widehat{\beta}_{MP} = \widehat{\beta}(\widehat{\tau}_{1,MP}, \widehat{\tau}_{2,reg}, \widehat{\alpha}_1, \widehat{\alpha}_2)$ donde $\widehat{\beta}(\tau_1, \tau_2, \alpha_1, \alpha_2)$ es la función definida en (2.19).

Siguiendo una vez más a Tchetgen Tchetgen (2009), se puede ver que cuando $p \in \mathcal{G}_1$, $\widehat{\tau}_{1,MP}$ es un estimador doble protegido para $\tau_1(g)$ el cual resulta consistente si además p pertenece a $\mathcal{G}_2 \cup \mathcal{H}_2$. Es decir, si $p \in \{\mathcal{G}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{G}_1 \cap \mathcal{H}_2\}$, $\widehat{\tau}_{1,MP}$ converge a $\tau_1(g)$. Este hecho, combinado con una generalización del Lema 2.1.2 que presentaremos a continuación, sugiere que $\widehat{\beta}_{MP}$ es consistente para $\beta(g)$ si se verifica alguna de las siguientes condiciones

- (i) los modelos dados en (2.8) para $\nu_1(g)$ y $\nu_2(g)$ son verdaderos,
- (ii) los modelos dados en (2.14) para h_1 y h_2 son verdaderos,
- (iii) los modelos asumidos para h_1 y $\nu_2(g)$ son verdaderos,
- (iv) los modelos asumidos para h_2 y $\nu_1(g)$ son verdaderos.

Dicho de otra manera: $\widehat{\beta}_{MP}$ converge a $\beta(g)$ si

$$p = gh_1h_2 \in \{\mathcal{H}_1 \cap \mathcal{H}_2\} \cup \{\mathcal{G}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{H}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{G}_1 \cap \mathcal{H}_2\},$$

donde \mathcal{G}_k y \mathcal{H}_k están definidos en (2.11) y (2.17), respectivamente. Para explicar este hecho, apelaremos al siguiente resultado que es una generalización del Lema 2.1.2, y cuya demostración presentamos en el Apéndice B.

Lema 2.1.3. *Sea $p = gh_1h_2$. Consideremos los modelos \mathcal{G}_k y \mathcal{H}_k , definidos en (2.11) y (2.17), respectivamente. Tenemos que $\beta(g)$ es el único valor que verifica:*

$$E_p(U(\beta(g), \tau_1, \tau_2, \alpha_1, \alpha_2)) = 0, \quad (2.24)$$

en cualquiera de los siguientes casos:

1. $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, $\tau_1 = \tau_1(g)$, $\tau_2 = \tau_2(g)$, $\alpha_1 \in \Xi_1$ y $\alpha_2 \in \Xi_2$. Es decir, si $p \in \mathcal{G}_1 \cap \mathcal{G}_2$ la ecuación (2.24) se verifica tomando $\tau_1 = \tau_1(g)$, $\tau_2 = \tau_2(g)$, para todo $\alpha_1 \in \Xi_1$, $\alpha_2 \in \Xi_2$.
2. $p \in \mathcal{H}_1 \cap \mathcal{H}_2$, $\alpha_1 = \alpha_1(h_1)$, $\alpha_2 = \alpha_2(h_2)$, $\tau_1 \in \Upsilon_1$ y $\tau_2 \in \Upsilon_2$. Es decir, si $p \in \mathcal{H}_1 \cap \mathcal{H}_2$ la ecuación (2.24) se verifica tomando $\alpha_1 = \alpha_1(h_1)$, $\alpha_2 = \alpha_2(h_2)$, para todo $\tau_1 \in \Upsilon_1$, $\tau_2 \in \Upsilon_2$.
3. $p \in \mathcal{H}_1 \cap \mathcal{G}_2$, $\alpha_1 = \alpha_1(h_1)$, $\tau_2 = \tau_2(g)$, $\alpha_2 \in \Xi_2$, y $\tau_1 \in \Upsilon_1$. Es decir, si $p \in \mathcal{H}_1 \cap \mathcal{G}_2$, la ecuación (2.24) se verifica tomando $\alpha_1 = \alpha_1(h_1)$, $\tau_2 = \tau_2(g)$, para todo $\alpha_2 \in \Xi_2$, y $\tau_1 \in \Upsilon_1$.
4. $p \in \mathcal{G}_1 \cap \mathcal{H}_2$, $\tau_1 = \tau_1(g)$, $\alpha_2 = \alpha_2(h_2)$, $\alpha_1 \in \Xi_1$, $\tau_2 \in \Upsilon_2$. Es decir, si $p \in \mathcal{G}_1 \cap \mathcal{H}_2$, la ecuación (2.24) se verifica tomando $\tau_1 = \tau_1(g)$, $\alpha_2 = \alpha_2(h_2)$ para todo $\alpha_1 \in \Xi_1$, $\tau_2 \in \Upsilon_2$.

Claramente los dos primeros items de este lema dicen lo mismo que el Lema 2.1.2, por eso es una generalización del mismo. Finalizaremos esta sección justificando la múltiple protección de $\widehat{\beta}_{MP}$ anunciada. El razonamiento es análogo al que realizamos

a la hora de estudiar la doble protección de $\widehat{\beta}_{DP}$. Como $\widehat{\beta}_{MP} = \widehat{\beta}(\widehat{\tau}_{1,MP}, \widehat{\tau}_{2,reg}, \widehat{\alpha}_1, \widehat{\alpha}_2)$, nuestro estimador converge a la solución en β de la ecuación

$$E_p (U (\beta, \tau_1^\ell, \tau_2^\ell, \alpha_1^\ell, \alpha_2^\ell)) = 0, \quad (2.25)$$

asumiendo que $\widehat{\tau}_{1,MP}$ converge a τ_1^ℓ , que $\widehat{\tau}_{2,reg}$ converge a τ_2^ℓ y que $\widehat{\alpha}_i$ converge a α_i^ℓ , para $i = 1, 2$. Cuando los modelos propuestos en (2.8) para $\nu_1(g)$ y $\nu_2(g)$ son correctos, tenemos que $p \in \mathcal{G}_1 \cap \mathcal{G}_2$ luego $\tau_2^\ell = \tau_2(g)$ y, debido a la doble protección de $\widehat{\tau}_{1,MP}$, también se verifica que $\tau_1^\ell = \tau_1(g)$. Apelando al caso tratado en el ítem 1 del Lema 2.1.3, deducimos que $\widehat{\beta}_{MP}$ converge a $\beta(g)$. Similarmente, cuando los modelos dados en (2.17) son verdaderos, $p \in \mathcal{H}_1 \cap \mathcal{H}_2$ y tenemos que $\alpha_i^\ell = \alpha_i(h_i)$, para $i = 1, 2$. Haciendo uso del ítem 2 del Lema 2.1.3, concluimos que $\widehat{\beta}_{MP}$ converge a $\beta(g)$. Si ahora asumimos que los modelos postulados para h_1 y $\nu_2(g)$ son correctos, tenemos que $p \in \mathcal{H}_1 \cap \mathcal{G}_2$, luego $\alpha_1^\ell = \alpha_1(h_1)$ y $\tau_2^\ell = \tau_2(g)$ y por el ítem 3 del Lema 2.1.3 obtenemos la consistencia de $\widehat{\beta}_{MP}$ a $\beta(g)$. Finalmente si los modelos postulados para h_2 y $\nu_1(g)$ son correctos resulta que $p \in \mathcal{G}_1 \cap \mathcal{H}_2$, luego $\alpha_2 = \alpha_2(h_2)$ y además por la doble protección de $\widehat{\tau}_{1,MP}$ tenemos que $\tau_1^\ell = \tau_1(g)$. Luego la consistencia de $\widehat{\beta}_{MP}$ a $\beta(g)$ se sigue del ítem 4 en el Lema 2.1.3.

2.2. Ejemplo 2: modelo de regresión parcialmente lineal con datos faltantes.

2.2.1. Modelo de regresión parcialmente lineal.

Consideremos un estudio en el que algunos individuos son asignados a cierto tratamiento mientras que otros conforman el grupo control. Luego, en todos los individuos se registra el valor de una variable respuesta de interés. Podemos pensar que los datos observados son n copias i.i.d del vector (W, X, Y^f) , donde W es un vector de variables pre-tratamiento, X es una variable binaria indicando si al individuo se le asigno el tratamiento, mientras que Y^f denota la variable respuesta escalar de interés. Si p^f es la distribución de (W, X, Y^f) , asumimos el modelo de regresión parcialmente lineal, es decir, vamos a suponer que

$$E_{p^f} (Y^f | X, W) = \gamma (p^f) X + t_{p^f} (W), \quad (2.26)$$

donde $\gamma(\cdot)$ es un funcional con imagen en \mathbb{R} y $t_{p^f}(\cdot)$ es una función desconocida. En las aplicaciones de este modelo a la inferencia causal basada en estudios observacionales, X es un tratamiento o intervención en el momento t_0 sobre el que se desea estudiar su efecto causal en el momento t_1 en una variable de respuesta Y^f (siendo $t_1 > t_0$ un tiempo especificado de antemano). El vector W contiene las variables confusoras medidas

en el momento t_0^- (es decir justo antes de asignar X). Una vía conceptual para cuantificar efectos causales asume la existencia de variables contrafactuales $\{Y_{(x)}^f : x \in \mathcal{X}\}$ donde \mathcal{X} es el conjunto de posibles valores de X . De este conjunto de variables contrafactuales, solo es posible observar una de ellas, la que corresponde al tratamiento que el paciente recibe en la realidad, de modo que Y^f (la variable factual o observada) es igual a $Y_{(X)}^f$. En el caso más elemental, que asumiremos en nuestra descripción de la aplicación del modelo de regresión lineal parcial a la inferencia causal, el tratamiento de interés es una intervención puntual en el momento t_0 y X es la variable indicadora de recibir intervención ($X = 1$ si el individuo recibe la intervención y $X = 0$ de otro modo). En este contexto, $Y^f = XY_{(1)}^f + (1 - X)Y_{(0)}^f$, la diferencia de medias condicionales $E(Y_{(1)}^f|W) - E(Y_{(0)}^f|W)$ cuantifica el efecto causal de la intervención en la subpoblación con variables confusoras W . El modelo de regresión parcialmente lineal en el presente contexto surge a partir de las siguientes dos suposiciones:

- (i) $E(Y_{(1)}^f|W) - E(Y_{(0)}^f|W) = \gamma$ no depende de W
 - (ii) $P(X = x|Y_{(x)}^f, W) = P(X = x|W)$, es decir $Y_{(x)}^f$ y X son independientes dado W .
- Bajo estas suposiciones,

$$\begin{aligned} \gamma &= E(Y_{(1)}^f|W) - E(Y_{(0)}^f|W) \\ &= E(Y_{(1)}^f|X = 1, W) - E(Y_{(0)}^f|X = 0, W) \\ &= E_{p^f}(Y^f|X = 1, W) - E_{p^f}(Y^f|X = 0, W) \end{aligned}$$

De donde,

$$E_{p^f}(Y^f|X = 1, W) = \gamma X + t_{p^f}(W) \quad (2.27)$$

siendo $t_{p^f}(W) = E(Y^f|X = 0, W)$ irrestricta bajo el modelo. Bajo la suposición (ii), $t_{p^f}(W) = E(Y_{(0)}^f|X = 0, W)$ de modo que en términos de las variables contrafactuales, el modelo asume que el efecto causal de la intervención es el mismo para todos los niveles de W (e igual a γ) pero no restringe la forma en la que la media de la variable de respuesta en ausencia de intervención ($Y_{(0)}^f$) varía con W . Generalizaciones de este modelo que contemplan la posibilidad de que W modifique el efecto de X son posibles, reemplazando a la constante γ por alguna función $q(W; \gamma)$, por ejemplo $q(W; \gamma) = \gamma_0 + \gamma_1^T W$, resultando en el modelo $E(Y^f|X, W) = \gamma_0 X + \gamma_1^T W X + t_{p^f}(W)$ para las variables fácticas. Si bien en aplicaciones generalmente se analizan datos asumiendo que $q(W; \gamma)$ es alguna función no constante de W , en nuestra presentación nos limitaremos a discutir el modelo donde $q(W; \gamma) = \gamma$ debido a que el mismo concentra el nudo conceptual de la propuesta de estimación múltiple protegida, eje de esta tesis. De este modo evitaremos complicar la notación y desviar la atención del punto central de nuestra investigación. Para facilitar la discusión y ayudar al lector a conceptualizar las dificultades específicas

que surgen en la aplicación de este modelo en la práctica damos a continuación un ejemplo de un estudio hipotético en el que el modelo de regresión lineal parcial tendría aplicación. Supongamos contar con información recogida en pacientes de una clínica para obesos. A algunos de estos pacientes se les ofrece la posibilidad de ser sometidos a un bypass gástrico. Algunos de aquellos que reciben la oferta, la aceptan. Los pacientes, hayan o no sido sometidos a la intervención, retornan con frecuencia a la clínica para su control. Supongamos que nos interesa utilizar los datos recogidos en los pacientes de la clínica a lo largo de los dos años desde la oferta de intervención para estudiar el efecto de la intervención en el peso del paciente al cabo de dos años. En este contexto, la población sobre la cual queremos estudiar el efecto causal es el conjunto de individuos obesos de una población dada que calificarían para recibir la oferta de intervención en caso de haber asistido a la clínica. La muestra es el conjunto de pacientes que efectivamente asiste a la clínica y recibe la oferta del bypass gástrico. El momento t_0 es el instante en el cual el paciente recibe la oferta de intervención. El momento t_1 es $t_0 + 2$ años. El vector W contiene todas las variables recogidas en el paciente hasta el momento de la oferta de intervención, e incluyen, además de variables demográficas y socioeconómicas como edad, sexo, estado civil, salario, etc, variables que cuantifican el estado de salud del paciente, como por ejemplo presión arterial y nivel de colesterol. La variable X es la indicadora de haber sido sometido al bypass gástrico. La variable Y^f es el peso del paciente (haya o no sido sometido a la intervención) en el momento t_1 (es decir al cabo de dos años de haber recibido la oferta). La suposición (ii) de independencia condicional sería válida en un estudio experimental (ensayo clínico) en el que en cada subpoblación con un valor de W distinto, la aceptación de la oferta de someterse al bypass gástrico hubiese sido aleatorizada con una probabilidad de aceptación que tal vez pudiera depender del valor de W . En la práctica este estudio experimental no se puede realizar. Sin embargo si W contiene todas las variables predictoras de $(Y_{(0)}^f, Y_{(1)}^f)$ que están correlacionadas con la decisión de someterse a la intervención, podemos conceptualizar a los datos observacionales (los datos recogidos en la clínica) como si fueran originados por un estudio experimental *natural* en el que la aleatorización surge a partir de la variabilidad en las preferencias y predisposición psicológica de los pacientes sobre someterse a cirugías, siendo estas variables incorrelacionadas con $(Y_{(0)}^f, Y_{(1)}^f)$. Así vemos que para que la condición (ii) sea razonablemente cierta W debiera ser un vector de dimensión alta, ya que deberá aproximadamente contener todas las variables que simultáneamente predicen, antes de iniciarse el tratamiento, el peso del paciente al cabo de dos años y la voluntad de someterse a la cirugía. Una vez asumidas (i) y (ii), el problema de inferencia se remite al problema clásico de estimación en el modelo de regresión parcial basado en una muestra de las variables (W, X, Y^f) donde W es un vector de dimensión alta. Sin embargo, el problema de inferencia que queremos discutir en el presente ejemplo tiene la complicación adicional de que la variable de respuesta Y^f no es observada en toda la muestra. En el contexto de nuestro ejemplo, supongamos que Y^f no es medida en

todos los pacientes porque algunos de ellos no retornan a la clínica al cabo de dos años para su control. Denotemos con $R = 1$ si Y^f es observada y $R = 0$ si Y^f es faltante. Luego, la variable observada es $Y = RY^f + (1 - R)*$ y nuestro problema es estimar a γ a partir de una muestra aleatoria del vector (Y, X, W, R) . Lamentablemente, sin suposiciones adicionales, este problema no tiene solución (es decir, no existe un estimador consistente de γ) ya que γ no está identificada, es decir, no es un funcional de la distribución de (Y, X, W, R) . Notemos que restringir el análisis estadístico a la muestra de pacientes con Y^f observado llevaría a la estimación de la esperanza condicional de la variable Y entre los pacientes que retornan a la clínica al cabo de dos años, es decir a $E(Y|R = 1, W, X) = E(Y^f|R = 1, X, W)$. Esta esperanza es igual a $E(Y^f|X, W)$ solo bajo la suposición adicional de que la subpoblación de individuos con un valor de (X, W) dado que retornan a la clínica al cabo de dos años tiene, en promedio, el mismo peso que la de la subpoblación de aquellos que no retornan a la clínica. Es razonable sospechar que, a pesar de condicionar en (X, W) , esta suposición no se cumplirá en la práctica porque las razones por las cuales los pacientes no retornan a la clínica vinculadas al peso no pueden ser explicadas por las variables (X, W) medidas dos años antes. Supongamos, sin embargo, que se cuenta con información auxiliar sobre el paciente recogida a lo largo de los dos años posteriores a t_0 y codificada en un vector de variables V que estamos dispuestos a asumir que, razonablemente, contiene todas las variables predictoras de Y^f que se correlacionan con R . Por ejemplo, V incluye el peso del paciente medido en una visita realizada al cabo de un año, las respuestas a un cuestionario sobre dieta completado a las 18 meses, etc. Formalmente, suponemos que

$$(iii) P(R = 1|Y^f, X, W, V) = P(R = 1|X, W, V).$$

Algunos autores (por ejemplo, Robins, Rotnitzky and Zhao, 1995) llaman a esta suposición la suposición de datos faltantes aleatoriamente (Missing at random). Esta denominación acuerda con la formulación general de la condición MAR dada por Rubin (1976) para contextos generales, no necesariamente relacionados a problemas de estimación en modelos de regresión. Queremos notar, sin embargo, que la acepción de esta denominación no es universal ya que algunos autores en el presente contexto denominarían MAR a la condición de que $P(R = 1|Y^f, X, W) = P(R = 1|X, W)$ (lo cual, como lo hemos notado anteriormente, llevaría a abordar el análisis basándose en los datos observados). Veremos más adelante que bajo (iii) (y suponiendo adicionalmente (i) y (ii)) $E(Y^f|X, W)$ esta identificada, es decir es un funcional de la distribución de (Y, X, W, V, R) , y por lo tanto γ está identificada. Finalmente, queremos hacer mención de un punto importante que puede confundir al lector no familiarizado con la inferencia causal. A pesar de haber obtenido mediciones del vector V , el objetivo de la inferencia continúa siendo la estimación de γ en el modelo para la esperanza condicional de Y^f dados (X, W) y no dados (X, V, W) porque V es una variable medida después de recibir la intervención (variable post-treatment). Es sabido que analizar modelos para

la esperanza condicional de Y^f dado X, V, W es incorrecto ya que a la dependencia de esta esperanza condicional en X no puede ascribirse una interpretación causal. Concretamente, cualquiera de las siguientes dos situaciones es posible:

a) X e Y^f son condicionalmente independientes dado W pero X e Y^f sean condicionalmente dependientes dado (W, V) .

b) X e Y^f son condicionalmente dependientes dado W pero X e Y^f sean condicionalmente independientes datao (W, V) .

El caso (a) conocido como la falacia de Berkson (1946) implica que aún si el modelo $E(Y^f|X, W, V) = \delta X + t(X, W)$ es válido, al parámetro δ no puede ascribirse una interpretación causal. Esto ocurre porque en el caso (a) $\delta \neq 0$ el tratamiento X no tiene efecto causal sobre Y^f ya que $Y_{(1)}^f$ e $Y_{(2)}^f$ tienen la misma distribución dado W .

El caso (b) ocurre cuando todo el efecto causal de X en Y^f está mediado por V (Pearl, 1995). En este caso, en el modelo $E(Y^f|X, W, V) = \delta X + t(X, W)$ el valor verdadero de δ es 0 pero sin embargo existe efecto causal de X en Y^f ya que $Y_{(1)}^f$ e $Y_{(2)}^f$ tienen distintas distribuciones dado W .

El modelo (2.26), cuando la variable respuesta es observada en todos los individuos de la muestra, fue estudiado por muchos autores, quienes presentaron distintas propuestas de estimación para $\gamma(g)$. Por ejemplo, Engle et al. (1986) propusieron estimar a $t_g(-)$ y $\gamma(g)$ en forma simultánea mediante el método de mínimos cuadrados penalizados. Speckman (1988) presenta dos métodos de estimación para $\gamma(g)$ y t_g , basados en utilizar mínimos cuadrados penalizados para estimar a t_g en el modelo simplificado que se obtiene si se reemplaza en (2.26) $\gamma(g)$ por cero. Por otro lado Robinson (1988) propone insertar de manera apropiada estimadores no paramétricos relacionados con la componente t_g en el modelo (2.26) y luego, estimar $\gamma(g)$ en un modelo de regresión paramétrico auxiliar. Específicamente, bajo (2.26) se puede ver que

$$t_{p^f}(W) = E_{p^f}(Y^f|W) - \gamma(p^f) E_{p^f}(X|W) \quad (2.28)$$

y por consiguiente

$$Y^f - E_{p^f}(Y^f|W) = \gamma(p^f)(X - E_{p^f}(X|W)) + U, \quad (2.29)$$

con U tal que $E_{p^f}(U|X, W) = 0$. Robinson propone estimar a $\phi_0(w) \equiv E_{p^f}(Y^f|W = w)$ y a $\phi_1(w) \equiv E_{p^f}(X|W = w)$ no paramétricamente mediante $\hat{\phi}_0(w)$ y $\hat{\phi}_1(w)$ (por ejemplo usando los estimadores por núcleos de Nadaraya-Watson) para luego estimar a $\gamma(g)$ en el modelo de regresión lineal dado por

$$Y^f - \hat{\phi}_0(W) = \gamma(p^f)(X - \hat{\phi}_1(W)) + U, \quad (2.30)$$

con el estimador de mínimos cuadrados. Bianco y Boente (2003) proponen procedimientos para estimar a $\gamma(g)$ y a t_g utilizando estimadores no paramétricos robustos para $\phi_0(w) \equiv E_{p^f}(Y^f|W=w)$ y $\phi_1(w) \equiv E_{p^f}(X|W=w)$, así como también estimadores robustos para $\gamma(p^f)$ en el modelo de regresión dado en (2.30).

Al igual que en el ejemplo 1, damos una posible factorización de p^f de manera que el funcional $\gamma(\cdot)$ depende solo de uno de los factores involucrados en la misma. Si llamamos $z = (w, x, y^f)$ y definimos

$$\begin{aligned} g(z) &\equiv p_{Y^f|W,X}(y^f|w, x) p_W(w), \\ h_1(z) &\equiv p_{X|W}(x|w), \end{aligned}$$

resulta que

$$p^f = gh_1, \quad (2.31)$$

y luego, por definición, $\gamma(\cdot)$ y $t_{p^f}(\cdot)$ dependen de p^f solo mediante g . Por lo tanto pondremos $\gamma(g)$ en lugar de $\gamma(p^f)$ y $t_g(-)$ en lugar de $t_{p^f}(-)$.

Cuando W es un vector en un espacio de alta dimensión los procedimientos de estimación descritos basados en estimadores no paramétricos para funciones de regresión en el vector W padecen la maldición de la dimensión. Una posible estrategia para sortear esta dificultad consiste en reducir la complejidad del problema, postulando un modelo para t_g . Se puede considerar que t_g es aditiva en las componentes de W , o bien, como consideraremos en este trabajo, que t_g pertenece a una clase paramétrica.

Las técnicas que usaremos para encontrar en este escenario estimadores doble protegidos, siguen la filosofía de trabajo de van der Laan M, Robins J. (2003), desarrollada en el Capítulo 1 sección 1.5.

Para presentar una primer propuesta de estimación, definimos $\nu_1(g) = t_g(-)$ y asumimos un modelo paramétrico para el funcional $\nu_1(g)$:

$$\nu_1(g) \in \mathcal{R}_{sub,1} \equiv \{b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\}, \quad (2.32)$$

donde $b_1(W; \cdot)$ es una función suave conocida salvo $\tau_1 \in \Upsilon_1$. Combinando el modelo de regresión original dado en (2.26) con este nuevo modelo, tenemos un modelo de regresión paramétrico para la esperanza de Y^f condicional a X y W , dado por

$$E_{p^f}(Y^f|X, W) = \gamma(g)X + b_1(W; \tau_1(g)).$$

Los parámetros de este nuevo modelo pueden ser consistentemente estimados por el método de mínimos cuadrados. Es decir, si consideramos $(\hat{\gamma}_{reg}, \hat{\tau}_{1,LS})$ definidos como una solución en (γ, τ_1) del sistema

$$\mathbb{P}_n \left(\begin{pmatrix} X \\ \frac{\partial b_1}{\partial \tau_1}(W, \tau_1) \end{pmatrix} (Y^f - \gamma X - b_1(W, \tau_1)) \right) = 0, \quad (2.33)$$

entonces, bajo condiciones de regularidad, $\hat{\gamma}_{reg}$ resulta consistente para $\gamma(g)$ bajo (2.32) y (2.26). Dicho de otra forma, resulta consistente si $p^f = gh_1$ pertenece a \mathcal{G}_1 , siendo

$$\mathcal{G}_1 = \{p \in \mathcal{M} : \nu_1(g) \in \mathcal{R}_{sub,1}\}. \quad (2.34)$$

Una segunda estrategia de estimación consiste en asumir un modelo paramétrico para $p_{X|W}$. Es decir, asumir que

$$h_1 \in \mathcal{V}_{sub,1} \equiv \{p_{X|W}(\cdot|\cdot; \alpha_1) : \alpha_1 \in \Xi_1\}, \quad (2.35)$$

donde para cada α_1 , $p_{X|W}(\cdot|\cdot; \alpha_1)$ es una función de probabilidad condicional de X dado W y $\Xi_1 \subseteq \mathbb{R}^{\ell_1}$. Al igual que en el ejemplo anterior, en adelante, utilizaremos la siguiente convención notacional:

$$E(X|W; \alpha_1) =: E_{h_1, \alpha_1}(X|W). \quad (2.36)$$

Un estimador consistente para $\gamma(g)$ llamado $\hat{\gamma}_G$ se obtiene resolviendo

$$\mathbb{P}_n(\{Y^f - \gamma X\} \{X - E(X|W; \hat{\alpha}_1)\}) = 0,$$

donde $\hat{\alpha}_1$ es el estimador de máxima verosimilitud de α_1 bajo el modelo dado en (2.35). Es decir $\hat{\gamma}_G$ es consistente para $\gamma(g)$ si $p^f = gh_1 \in \mathcal{H}_1$ donde

$$\mathcal{H}_1 = \{p \in \mathcal{M} : h_1 \in \mathcal{V}_{sub,1}\}. \quad (2.37)$$

Para estudiar la consistencia del estimador $\hat{\gamma}_G$, notemos que si $p^f = gh_1$ pertenece a \mathcal{H}_1 entonces h_1 pertenece al conjunto $\mathcal{V}_{sub,1}$. Luego existe un único elemento $\alpha_1(h_1)$ en Ξ_1 que determina a h_1 : $\alpha_1(h_1)$ es tal que $h_1 = h_{1, \alpha_1(h_1)}$. Tenemos entonces que

$$E_{gh_1, \alpha_1(h_1)}\left(\left((Y^f - \gamma(g)X)(X - E(X|W; \alpha_1(h_1)))\right)\right) = E\left(t_g(W)(X - E(X|W))\right) = 0. \quad (2.38)$$

Por último, como $\hat{\gamma}_G$ resuelve la versión empírica de la ecuación (2.38), imputando el estimador de máxima verosimilitud en el lugar de $\alpha_1(h_1)$, es de esperar que cuando p pertenezca al modelo \mathcal{H}_1 , $\hat{\gamma}_G$ converja a $\gamma(g)$.

Al igual que en el ejemplo anterior, para garantizar la consistencia de los estimadores $\hat{\gamma}_{reg}$ y $\hat{\gamma}_G$ se requiere que los respectivos modelos asumidos sean verdaderos. En otras palabras: $\hat{\gamma}_{reg}$ no es consistente si (2.32) es falso y $\hat{\gamma}_G$ no es consistente si (2.35) no es verdadero. Sin embargo hay una tercer propuesta de estimación para $\gamma(g)$ que confiere mayor protección. Es decir, hay un tercer estimador denotado por $\hat{\gamma}_{DP}$ que resulta consistente cuando (2.32) o (2.35) es correcto, sin necesidad de especificar cual de los dos modelos es el verdadero. Dicho estimador se calcula como la solución en γ de la

ecuación $\mathbb{P}_n(\{Y^f - \gamma X - b_1(W; \hat{\tau}_{1,LS})\} \{X - E(X|W; \hat{\alpha}_1)\}) = 0$, donde $\hat{\tau}_{1,LS}$ junto con $\hat{\gamma}_{reg}$ resuelven la ecuación (2.33).

Como en el primer ejemplo de este capítulo, la doble protección de $\hat{\gamma}_{DP}$ se debe al siguiente hecho: si llamamos

$$\mathcal{U}(\gamma, \tau_1, \alpha_1) = \{Y^f - \gamma X - b_1(W; \tau_1)\} \{X - E(X|W; \alpha_1)\},$$

tenemos que $\gamma(g)$ es el único valor satisfaciendo

$$E_{p^f}(\mathcal{U}(\gamma(g), \tau_1, \alpha_1)) = 0, \quad (2.39)$$

en cualquiera de los siguientes casos:

1. $p^f \in \mathcal{G}_1$, $\tau_1 = \tau_1(g)$ cualquiera sea el valor para α_1 .
2. $p^f \in \mathcal{H}_1$, $\alpha_1 = \alpha_1(h_1)$ cualquiera sea el valor para τ_1 .

Por el momento presentamos tres propuestas de estimación consistentes para $\gamma(g)$ bajo el modelo dado en (2.26), basadas en los modelos (2.32) y (2.35). Los estimadores son

1. $\hat{\gamma}_{reg}$, consistente si $p^f \in \mathcal{G}_1$,
2. $\hat{\gamma}_G$, consistente si $p^f \in \mathcal{H}_1$,
3. $\hat{\gamma}_{DP}$, consistente si $p^f \in \mathcal{G}_1 \cup \mathcal{H}_1$.

En la próxima sección vamos a estimar bajo el mismo modelo (2.26), pero en un escenario donde la variable respuesta no va a ser observada en todos los individuos.

2.2.2. Modelo de regresión parcialmente lineal con datos faltantes

En diferentes aplicaciones la variable respuesta no es observada en todos los individuos. En compensación, se dispone de información adicional dada mediante un vector V de variables post tratamiento, medidas antes que Y^f y después de X . En esta sección abordamos el siguiente escenario: nuestro objeto de interés es $\gamma(p^f)$ definido en (2.26), siendo que la variable Y^f no es observada en todos los individuos pero contamos además con un nuevo vector denotado por V post tratamiento y previo al instante en el que se mide la variable Y^f .

En la literatura hay distintas propuestas de estimación bajo el modelo (2.26) cuando Y^f es faltante. Por ejemplo, Wang Q y Sun Z, (2006), presentan estimadores para $\gamma(g)$ y $t_g(-)$ combinando métodos no paramétricos de estimación de densidad con métodos que imputan, para cada dato faltante, posibles valores para Y^f y luego, hacen un

análisis como si la muestra fuera completa. Otro enfoque es el de Wang, Q, Linton O y Hardle (2004), quienes para estimar a la media de Y^f bajo el modelo (2.26) cuando Y^f es faltante, proponen estimadores de $\gamma(g)$ y de $t_g(-)$ extendiendo el procedimiento descrito por Robinson (1988) al caso en que Y^f es faltante. Mencionamos también la propuesta dada por Bianco A, Boente G, Manteiga W y González A (2009), quienes para estimar en forma resistente una medida de posición asociada a Y^f cuando dicha variable es faltante, presentan una propuesta de estimación robusta para $\gamma(p^f)$ y para ciertos funcionales asociados a $t_g(-)$ a partir del procedimiento presentado por Wang, Q, Linton O y Hardle W, (2004).

Nosotros, a continuación, daremos un enfoque paramétrico al problema de estimar a $\gamma(g)$ bajo el modelo (2.26) cuando Y^f está perdida en una submuestra, que incluye procedimientos doble y múltiple protegidos. Nuevamente, las técnicas usadas para encontrar en este escenario estimadores doble protegidos, siguen la filosofía de trabajo de van der Laan M, Robins J. (2003) Capítulo 1, sección 1.5.

Siguiendo las ideas presentadas en el ejemplo en donde se quería medir el efecto del bypass gástrico en el peso de los pacientes de una cierta clínica, recordamos que si denotamos con R a la variable binaria que indica para que individuos se observó Y^f , tenemos que los datos observados son ahora n copias i.i.d. del vector $Z = (W, X, R, Y)$, donde $Y = RY^f + (1 - R) * .$ Como mencionamos en dicho ejemplo, sin suposiciones adicionales el parámetro de interés no está identificado con lo cual asumiremos que contamos con un vector de variables adicionales denotado por V las cuales fueron medidas entre X e Y^f , y asumiremos que son todas las variables predictoras para Y^f que se correlacionan con R , es decir supondremos la condición de MAR que en este contexto consiste en suponer que

$$P(R = 1|V, X, W, Y^f) = P(R = 1|V, X, W). \quad (2.40)$$

El próximo resultado, cuya demostración incluimos en el Apéndice B, brinda una expresión para $\gamma(p^f)$ que depende exclusivamente de la distribución de los datos observados bajo la suposición de MAR, implicando así la identificación de $\gamma(g)$. Cabe destacar que en el presente contexto los datos conforman una muestra del vector (W, X, V, R, Y) . Sin embargo, considerando que el mismo se obtiene mediante una transformación del vector (W, X, V, R, Y^f) , y dado que la condición MAR involucra la distribución de (W, X, V, R, Y^f) , resulta de utilidad considerar la distribución de este último vector para demostrar la identificabilidad de $\gamma(p^f)$ en términos de la distribución del vector (W, X, V, R, Y) , como se constata en el siguiente resultado.

Proposición 2.2.1. *Definimos*

$$\mathcal{F}^f \equiv \{p^f : p^f \sim (W, X, V, R, Y^f) \text{ donde } E_{p^f}(Y^f|X, W) = \gamma(p^f)X + t_{p^f}(W)\},$$

$$\mathcal{F} \equiv \{p : p \sim (W, X, V, R, Y) \text{ donde } E_p(E_p(Y|R = 1, V, X, W)|X, W) = \beta(p)X + e_p(W)\}.$$

Si el vector (W, X, V, R, Y^f) tiene una distribución $p^f \in \mathcal{F}^f$ y además se satisface la condición dada en (2.40) entonces, la distribución p del vector (W, X, V, R, Y) , con $Y = RY^f + (1 - R) * pertenece a \mathcal{F} y se verifica que $\gamma(p^f) = \beta(p)$ y $t_{p^f}(W) = e_p(W)$.$

El teorema anterior garantiza la identificabilidad del funcional $\gamma(p^f)$. Es decir, bajo el supuesto de MAR (2.40), tenemos que $\gamma(p^f) = \beta(p)$, y por consiguiente el funcional de interés queda determinado por la distribución de los datos observados. Además transforma el problema original en estimar $\beta(p)$ bajo el modelo

$$E_p(E_p(Y|R=1, V, X, W) | X, W) = \beta(p) X + e_p(W), \quad (2.41)$$

donde p es la distribución de $Z = (W, X, V, R, Y)$ y $e_p(\cdot)$ es una función desconocida. O sea, nuestro nuevo problema es el siguiente: dadas n copias i.i.d. del vector $Z = (W, X, V, R, Y)$ con distribución p , verificando

$$E_p(E_p(Y|R=1, V, X, W) | X, W) = \beta(p) X + e_p(W), \quad (2.42)$$

nos proponemos estimar a $\beta(p)$. Notar que de esta manera dejamos de lado al funcional $\gamma(p^f)$ y con él al problema de datos faltantes, enfocandonos ahora en la estimación de $\beta(p)$ bajo el modelo (2.42), a partir de una muestra del vector (W, X, V, R, Y) .

Al igual que en el ejemplo 1 damos una posible factorización de p de manera que el funcional $\beta(\cdot)$ depende solo de uno de los factores involucrados en la misma. A saber: $p = gh_1h_2$ donde

$$\begin{aligned} g(z) &= p_{Y|R,W,X,V}(y|r, v, x, w) p_{V|X,W}(v|x, w) p_W(w), \\ h_1(z) &= p_{X|W}(x|w), h_2(z) = p_{R|V,X,W}(r|v, x, w), \end{aligned}$$

luego, por definición, $\beta(\cdot)$ depende de p solo mediante g por lo cual pondremos $\beta(g)$ en lugar de $\beta(p)$. Lo mismo ocurre con $e_p(W)$, por lo tanto, de ahora en más pondremos $e_g(W)$.

Usando la factorización dada para p , el modelo (2.42) impone la siguiente restricción: existe un número $\beta(g) \in \mathbb{R}$, y una función $e_g(-)$ tales que

$$\nu_1(g) \equiv E_g(E_g(Y - \beta(g) X | R=1, V, X, W) | X, W) \equiv e_g(W), \quad (2.43)$$

es decir, existe un número $\beta(g)$ tal que

$$\nu_1(g) \equiv E_g(E_g(Y - \beta(g) X | R=1, V, X, W) | X, W), \quad (2.44)$$

no depende de X .

Si llamamos $Y^* = E_p(Y|R=1, V, X, W)$, resulta que el modelo (2.42) puede reescribirse como un modelo parcialmente lineal de la siguiente manera

$$E_p(Y^* | X, W) = \beta(p) X + e_p(W).$$

Si Y^* fuera una variable observada, podríamos aplicar los procedimientos de estimación desarrollados en la sección (2.2.1). Al no conocer dicha variable, en la próxima sección vamos a adaptar las técnicas presentadas en (2.2.1), reemplazando la variable Y^* por un buen predictor de la misma. Una forma posible de obtener un buen predictor es imponer un modelo paramétrico para $E_p(Y|R=1, V, X, W)$. Desarrollamos estas ideas en la próxima sección.

2.2.3. Estimadores basados en una primera regresión.

Una primer familia de estimadores para $\beta(g)$, surge al considerar un modelo paramétrico para

$$\nu_2(g) \equiv E_g(Y|R=1, W=\cdot, X=\cdot, V=\cdot). \quad (2.45)$$

Es decir, asumir que

$$\nu_2(g) \in \mathcal{R}_{sub,2} \equiv \{b_2(\cdot, \cdot, \cdot; \tau_2) : \tau_2 \in \Upsilon_2\}, \quad (2.46)$$

donde $b_2(W, X, V; \cdot)$ es una función suave conocida salvo por el parámetro finito-dimensional τ_2 y $\Upsilon_2 \subseteq \mathbb{R}^{r_2}$. En adelante, si $\nu_2(g) \in \mathcal{R}_{sub,2}$ utilizaremos $\tau_2(g)$ para denotar al único elemento para el cual $\nu_2(g) = b_2(\cdot, \cdot, \cdot; \tau_2(g))$. A continuación vamos a ver tres escenarios de estimación de $\beta(g)$ bajo el modelo considerado en (2.46). En los tres hacemos uso del estimador de regresión por mínimos cuadrados de $\tau_2(g)$, denotado por $\hat{\tau}_{2,LS}$, el cual se obtiene regresando a Y en (W, X, V) entre los individuos que tienen $R=1$ bajo el modelo dado por (2.46). Más precisamente: $\hat{\tau}_{2,LS}$ resuelve la ecuación en τ_2 dada por

$$\mathbb{P}_n \left(R \frac{\partial b_2(W, X, V; \tau_2)}{\partial \tau_2} (Y - b_2(W, X, V; \tau_2)) \right) = 0. \quad (2.47)$$

Una vez computado $\hat{\tau}_{2,LS}$, construimos $\tilde{Y} = b_2(W, X, V; \hat{\tau}_{2,LS})$. Además, en forma análoga a lo hecho en el ejemplo anterior definimos

$$\mathcal{G}_2 = \{p \in \mathcal{M} : \nu_2(g) \in \mathcal{R}_{sub,2}\}. \quad (2.48)$$

Construcción de $\hat{\beta}_1$

Recordamos que bajo el modelo (2.42), definimos $\nu_1(g) = e_g(-)$. Vamos ahora a asumir un modelo paramétrico para el funcional $\nu_1(g)$:

$$\nu_1(g) \in \mathcal{R}_{sub,1} \equiv \{b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\}, \quad (2.49)$$

donde $b_1(W; \cdot)$ es una función suave conocida, salvo $\tau_1 \in \Upsilon_1$. También definimos \mathcal{G}_1 siendo

$$\mathcal{G}_1 = \{p \in \mathcal{M} : \nu_1(g) \in \mathcal{R}_{sub,1}\}, \quad (2.50)$$

de forma tal que si $p = gh_1h_2$ pertenece a \mathcal{G}_1 , resulta que el funcional $\nu_1(g)$ pertenece al conjunto $\mathcal{R}_{sub,1}$. Luego, denotamos con $\tau_1(g)$ al único elemento de Υ_1 para el cual se verifica que $\nu_1(g) = b_1(\cdot; \tau_1(g))$.

En esta primer propuesta de estimación vamos a asumir que $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, es decir, asumiremos que los modelos (2.46) y (2.49) son ciertos.

Definimos los estimadores $(\hat{\beta}_1, \hat{\tau}_{1,LS})$ como una solución (β, τ_1) del sistema dado por

$$\mathbb{P}_n \left(\left(\begin{array}{c} X \\ \frac{\partial b_1}{\partial \tau_1}(W, \tau_1) \end{array} \right) (\tilde{Y} - \beta X - b_1(W, \tau_1)) \right) = 0. \quad (2.51)$$

Bajo condiciones de regularidad $\hat{\beta}_1$ será consistente para $\beta(g)$ si los modelos (2.46) y (2.49) son ambos correctos. La consistencia de esta propuesta se debe a que si $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, tenemos que

$$E_{gh_1h_2} \left(b_2(W, X, V; \tau_2(g)) | X, W \right) = \beta(g)X + b_1(W, \tau_1(g)), \quad (2.52)$$

de donde obtenemos que

$$E_{gh_1h_2} \left(\left(\begin{array}{c} X \\ \frac{\partial b_1}{\partial \tau_1}(W, \tau_1(g)) \end{array} \right) (b_2(W, X, V; \tau_2(g)) - \beta(g)X - b_1(W, \tau_1(g))) \right) = 0. \quad (2.53)$$

Como $(\hat{\beta}_1, \hat{\tau}_{1,LS})$ resuelve una versión empírica de la ecuación (2.53) imputando un estimador consistente de $\tau_2(g)$, es de esperar que bajo condiciones de unicidad y regularidad, $\hat{\beta}_1$ converja a $\beta(p)$.

Construcción de $\hat{\beta}_2$

Ahora asumiremos que $p = gh_1h_2 \in \mathcal{H}_1 \cap \mathcal{G}_2$, es decir vamos a suponer que los modelos (2.35) y (2.46) son correctos. En ese caso

$$E_{gh_1h_2} \left((b_2(W, X, V; \tau_2(g)) - \beta(g)X)(X - E(X|W; \alpha_1(g))) \right) = E_{gh_1h_2} \left(e_g(W)(X - E_{h_1}(X|W)) \right) = 0.$$

Esta última ecuación sugiere resolver la versión empírica de la misma. Es decir, definimos $\hat{\beta}_2$ siendo la solución en β de la ecuación dada por

$$\mathbb{P}_n \left(\left\{ \tilde{Y} - \beta X \right\} \left\{ X - E(X|W; \hat{\alpha}_1) \right\} \right) = 0.$$

Bajo condiciones de regularidad, si los modelos asumidos en (2.46) y (2.35) son ambos correctos, $\hat{\beta}_2$ converge a $\beta(g)$.

Construcción de $\widehat{\beta}_3$

Definimos ahora $\widehat{\beta}_3$, el tercer estimador de esta sección siendo la solución en β de la ecuación

$$\mathbb{P}_n \left(\left\{ \widetilde{Y} - \beta X - b_1(W; \widehat{\tau}_{1,LS}) \right\} \left\{ X - E(X|W; \widehat{\alpha}_1) \right\} \right) = 0,$$

donde $\widehat{\tau}_{1,LS}$ es el estimador definido como solución de la ecuación (2.51). Bajo condiciones de regularidad, $\widehat{\beta}_3$ será consistente si $p = gh_1h_2 \in \{\mathcal{H}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{G}_1 \cap \mathcal{G}_2\}$. Con lo cual, $\widehat{\beta}_3$ resulta doble protegido, pues es consistente si el modelo (2.46) es verdadero y alguno de los modelos dados por (2.49) o (2.35) es correcto. La doble protección de este estimador se debe a que, bajo condiciones de regularidad, si llamamos

$$U_3(\beta, \tau_1, \tau_2, \alpha_1) = \{b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1)\} \{X - E(X|W; \alpha_1)\},$$

tenemos que $\beta(g)$ es el único valor satisfaciendo

$$E_p(U_3(\beta(g), \tau_1, \tau_2, \alpha_1)) = 0, \quad (2.54)$$

en cualquiera de los siguientes casos:

1. $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, $\tau_1 = \tau_1(g)$, $\tau_2 = \tau_2(g)$, cualesquiera sea el valore para α_1 .
2. $p \in \mathcal{H}_1 \cap \mathcal{G}_2$, $\alpha_1 = \alpha_1(h_1)$, $\tau_2 = \tau_2(g)$, cualesquiera sea el valore para τ_1 .

Por el momento construimos 3 estimadores, a saber:

1. $\widehat{\beta}_1$ consistente si $p \in \mathcal{G}_1 \cap \mathcal{G}_2$.
2. $\widehat{\beta}_2$ consistente si $p \in \mathcal{H}_1 \cap \mathcal{G}_2$.
3. $\widehat{\beta}_3$ consistente si $p \in \{\mathcal{G}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{H}_1 \cap \mathcal{G}_2\}$.

Al asumir como válido el modelo (2.46) los estimadores $\widehat{\beta}_1$, $\widehat{\beta}_2$ y $\widehat{\beta}_3$ se construyeron como $\widehat{\beta}_{reg}$, $\widehat{\beta}_G$ y $\widehat{\beta}_{DP}$ pero reemplazando Y^f por $b_2(W, X, V; \widehat{\tau}_{2,LS})$. Esa fue la estrategia de esta sección para lidiar con la pérdida de la variable respuesta. Veamos otra estrategia posible basada en pesos.

2.2.4. Estimadores basados en pesos.

En esta sección, al igual que en la anterior, asumiremos que vale el modelo (2.42) y procuramos nuevas estrategias de estimación para $\beta(g)$. Para ello, incorporamos un

nuevo modelo paramétrico, esta vez para h_2 . Es decir, si $p = gh_1h_2$ de ahora en más a lo largo de la sección asumiremos que

$$h_2 \in \mathcal{V}_{sub,2} \equiv \{p_{R|X,W,V}(\cdot|\cdot, \cdot, \cdot; \alpha_2) : \alpha_2 \in \Xi_2\} \quad (2.55)$$

donde $\Xi_2 \subseteq \mathbb{R}^{\ell_2}$ y para cada $\alpha_2 \in \Xi_2$ resulta que $p_{R|X,W,V}(\cdot|\cdot, \cdot, \cdot; \alpha_2)$ es una probabilidad condicional de R dado (X, W, V) .

Denotaremos con $\alpha_2(h_2)$ al único elemento de Ξ_2 que identifica a h_2 : $h_2 = h_{2,\alpha_2(h_2)}$. Además, denotaremos con $\hat{\alpha}_2$ al estimador de máxima verosimilitud de $\alpha_2(h_2)$ bajo el modelo dado en (2.55). Como en los casos anteriores definimos

$$\mathcal{H}_2 = \{p \in \mathcal{M} : h_2 \in \mathcal{V}_{sub,2}\}. \quad (2.56)$$

Al igual que lo hecho durante el ejemplo 1, en adelante, utilizaremos la siguiente notación:

$$E(R|X, W, V; \alpha_2) =: E_{h_2, \alpha_2}(R|X, W, V).$$

Brindaremos ahora tres propuestas de estimación basadas en asumir (2.55). Para eso enunciamos el siguiente resultado, cuya demostración presentamos en el Apéndice B.

Lema 2.2.1. *Para cualquier función $\psi(W, X, V, Y)$ tenemos que*

$$E\left(\psi(W, X, V, Y) \frac{R}{E(R|W, X, V)}\right) = E\left(E(\psi(W, X, V, Y)|R=1, W, X, V)\right). \quad (2.57)$$

Ahora si, construyamos nuevas propuestas de estimación basadas en asumir como verdadero el modelo dado en (2.55).

Construcción de $\tilde{\beta}_1$

Supongamos que $p = gh_1h_2 \in \mathcal{G}_1 \cap \mathcal{H}_2$. Usando el Lema 2.2.1 resulta

$$E_{gh_1h_2}\left(\frac{R}{E(R|X, W, V; \alpha_2(h_2))} \left(\frac{\partial b_1}{\partial \tau_1}(W, \tau_1(g))\right) (Y - \beta(g)X - b_1(W, \tau_1(g)))\right) = 0,$$

de donde, para estimar a $\beta(g)$, proponemos considerar $(\tilde{\beta}_1, \tilde{\tau}_{1,LS})$ solución de la versión empírica de la ecuación anterior. Es decir tomamos la solución en (β, τ_1) del sistema

$$\mathbb{P}_n\left(\frac{R}{E(R|X, W, V; \hat{\alpha}_2)} \left(\frac{\partial b_1}{\partial \tau_1}(W, \tau_1)\right) (Y - \beta X - b_1(W, \tau_1))\right) = 0, \quad (2.58)$$

en vista de lo cual, si $p = gh_1h_2 \in \mathcal{G}_1 \cap \mathcal{H}_2$, bajo condiciones de regularidad, $\tilde{\beta}_1$ converge a $\beta(g)$.

Construcción de $\tilde{\beta}_2$

Ahora vamos a suponer que se verifica el modelo dado en (2.35), o sea asumimos que $p = gh_1h_2 \in \mathcal{H}_1 \cap \mathcal{H}_2$. De ser así, usando el Lema 2.2.1 y la definición de esperanza condicional resulta que

$$E_{gh_1h_2} \left(\frac{R}{E(R|X, W, V; \alpha_2(g))} \{Y - \beta(g)X\} \{X - E(X|W; \alpha_1(g))\} \right) = E(\{e_p(W)\} \{X - E(X|W)\}) = 0. \quad (2.59)$$

Esto sugiere que, bajo hipótesis de regularidad, un estimador consistente de $\beta(g)$ se obtiene al resolver una versión empírica de la ecuación dada en (2.59). Definimos entonces $\tilde{\beta}_2$ siendo la solución de

$$\mathbb{P}_n \left(\frac{R}{E(R|X, W, V; \hat{\alpha}_2)} \{Y - \beta X\} \{X - E(X|W; \hat{\alpha}_1)\} \right) = 0.$$

Construcción de $\tilde{\beta}_3$

El último estimador de esta sección es $\tilde{\beta}_3$, el cual se obtiene resolviendo en β la ecuación

$$\mathbb{P}_n \left(\frac{R}{E(R|X, W, V; \hat{\alpha}_2)} \{Y - \beta X - b_1(W; \tilde{\tau}_{1,LS})\} \{X - E(X|W; \hat{\alpha}_1)\} \right) = 0$$

donde $\tilde{\tau}_{1,LS}$ es el estimador definido como la solución de la ecuación (2.58). Este estimador resulta consistente si se verifican los modelos (2.49) y (2.55) o se verifican (2.35) y (2.55), con lo cual resulta doble protegido. Esto se debe a que si llamamos

$$\tilde{U}(\beta, \tau_1, \alpha_1, \alpha_2) = \frac{R}{E(R|X, W, V; \alpha_2)} \{Y - \beta X - b_1(W; \tau_1)\} \{X - E(X|W; \alpha_1)\}$$

y $p = gh_1h_2$, entonces $\beta(g)$ es el único valor satisfaciendo

$$E_p \left(\tilde{U}(\beta(g), \tau_1, \alpha_1, \alpha_2) \right) = 0, \quad (2.60)$$

en cualquiera de los siguientes casos:

1. $p \in \mathcal{G}_1 \cap \mathcal{H}_2$, $\tau_1 = \tau_1(g)$, $\alpha_2 = \alpha_2(h_2)$, cualquiera sea el valor para α_1 .
2. $p \in \mathcal{H}_1 \cap \mathcal{H}_2$, $\alpha_1 = \alpha_1(h_1)$, $\alpha_2 = \alpha_2(h_2)$, cualquiera sea el valor para τ_1 .

En esta sección construimos 3 estimadores más, a saber

1. $\tilde{\beta}_1$ consistente si $p \in \mathcal{G}_1 \cap \mathcal{H}_2$.
2. $\tilde{\beta}_2$ consistente si $p \in \mathcal{H}_1 \cap \mathcal{H}_2$.
3. $\tilde{\beta}_3$ consistente si $p \in \{\mathcal{G}_1 \cap \mathcal{H}_2\} \cup \{\mathcal{H}_1 \cap \mathcal{H}_2\}$.

Al asumir como válido el modelo (2.55), los estimadores $\tilde{\beta}_1$, $\tilde{\beta}_2$ y $\tilde{\beta}_3$ se construyeron como $\hat{\beta}_{reg}$, $\hat{\beta}_G$ y $\hat{\beta}_{DP}$ pero reemplazando Y^f por Y multiplicado por $R/E(R|X, W, V; \hat{\alpha}_2)$, que funciona como un peso para la variable Y .

A lo largo de las últimas dos secciones, hemos presentado diferentes propuestas para estimar $\beta(p)$ bajo el modelo (2.42). Notablemente, en la próxima sección presentamos una séptima propuesta de estimación que confiere más protección que las conferidas por las construcciones presentadas hasta el momento.

2.2.5. Propuesta cuádruple protegida.

En esta sección construiremos $\hat{\beta}_{MP}$, un estimador consistente para $\beta(g)$ cuando se verifican las condiciones bajo las cuales $\hat{\beta}_3$ o $\tilde{\beta}_3$ resultan consistentes, sin necesidad de especificar cual de los escenarios es verdadero. Como en el ejemplo 1, la construcción de $\hat{\beta}_{MP}$ se basa en la existencia un estimador doble protegido $\hat{\tau}_{1,MP}$ de $\tau_1(g)$. Es decir $\hat{\tau}_{1,MP}$ es consistente para $\tau_1(g)$ si $p \in \{\mathcal{G}_1 \cap \mathcal{H}_2\} \cup \{\mathcal{G}_1 \cap \mathcal{G}_2\}$. En este caso, el par $(\hat{\tau}_{1,MP}, \hat{\beta}_{MP})$ se computa en forma conjunta. Más precisamente $(\hat{\beta}_{MP}, \hat{\tau}_{1,MP})$ es la solución del sistema dado por $\mathbb{P}_n(U(\beta, \tau_1, \hat{\tau}_{2,LS}, \hat{\alpha}_1, \hat{\alpha}_2)) = 0$ y $\mathbb{P}_n(T(\beta, \tau_1, \hat{\tau}_{2,LS}, \hat{\alpha}_1, \hat{\alpha}_2)) = 0$ donde

$$U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) \equiv H(\beta, \tau_1, \tau_2, \alpha_2) \{X - E(X|W; \alpha_1)\}, \quad (2.61)$$

$$T(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) \equiv H(\beta, \tau_1, \tau_2, \alpha_2) \{X - E(X|W; \alpha_1)\} d(X, W, \beta, \tau_1, \tau_2, \alpha_1, \alpha_2), \quad (2.62)$$

con $d(X, W, \beta, \tau_1, \tau_2, \alpha_1, \alpha_2) = \{X - E(X|W; \alpha_1)\} \frac{\partial b_1}{\partial \tau_1}(W, \tau_1)$ y

$$H(\beta, \tau_1, \tau_2, \alpha_2) \equiv \frac{R \{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1),$$

mientras que $\hat{\alpha}_i$ denota al estimador de máxima verosimilitud bajo el modelo \mathcal{H}_i y $\hat{\tau}_{2,LS}$ es el estimador que resuelve la ecuación (2.47).

El estimador $\hat{\beta}_{MP}$ converge a $\beta(g)$ si

$$p = gh_1h_2 \in \{\mathcal{H}_1 \cap \mathcal{H}_2\} \cup \{\mathcal{G}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{H}_1 \cap \mathcal{G}_2\} \cup \{\mathcal{G}_1 \cap \mathcal{H}_2\}, \quad (2.63)$$

donde \mathcal{G}_k , $k = 1, 2$ y \mathcal{H}_k , $k = 1, 2$ están definidos en (2.49), (2.46), (2.35) y (2.55) respectivamente. Este hecho se deduce al adaptar el Lema 2.1.3 al presente contexto. Veremos en el Capítulo 4 que

$$E_p(U(\beta(g), \tau_1, \tau_2, \alpha_1, \alpha_2)) = 0, \quad (2.64)$$

en cualquiera de los siguientes casos:

1. $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, $\tau_1 = \tau_1(g)$, $\tau_2 = \tau_2(g)$, $\alpha_1 \in \Xi_1$ y $\alpha_2 \in \Xi_2$. Es decir, si $p \in \mathcal{G}_1 \cap \mathcal{G}_2$ la ecuación (2.64) se verifica tomando $\tau_1 = \tau_1(g)$, $\tau_2 = \tau_2(g)$, para todo $\alpha_1 \in \Xi_1$, $\alpha_2 \in \Xi_2$.
2. $p \in \mathcal{H}_1 \cap \mathcal{H}_2$, $\alpha_1 = \alpha_1(h_1)$, $\alpha_2 = \alpha_2(h_2)$, $\tau_1 \in \Upsilon_1$ y $\tau_2 \in \Upsilon_2$. Es decir, si $p \in \mathcal{H}_1 \cap \mathcal{H}_2$ la ecuación (2.64) se verifica tomando $\alpha_1 = \alpha_1(h_1)$, $\alpha_2 = \alpha_2(h_2)$, para todo $\tau_1 \in \Upsilon_1$, $\tau_2 \in \Upsilon_2$.
3. $p \in \mathcal{H}_1 \cap \mathcal{G}_2$, $\alpha_1 = \alpha_1(h_1)$, $\tau_2 = \tau_2(g)$, $\alpha_2 \in \Xi_2$, y $\tau_1 \in \Upsilon_1$. Es decir, si $p \in \mathcal{H}_1 \cap \mathcal{G}_2$, la ecuación (2.64) se verifica tomando $\alpha_1 = \alpha_1(h_1)$, $\tau_2 = \tau_2(g)$, para todo $\alpha_2 \in \Xi_2$, y $\tau_1 \in \Upsilon_1$.
4. $p \in \mathcal{G}_1 \cap \mathcal{H}_2$, $\tau_1 = \tau_1(g)$, $\alpha_2 = \alpha_2(h_2)$, $\alpha_1 \in \Xi_1$, $\tau_2 \in \Upsilon_2$. Es decir, si $p \in \mathcal{G}_1 \cap \mathcal{H}_2$, la ecuación (2.64) se verifica tomando $\tau_1 = \tau_1(g)$, $\alpha_2 = \alpha_2(h_2)$ para todo $\alpha_1 \in \Xi_1$, $\tau_2 \in \Upsilon_2$.

El procedimiento de estimación para $\beta(g)$ recién descrito puede resumirse de la siguiente manera: construimos $\hat{\tau}_{1,MP}$, $\hat{\tau}_{2,LS}$, $\hat{\beta}_{MP}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$ resolviendo conjuntamente para (β, τ, α) , con $\alpha = (\alpha_1, \alpha_2)$ y $\tau = (\tau_1, \tau_2)$ el sistema de ecuaciones

$$\mathbb{P}_n(G(\beta, \tau, \alpha)) = 0, \quad (2.65)$$

donde $G = (U', T', S')'$ con $S \equiv (S_1, S_2)$, $S_j(\alpha_j) = \partial \log h_{j,\alpha_j} / \partial \alpha_j$ y $T \equiv (T_1, T_2)$ con $T_1(\beta, \tau, \alpha)$ como la función T definida en (2.62) y $T_2(\beta, \tau, \alpha) = T_2(\tau_2) = R \frac{\partial b_2(W, X, V; \tau_2)}{\partial \tau_2} (Y - b_2(W, X, V; \tau_2))$.

2.3. Estudio de Monte Carlo

En esta sección vamos a exponer los resultados que obtuvimos en un estudio de simulación diseñado con el objetivo de evaluar el comportamiento con muestras finitas del estimador múltiple protegido expuesto en la sección anterior. El contexto es el de regresión semiparamétrica con datos faltantes. Generamos las variables involucradas de la siguiente manera:

1. $W = (W_1, W_2) \sim N(\mu_W, \Sigma_W)$, donde usamos $\mu_W = (0, 0)^t$, $\Sigma_W = I_2$, con I_n la matriz identidad en $\mathbb{R}^{n \times n}$.
2. X binaria tomando valores 0 y 1 de acuerdo con

$$P(X = 1 | W_1 = w_1, W_2 = w_2) = \text{expit}(b_1 + b_2 w_1 + b_3 w_2 + b_4 w_1 w_2),$$

donde $\text{expit}(u) = \frac{e^u}{1+e^u}$ y tomamos $b = (-0.1, -0.5, 1.5, 1)$.

3. $V \sim \text{expit}((\gamma_1 + \gamma_2 X)T + (\gamma_3 + \gamma_4 X))$, con $T \sim N(0, 1)$ independiente de (X, W) . Tomamos $\gamma = (1, 0.5, -0.5, 0.4)$.
4. R binaria tomado valores 0 y 1 de acuerdo con

$$P(R = 1 | W = (w_1, w_2), X = x, V = v) = \text{expit}(d_1 + d_2 x + d_3 v + d_4 x v + d_5 v^2),$$

donde tomamos $d = (0.2, 0.5, 0.25, -1, -0.5)$.

5. Finalmente generamos Y en la forma

$$Y = a_1 + a_2 X + a_3 V + a_4 V^2 + a_5 W_1 + a_6 W_2 + a_7 W_1 W_2 + U,$$

donde $U \sim N(0, \sigma_1^2)$, independiente de (W, X, V, R) . Tomamos $a = (0.2, 1, 0.2, -1.4, 0.5, 0.25, -1.5)$ y $\sigma_1 = 1$.

De acuerdo a estas distribuciones, la probabilidad de que R sea 0 es aproximadamente 0.54.

La variable V fue generada como una transformación de una variable normal T y la variable dicotómica X de forma tal que resulte una variable continua, acotada ($0 < V < 1$) y correlacionada con X . Construir V acotada garantiza que $P(R = 1 | W = (w_1, w_2), X = x, V = v) = \text{expit}(0.2 + 0.5x + 0.25v - xv - 0.5v^2)$ está acotada inferiormente por una constante positiva. De hecho, estudiando para $x = 0$ y $x = 1$ la función $\zeta(v) = \text{expit}(0.2 + 0.5x + 0.25v - xv - 0.5v^2)$ resulta que

$$\sigma = 0,36 < P(R = 1 | W = (w_1, w_2), X = x, V = v).$$

De ahora en más denotaremos con $p = gh_1h_2$ a la distribución de (W, X, V, R, Y) , donde

$$\begin{aligned} g(z) &= p_{Y|R,W,X,V}(y|r, v, x, w) p_{V|X,W}(v|x, w) p_W(w), \\ h_1(z) &= p_{X|W}(x|w), h_2(z) = p_{R|V,X,W}(r|v, x, w). \end{aligned}$$

Para verificar que los datos generados con el procedimiento descrito satisfacen el modelo (2.41) y determinar el valor del parámetro $\beta(g)$, necesitamos una expresión para $E(E(Y|R=1, V, X, W)|W, X)$. Al ser $Y = 0,2 + X + 0,2V - 1,4V^2 + 0,5W_1 + 0,25W_2 - 1,5W_1W_2 + U$, con U independiente de (W, X, V, R) , resulta que

$$Z \equiv E(Y|R=1, V, X, W) = 0,2 + X + 0,2V - 1,4V^2 + 0,5W_1 + 0,25W_2 - 1,5W_1W_2.$$

Luego, para calcular $E(Z|XW)$ necesitamos $E(V|XW)$ y $E(V^2|XW)$. Ahora bien, tenemos que

$$V = \text{expit}(-0,5 + 0,4X + T + 0,5XT),$$

con $T \sim N(0, 1)$ independiente de (X, W) . Consideremos los siguientes cálculos.

1. para $X = 0$

$$E(V|X=0, W) = E(\text{expit}(-0,5 + T)|X=0W) = E(\text{expit}(-0,5 + T)) \equiv \ell_1.$$

2. Para $X = 1$

$$E(V|X=1, W) = E(\text{expit}(-0,1 + 1,5T)|X=1W) = E(\text{expit}(-0,1 + 1,5T)) \equiv \ell_2.$$

3. Para $X = 0$

$$E(V^2|X=0, W) = E(\{\text{expit}(-0,5 + T)\}^2|X=0W) = E(\{\text{expit}(-0,5 + T)\}^2) \equiv \ell_3.$$

4. Para $X = 1$

$$E(V^2|X=1, W) = E(\{\text{expit}(-0,1 + 1,5T|X=0W)\}^2) = E(\{\text{expit}(-0,1 + 1,5T)\}^2) \equiv \ell_4.$$

Luego por los items 1) y 2) resulta que

$$E(V|XW) = \ell_1(1 - X) + \ell_2X,$$

y por los items 3) y 4) resulta que

$$E(V^2|XW) = \ell_3(1 - X) + \ell_4X.$$

Podemos entonces concluir que

$$E[E(Y|R=1, V, X, W) | X, W] = \beta(g)X + e_g(W), \quad (2.66)$$

siendo

$$\beta(g) = 1 + 0,2(\ell_2 - \ell_1) - 1,4(\ell_4 - \ell_3) \quad (2.67)$$

$$e_g(W) = 0,2(1 + \ell_1) - 1,4\ell_3 + 0,5W_1 + 0,25W_2 - 1,5W_1W_2. \quad (2.68)$$

Hemos aproximado el valor de ℓ_i , $1 \leq i \leq 4$, calculando una versión empírica de la expresión que determina a cada uno de ellos, con tamaño muestra igual a $N_\ell = 1000000$, obteniendo los siguientes resultados $\ell_1 \approx 0,3997$, $\ell_2 \approx 0,4808$, $\ell_3 \approx 0,2035$ y $\ell_4 \approx 0,3100$. Finalmente, asumiendo de ahora en adelante como verdaderos los valores aproximados para ℓ_i , obtenemos que

$$\beta(g) = 0,86712, \quad e_g(W) = -0,00496 + 0,5W_1 + 0,25W_2 - 1,5W_1W_2.$$

A modo de resumen, tenemos que

- ★ $\beta(g) = 0,867$,
- ★ $e_g(W) = -0,00496 + 0,5W_1 + 0,25W_2 - 1,5W_1W_2$,
- ★ $E(Y|R=1, V, X, W) = 0,2 + X + 0,2V - 1,4V^2 + 0,5W_1 + 0,25W_2 - 1,5W_1W_2$,
- ★ $P(X=1|W_1=w_1, W_2=w_2) = \text{expit}(-0,1 - 0,5w_1 + 1,5w_2 + w_1w_2)$,
- ★ $P(R=1|W=(w_1, w_2), X=x, V=v) = \text{expit}(0,2 + 0,5x + 0,25v - xv - 0,5v^2)$.

Llamamos

$$h_1 \equiv P(X=1|W_1=\cdot, W_2=\cdot),$$

$$h_2 \equiv P(R=1|W=\cdot, X=\cdot, V=\cdot),$$

$$\nu_1(g) \equiv E_g(E_g(Y - \beta(g)X|R=1, V=\cdot, X=\cdot, W=\cdot) | X=\cdot, W=\cdot) \equiv e_g(W=\cdot),$$

$$\nu_2(g) \equiv E_g(Y|R=1, X=\cdot, V=\cdot, W=\cdot),$$

y recordamos que h_{1,α_1} es un modelo paramétrico para h_1 , h_{2,α_2} es un modelo paramétrico para h_2 , $b_1(W, \tau_1)$ es un modelo paramétrico para $\nu_1(g)$ y finalmente $b_2(W, X, V, \tau_2)$ es un modelo paramétrico para $\nu_2(g)$. Luego, en vista de como generamos a las variables de esta simulación, ajustaremos el modelo correcto si en cada caso tomamos

$$h_{1,\alpha_1} = \text{expit}(\alpha_1^1 + \alpha_1^2W_1 + \alpha_1^3W_2 + \alpha_1^4W_1W_2),$$

$$h_{2,\alpha_2} = \text{expit}(\alpha_2^1 + \alpha_2^2 * X + \alpha_2^3V + \alpha_2^4XV + \alpha_2^5V^2),$$

$$b_1(W, \tau_1) = \tau_1^1 + \tau_1^2W_1 + \tau_1^3W_2 + \tau_1^4W_1W_2,$$

$$b_2(W, X, V, \tau_2) = \tau_2^1 + \tau_2^2X + \tau_2^3V + \tau_2^4V^2 + \tau_2^5W_1 + \tau_2^6W_2 + \tau_2^7W_1W_2.$$

Cada modelo estará mal especificado si ajustamos

$$\begin{aligned} h_{1,\alpha_1} &= \text{expit}(\alpha_1^1 + \alpha_1^2 W_1 + \alpha_1^3 W_2), \\ h_{2,\alpha_2} &= \text{expit}(\alpha_2^1 + \alpha_2^2 * X + \alpha_2^3 V), \\ b_1(W, \tau_1) &= \tau_1^1 + \tau_1^2 W_1 + \tau_1^3 W_2, \\ b_2(W, X, V, R, \tau_2) &= \tau_2^1 + \tau_2^2 X + \tau_2^3 V + \tau_2^4 W_1 + \tau_2^5 W_2. \end{aligned}$$

Básicamente, en este trabajo consideramos que nos equivocamos omitiendo los términos cuadráticos. Otras elecciones son posibles, como por ejemplo omitir algún término lineal. La simulación consistió en generar 5000 muestras de tamaño 1500 y 500 cada una y para cada tamaño de muestra calculamos los estimadores $\widehat{\beta}_{MP}, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3$ propuestos en las secciones anteriores en los siguientes escenarios: si $p = gh_1 h_2$ es la distribución de (W, X, V, R, Y) entonces

- ★ en el escenario 1 ajustamos los modelos bien especificado para h_2 y $\nu_1(g)$ y mal especificados para h_1 y $\nu_2(g)$,
- ★ en el escenario 2 ajustamos los modelos bien especificados para h_1 y h_2 y mal especificados para $\nu_1(g)$ y $\nu_2(g)$,
- ★ en el escenario 3 ajustamos los modelos bien especificados para $\nu_1(g)$ y $\nu_2(g)$ y mal especificados para h_1 y h_2 ,
- ★ en el escenario 4 ajustamos los modelos bien especificados para h_1 y $\nu_2(g)$ y mal especificados para h_2 y $\nu_1(g)$.

Si bien en otros escenarios podemos obtener estimadores consistentes (un escenario alternativo al 1 sería, por ejemplo, considerar bien especificados los modelos para $h_2, \nu_1(g)$ y h_1 y mal especificado el modelo para $\nu_2(g)$) elegimos ponernos en los peores casos en donde tenemos consistencia. Entendemos por peor caso, un escenario donde se verifican mínimas condiciones para que el estimador $\widehat{\beta}_{MP}$ sea consistente. Es decir, elegimos escenarios donde tenemos exactamente dos modelos bien especificados de los cuatro objetos a modelar. Con menos de dos modelos bien especificados no tenemos ningún resultado de consistencia, por eso entendemos que son condiciones mínimas.

Tenemos entonces para cada tamaño de muestra, 5000 realizaciones de cada uno de los estimadores vistos en la sección anterior. En las tablas dadas a continuación reportamos las siguientes medidas resumen muestrales para cada muestra de estimadores:

- 1) **Mediana** En esta columna calculamos la mediana muestral multiplicada por 100 para cada estimador.

- 2) **Media.** En esta columna calculamos la media empírica multiplicada por 100 de cada uno de los estimadores.
- 3) **Media 0.025.** En esta columna calculamos la media 0.025 podada de cada uno de los estimadores multiplicado por 100. Para calcular esta medida de posición, ordenamos la muestra y descartamos los primeros y los últimos 62 elementos de la muestra ordenada, tomando el promedio de los 5000 - 2 * 62 valores centrales. Para fijar ideas, si por ejemplo, $\hat{\beta}_1, \dots, \hat{\beta}_{5000}$ son las realizaciones del estimador $\hat{\beta}_{MP}$ en el escenario 1 para tamaño muestral $n = 1500$, luego

$$\text{Media } 0,025. = 85.95 = 100 * \frac{1}{4875} \sum_{i=63}^{4938} \hat{\beta}^{(i)},$$

donde $\hat{\beta}^{(i)}$ es el i -ésimo estadístico de orden de la muestra $\hat{\beta}_1, \dots, \hat{\beta}_{5000}$.

- 4) **RECM.** En esta columna calculamos la raíz cuadrada del error cuadrático medio muestral de cada uno de los estimadores multiplicado por 100. Es decir, si por ejemplo $\hat{\beta}_1, \dots, \hat{\beta}_{5000}$ son las realizaciones del estimador $\hat{\beta}_{MP}$ en el escenario 1 para tamaño muestral $n = 1500$, entonces

$$\text{RECM} = 13,69 = 100 * \sqrt{\frac{1}{5000} \sum_{i=1}^{5000} (\hat{\beta}_i - 0,867)^2}$$

- 5) **RECM 0.05.** En esta columna calculamos la raíz cuadrada del error cuadrático medio 0.05 podado de cada uno de los estimadores multiplicada por 100. Para calcular esta medida de escala, armamos una nueva muestra que consta del cuadrado de la diferencia entre cada realización del estimador y 0,867. Luego ordenamos la nueva muestra y descartamos los primeros y los últimos 125 elementos de la nueva muestra, tomando la raíz cuadrada del promedio de lo que queda. Para fijar ideas si por ejemplo, $\hat{\beta}_1, \dots, \hat{\beta}_{5000}$ son las realizaciones del estimador $\hat{\beta}_{MP}$ en el escenario 1 para tamaño muestral $n = 1500$, llamamos $w_i = (\hat{\beta}_i - 0,867)^2$, luego

$$\text{RECM } 0.05. = 12,17 = 100 * \sqrt{\frac{1}{4749} \sum_{i=126}^{4875} w^{(i)},}$$

donde $w^{(i)}$ es el i -ésimo estadístico de orden de la muestra w_1, \dots, w_{5000} .

- 6) **RECM R.** En esta columna calculamos la raíz cuadrada del error cuadrático medio robusto de cada uno de los estimadores multiplicada por 100. Específicamente, armamos una nueva muestra que consta del modulo de la diferencia entre cada realización del estimador y 0,867. Luego, el error cuadrático medio robusto reportado

es el cuadrado de la mediana de esa nueva muestra. Por ejemplo, si $\widehat{\beta}_1, \dots, \widehat{\beta}_{5000}$ son las realizaciones del estimador $\widehat{\beta}_{MP}$ en el escenario 1 para tamaño muestral $n = 1500$, entonces

$$\text{RECM R.} = 9,16 = 100 * \text{mediana}(|\widehat{\beta}_1 - 0,867|, |\widehat{\beta}_2 - 0,867|, \dots, |\widehat{\beta}_{5000} - 0,867|)$$

- 7) SD. En la columna SD reportamos el desvío standard muestral de cada estimador multiplicado por 100.

Al estudiar los resultados obtenidos en la simulación, observamos el buen desempeño del estimador $\widehat{\beta}_{MP}$ en el siguiente sentido: en los cuatro escenarios presentados, si bien en algunos casos se observaron realizaciones atípicas, la distribución del estimador se encuentra concentrada alrededor del valor de $\beta(p)$. De hecho, la mediana de $\widehat{\beta}_{MP}$ resulta siempre cercana a $\beta(p)$ y lo mismo ocurre con la media 0.025 podada. Por otro lado ninguno de los restantes estimadores tiene distribución concentrada alrededor de $\beta(p)$ en TODOS los escenarios contemplados. Esto ilustra la múltiple protección de $\widehat{\beta}_{MP}$ y su ventaja por sobre el resto de los estimadores. El lector podrá observar que en todos los escenarios hay un estimador más preciso que $\widehat{\beta}_{MP}$ (por ejemplo en el escenario 1, para tamaño muestral $n = 1500$, $\widetilde{\beta}_3$ tiene esta característica.) Esto no es sorprendente ya que evidentemente ilustra el sacrificio (“trade off ”) en términos de varianza que conlleva la múltiple protección.

A continuación para cada escenario presentamos las tablas con los cálculos de las medidas resumen muestrales descritas anteriormente, histogramas y diagramas de caja (o boxplots, usaremos indistintamente ambas expresiones) para cada muestra de estimadores y para cada tamaño muestral.

Escenario 1

Bien especificado los modelos para h_2 y $\nu_1(g)$.

Es de esperar que los siguientes estimadores sean consistentes: $\widehat{\beta}_{MP}$, $\widetilde{\beta}_1$, $\widetilde{\beta}_3$.

Obtenemos las siguientes tablas

a) caso $n = 1500$:

	Mediana	Media	Media 0.025.	RECM	RECM 0.05.	RECM R.	SD.
$\widehat{\beta}_{MP}$	85,87	85,97	85,95	13,69	12,17	9,16	13,67
$\widetilde{\beta}_1$	5,58	5,49	5,52	82,43	81,91	81,11	14,21
$\widetilde{\beta}_2$	5,55	5,44	5,48	82,48	81,95	81,14	14,22
$\widetilde{\beta}_3$	24,56	24,41	24,41	63,24	62,86	62,14	10,95
$\widetilde{\beta}_1$	95,06	94,96	94,96	11,96	11	8,86	8,96
$\widetilde{\beta}_2$	6,24	57,89	-0,0810	$4,5 * 10^3$	107,78	95,94	4632
$\widetilde{\beta}_3$	86,54	86,52	86,52	8,54	7,68	5,75	8,51

b) Caso $n = 500$:

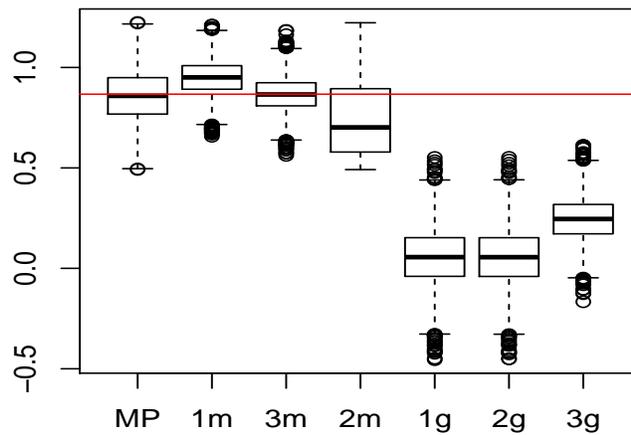
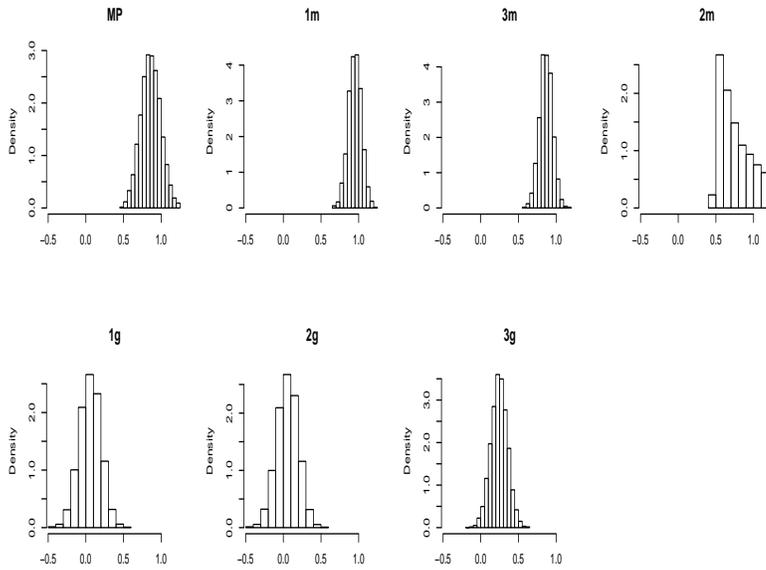
	Mediana	Media	Media 0.025.	RECM	RECM 0.05.	RECM R.	SD.
$\widehat{\beta}_{MP}$	86,11	87,06	86,76	39,82	22,02	15,74	39,8
$\widetilde{\beta}_1$	7,03	6,48	6,58	84	82,51	79,67	24,92
$\widetilde{\beta}_2$	6,65	6,32	6,42	84,15	82,63	80,04	24,92
$\widetilde{\beta}_3$	25,37	25,39	25,41	64,24	63,11	61,32	19,22
$\widetilde{\beta}_1$	94,78	94,81	94,82	17,29	15,56	11,86	17,26
$\widetilde{\beta}_2$	5,62	334,1	-1,31	7071	100	95,43	22859
$\widetilde{\beta}_3$	86,67	86,6	86,57	14,93	13,27	9,88	14,92

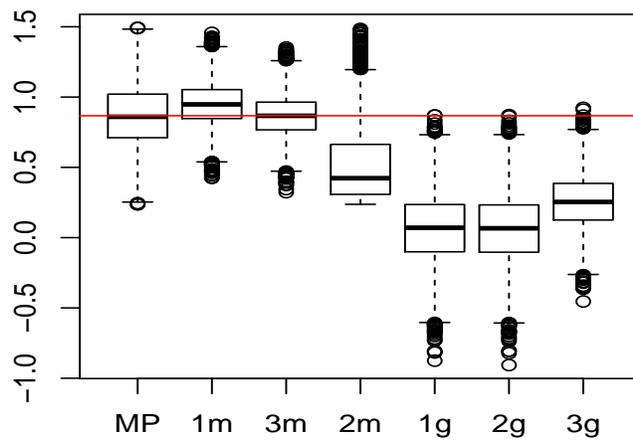
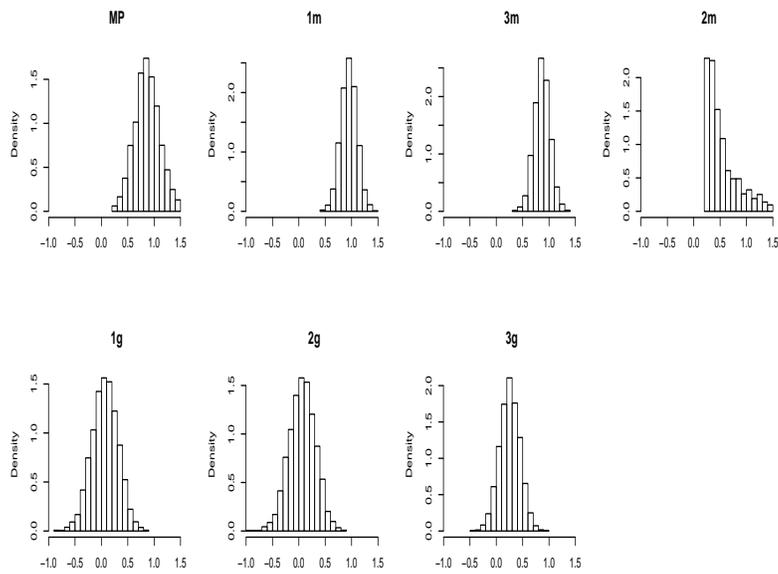
A continuación damos cuatro gráficos. En el primero presentamos los histogramas de cada uno de los estimadores calculados en el escenario 1, para tamaño de muestra $n = 1500$, habiendo omitido los outliers de los estimadores $\widehat{\beta}_{MP}$, $\widetilde{\beta}_1$, $\widetilde{\beta}_2$ y $\widetilde{\beta}_3$. Decidimos que una observación resultaba atípica si la misma es menor que el primer cuantil de la muestra menos 1.5 por la distancia intercuantil de la muestra o mayor que el tercer cuantil de la muestra más 1.5 por la distancia intercuantil asociada a dicha muestra de estimadores. Con ese criterio, en este escenario, el porcentaje mayor de observaciones atípicas fue de 0.86 % para tamaño de muestra $n = 1500$ y ocurrió para el estimador $\widehat{\beta}_{MP}$.

En el segundo gráfico, presentamos los diagramas de caja de dichos estimadores junto con una recta horizontal sobre el valor verdadero del parámetro (0,867), con $\widehat{\beta}_{MP}$, $\widetilde{\beta}_1$, $\widetilde{\beta}_2$ y $\widetilde{\beta}_3$ desprovistos de outliers, para tamaño de muestra $n = 1500$.

En el tercer y cuarto gráfico reportamos lo mismo que en el primero y segundo respectivamente pero para tamaño de muestra $n = 500$. Eliminamos outliers con el mismo criterio que en el caso $n = 1500$ obteniendo en este escenario que el porcentaje mayor de observaciones atípicas fue de 1.7 % y ocurrió para el estimador $\widehat{\beta}_{MP}$.

Adoptamos la siguiente convención notacional para los gráficos: para cada $1 \leq i \leq 3$ pondremos im para denotar al estimador $\hat{\beta}_i$, pondremos ig para denotar al estimador $\hat{\beta}_i$, y pondremos MP para denotar al estimador $\hat{\beta}_{MP}$. Por ejemplo, si en un gráfico aparece la expresión $1m$, se está haciendo alusión al estimador $\hat{\beta}_1$.





Escenario 2

Bien especificados los modelos para h_1 y h_2 .

Es de esperar que los siguientes estimadores sean consistentes: $\hat{\beta}_{MP}$, $\tilde{\beta}_2$, $\tilde{\beta}_3$.

En este caso los resultados son:

a) $n = 1500$

	Mediana	Media	Media 0.025.	RECM	RECM 0.05.	RECM R.	SD
$\widehat{\beta}_{MP}$	86,65	81,36**	86,82	727,03	25,88	12,79	727,02
$\widehat{\beta}_1$	14,6	14,55	14,60	73,52	72,94	72,09	14,15
$\widehat{\beta}_2$	5,63	5,49	5,53	82,44	81,91	81,06	14,22
$\widehat{\beta}_3$	5,63	5,49	5,53	82,44	81,91	81,06	14,22
$\widetilde{\beta}_1$	13,21	13,05	13,08	75,67	74,48	73,49	14,50
$\widetilde{\beta}_2$	86,51	86,38	86,38	11,62	10,4	7,81	11,59
$\widetilde{\beta}_3$	86,39	86,36	86,37	11,02	9,85	7,49	10,98

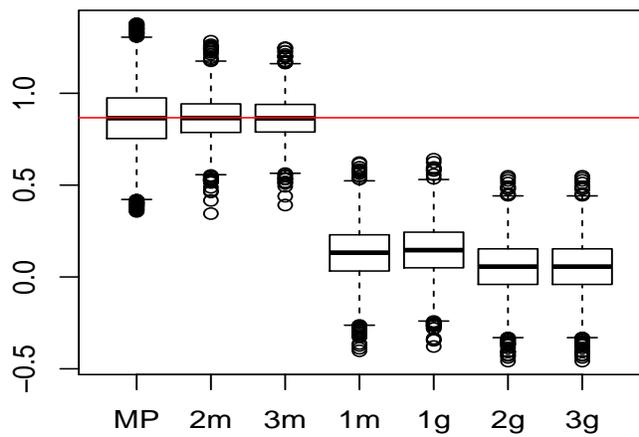
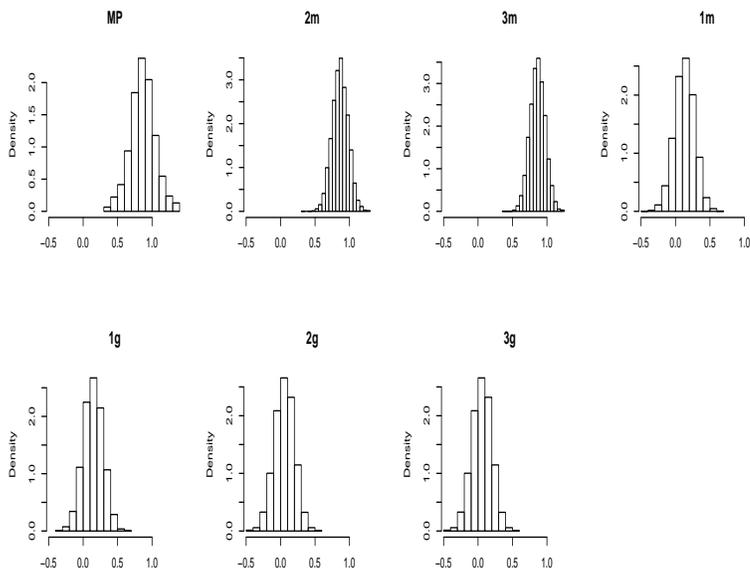
b) $n = 500$

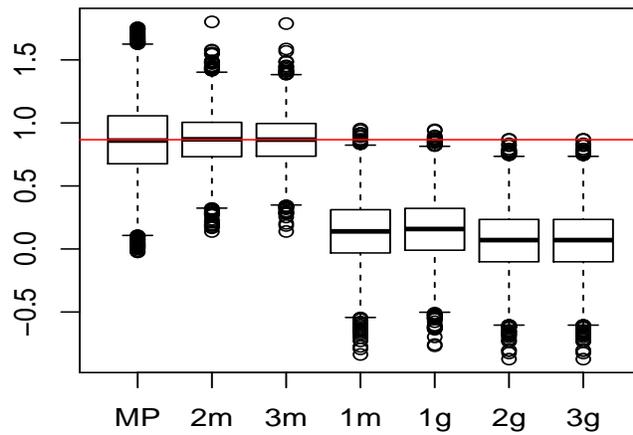
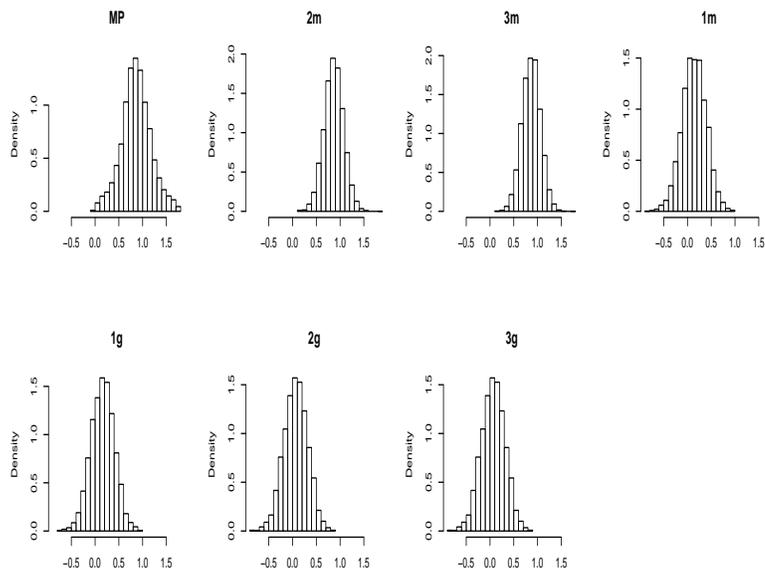
	Mediana	Media	Media 0.025.	RECM	RECM 0.05.	RECM R.	SD.
$\widehat{\beta}_{MP}$	85,84	80,67	85,73	324,29	48,12	22,34	324,26
$\widehat{\beta}_1$	15,84	15,51	15,61	75,37	73,68	70,85	24,79
$\widehat{\beta}_2$	7,05	6,50	6,60	83,98	82,46	79,65	24,93
$\widehat{\beta}_3$	7,05	6,50	6,60	83,98	82,46	79,65	24,93
$\widetilde{\beta}_1$	13,96	13,82	13,91	77,21	75,45	72,73	25,49
$\widetilde{\beta}_2$	87	86,8	86,79	20,45	18,19	13,62	20,40
$\widetilde{\beta}_3$	86,74	86,62	86,64	19,24	17,18	12,94	19,23

A continuación damos cuatro gráficos. En el primero presentamos los histogramas de cada uno de los estimadores calculados en el escenario 2, para tamaño de muestra $n = 1500$, habiendo omitido los outliers del estimador $\widehat{\beta}_{MP}$ con el criterio usado en el escenario 1. En este caso omitimos el 10% de las observaciones.

En el segundo gráfico presentamos los boxplot de cada uno de los estimadores calculados en el escenario 2 junto con una recta horizontal sobre el valor verdadero del parámetro (0,867), con $\widehat{\beta}_{MP}$ desprovisto de outliers para tamaño de muestra $n = 1500$.

En el tercer y cuarto gráfico reportamos lo mismo que en el primero y segundo pero para tamaño de muestra $n = 500$, omitiendo outliers del estimador $\widehat{\beta}_{MP}$ con el criterio usado anteriormente. En este caso descartamos el 6.12% de los datos. Recordamos nuevamente que adoptamos la siguiente convención notacional para los gráficos: para cada $1 \leq i \leq 3$ pondremos im para denotar al estimador $\widetilde{\beta}_i$, pondremos ig para denotar al estimador $\widehat{\beta}_i$ y pondremos MP para denotar al estimador $\widehat{\beta}_{MP}$. Por ejemplo, si en un gráfico aparece 1m se está haciendo alusión al estimador β_1 .





Escenario 3

Bien especificados los modelos para $\nu_1(g)$ y $\nu_2(g)$.

Es de esperar que los siguientes estimadores sean consistentes: $\hat{\beta}_{MP}$, $\hat{\beta}_1$, $\hat{\beta}_3$

Nos queda:

a) Caso $n = 1500$

	Mediana	Media	Media 0.025.	RECM	RECM 0.05.	RECM R.	SD
$\widehat{\beta}_{MP}$	86,54	83,27	86,47	261,68	12,41	8,34	261,65
$\widehat{\beta}_1$	86,43	86,51	86,52	8,42	7,48	5,62	8,39
$\widehat{\beta}_2$	3,62	3,49	3,53	83,98	83,68	83,07	11,45
$\widehat{\beta}_3$	86,64	86,62	86,63	6,72	6	4,57	6,71
$\widetilde{\beta}_1$	27,75	27,75	27,74	59	58,97	58,95	2,54
$\widetilde{\beta}_2$	6,48	6,30	6,33	81,7	81,2	80,21	14,56
$\widetilde{\beta}_3$	24,16	24,15	24,1	62,59	62,56	62,54	2,32

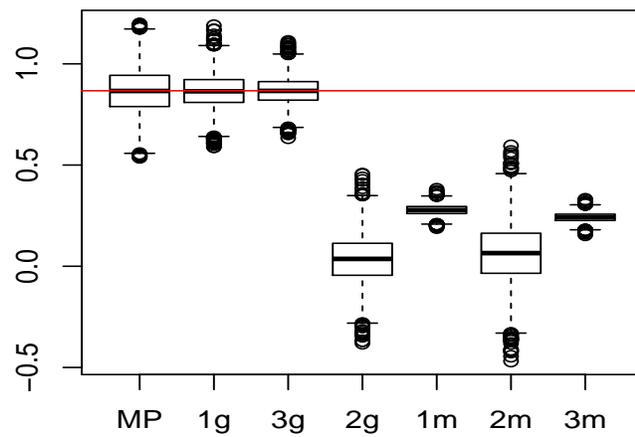
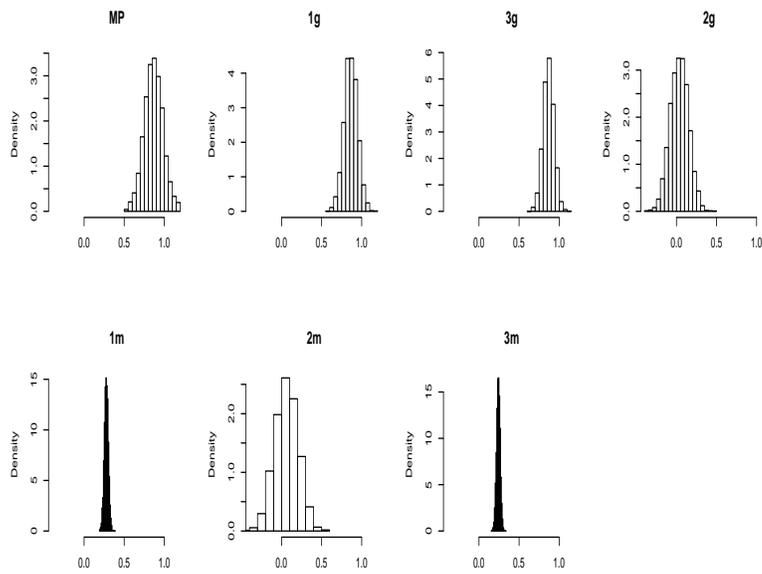
b) Caso $n = 500$

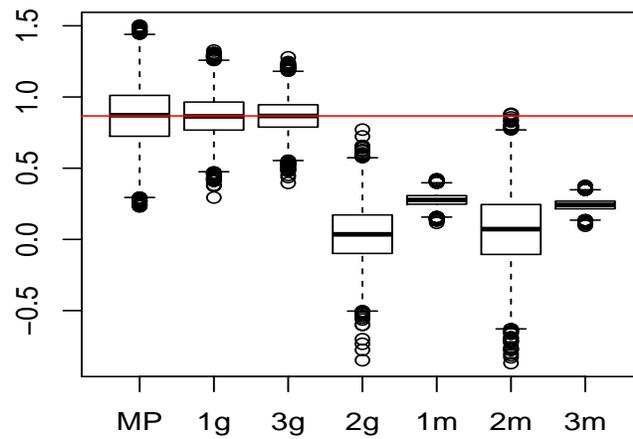
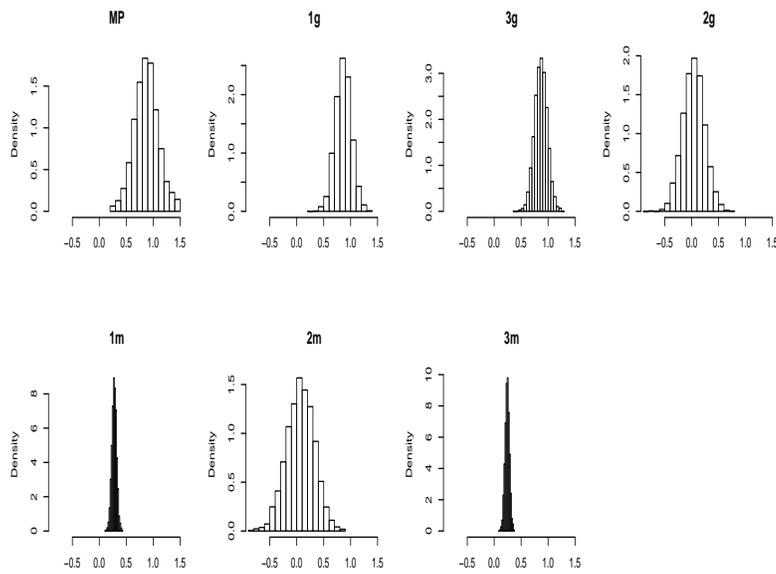
	Mediana	Media	Media 0.025.	RECM	ECM 0.05.	RECM R.	SD.
$\widehat{\beta}_{MP}$	87,13	244	86,97	11668	26,72	15,90	11668
$\widehat{\beta}_1$	86,39	86,49	86,49	14,8	13,19	9,88	14,79
$\widehat{\beta}_2$	3,54	3,63	3,63	85,45	84,55	83,15	20,10
$\widehat{\beta}_3$	86,68	86,62	86,62	11,83	10,53	7,84	11,82
$\widetilde{\beta}_1$	27,68	27,71	27,71	59,15	59,08	59,03	4,48
$\widetilde{\beta}_2$	7,20	6,77	6,86	83,97	82,32	70,51	25,76
$\widetilde{\beta}_3$	24,23	24,16	24,17	62,65	62,60	62,46	4,05

A continuación damos cuatro gráficos. En el primero presentamos los histogramas de cada uno de los estimadores calculados en el escenario 3, para tamaño de muestra $n = 1500$, habiendo omitido los outliers del estimador $\widehat{\beta}_{MP}$ con el criterio usado anteriormente. En este caso el porcentaje de observaciones atípicas fue 5.92%.

En el segundo diagrama, presentamos los boxplot de cada uno de los estimadores calculados en el escenario 3, junto con una recta horizontal sobre el valor verdadero del parámetro (0,867), con $\widehat{\beta}_{MP}$ desprovisto de outliers para tamaño de muestra $n = 1500$.

En el tercer y cuarto gráfico reportamos lo mismo que en el primero y segundo pero para tamaño de muestra $n = 500$, y la omisión de outliers del estimador $\widehat{\beta}_{MP}$ se realizó con el criterio usado anteriormente, obteniéndose un 8.06% de datos atípicos.





Escenario 4

Bien especificados los modelos para h_1 y $\nu_2(g)$.

Es de esperar que los siguientes estimadores sean consistentes $\hat{\beta}_{MP}$, $\hat{\beta}_2$, $\hat{\beta}_3$.

Obtenemos

a) $n = 1500$

	Mediana	Media	Media 0.025.	RECM	RECM 0.05.	RECM R.	SD.
$\widehat{\beta}_{MP}$	86,80	91,80	86,76	331,66	20,12	10,78	331,59
$\widehat{\beta}_1$	4,58	4,48	4,49	83,05	82,73	82,12	11,77
$\widehat{\beta}_2$	86,44	86,51	86,52	8,42	7,54	5,62	8,41
$\widehat{\beta}_3$	86,44	86,51	86,52	8,42	7,54	5,62	8,41
$\widetilde{\beta}_1$	10,87	10,74	10,78	77,3	76,73	75,82	14,40
$\widetilde{\beta}_2$	6,24	57,89	-0,081	4582	107,78	95,94	1330
$\widetilde{\beta}_3$	26,95	27	26,96	59,82	59,79	59,74	3,33

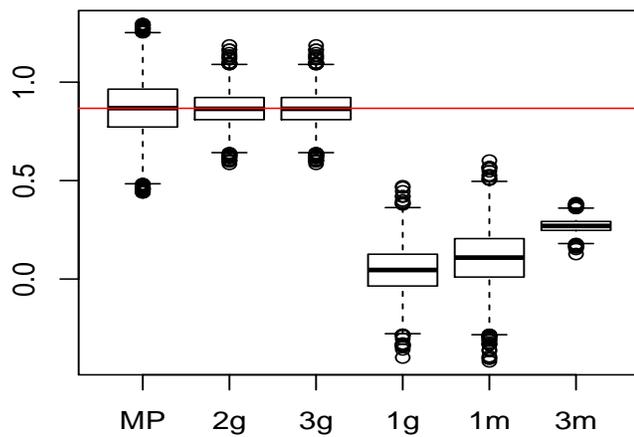
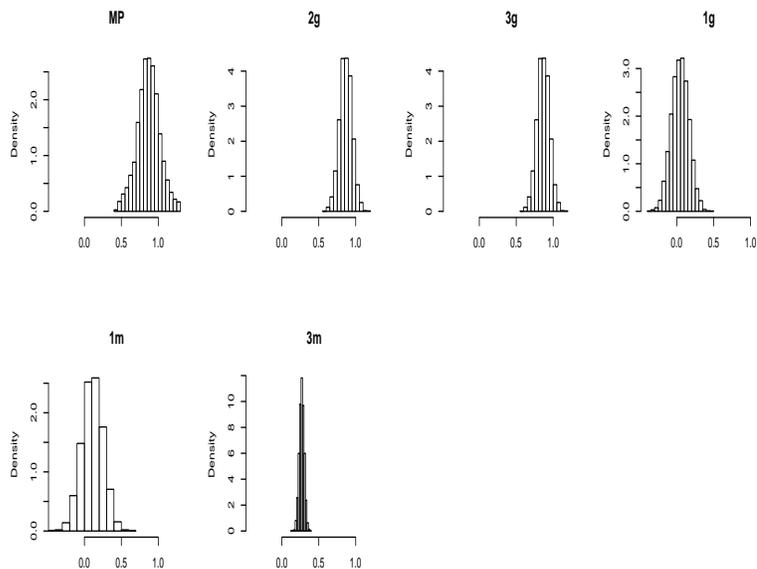
b) $n = 500$

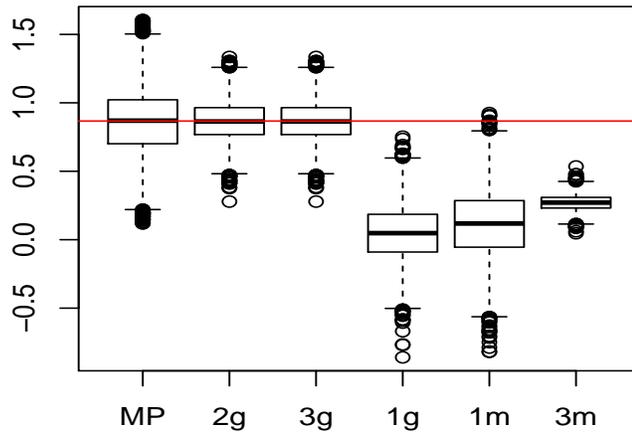
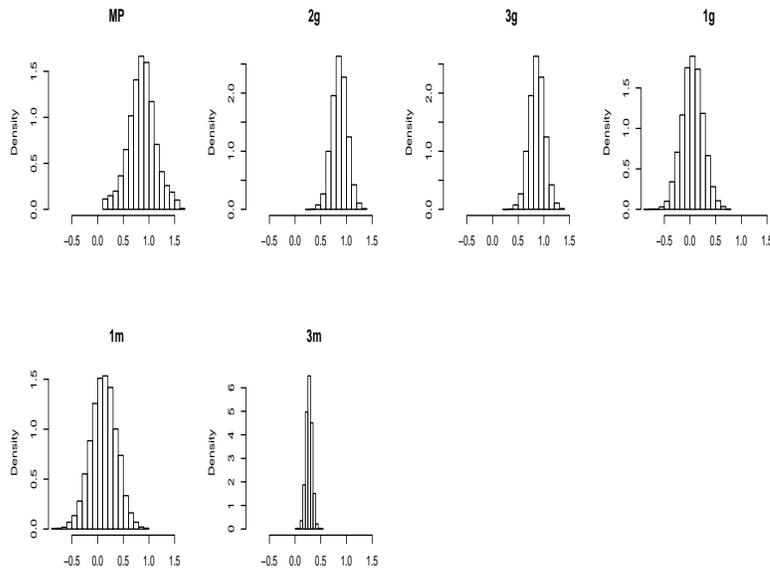
	Mediana	Media	Media 0.025.	RECM	RECM 0.05.	RECM R.	SD.
$\widehat{\beta}_{MP}$	87	82	86,66	411,5	37,87	18,69	411,1
$\widehat{\beta}_1$	4,78	4,65	4,66	84,68	90,50	71,59	21,32
$\widehat{\beta}_2$	86,37	86,5	86,49	14,78	13,19	9,86	14,76
$\widehat{\beta}_3$	86,37	86,5	86,49	14,78	13,19	9,86	14,76
$\widetilde{\beta}_1$	11,85	11,52	11,61	79,3	77,64	74,85	25,32
$\widetilde{\beta}_2$	5,62	334	-1,31	10^6	107,58	95,43	22860
$\widetilde{\beta}_3$	27,11	27,04	27,05	59,94	59,82	59,59	5,45

A continuación damos cuatro gráficos. En el primero presentamos los histogramas de cada uno de los estimadores calculados en el escenario 4, para tamaño de muestra $n = 1500$, salvo para el estimador $\widehat{\beta}_2$ quien presenta un comportamiento sumamente errático, debido a lo cual, tuvimos que omitirlo del análisis. Nuevamente, para hacer este gráfico descartamos los outliers del estimador $\widehat{\beta}_{MP}$ con el mismo criterio usado anteriormente. Tuvimos un 9.6% de datos atípicos.

En el segundo diagrama, presentamos los boxplot de cada uno de los estimadores calculados en el escenario 4, salvo para el estimador $\widehat{\beta}_2$, junto con una recta horizontal sobre el valor verdadero del parámetro (0,867), con $\widehat{\beta}_{MP}$ desprovisto de outliers para el mismo tamaño muestral.

En el tercer y cuarto gráfico reportamos lo mismo que en el primero y segundo pero para tamaño de muestra $n = 500$. En este caso las observaciones atípicas fueron el 7.6%.





Como anticipamos, en los cuatro escenarios el estimador $\hat{\beta}_{MP}$ resulta estar razonablemente cerca del valor del parámetro de interés. Mirando las tablas observamos que en el escenario dos y en el escenario cuatro $\hat{\beta}_2$ y $\hat{\beta}_3$ dan lo mismo. Esto se debe a que la elección particular de los modelos para h_1 y para $\nu_1(g)$ en estos escenarios hace que los estimadores $\hat{\beta}_2$ y $\hat{\beta}_3$ sean algebraicamente iguales.

En efecto, el estimador $\widehat{\beta}_2$ resuelve la ecuación

$$\mathbb{P}_n \left(\left\{ \widetilde{Y} - \beta X \right\} \left\{ X - E(X|W; \widehat{\alpha}_1) \right\} \right) = 0,$$

mientras que el estimador $\widehat{\beta}_3$ resuelve la ecuación

$$\mathbb{P}_n \left(\left\{ \widetilde{Y} - \beta X - b_1(W, \widehat{\tau}_1) \right\} \left\{ X - E(X|W; \widehat{\alpha}_1) \right\} \right) = 0,$$

donde $\widetilde{Y} = b_2(W, X, V; \widehat{\tau}_2)$. Estas ecuaciones son idénticas (y por lo tanto $\widehat{\beta}_2$ y $\widehat{\beta}_3$ son iguales) si se verifica

$$\mathbb{P}_n (b_1(W, \widehat{\tau}_1) \{X - E(X|W; \widehat{\alpha}_1)\}) = 0.$$

Por otro lado, en ambos escenarios

$$E(X|W, \alpha_1) = \text{expit}(\alpha_1^1 + \alpha_1^2 W_1 + \alpha_1^3 W_2 + \alpha_1^4 W_1 W_2),$$

en el escenario 2

$$b_1(W, \tau_1) = \tau_1^1 + \tau_1^2 W_1 + \tau_1^3 W_2 + \tau_1^4 W_1 W_2,$$

y en el escenario 4

$$b_1(W, \tau_1) = \tau_1^1 + \tau_1^2 W_1 + \tau_1^3 W_2.$$

Como $\widehat{\alpha}_1$ resuelve la ecuación score en la regresión logística de X en el vector $(1, W_1, W_2, W_1 W_2)$, se verifica que

$$\mathbb{P}_n \left(\left(\begin{array}{c} 1 \\ W_1 \\ W_2 \\ W_1 W_2 \end{array} \right) (X - \text{expit}(\widehat{\alpha}_1^1 + \widehat{\alpha}_1^2 W_1 + \widehat{\alpha}_1^3 W_2 + \widehat{\alpha}_1^4 W_1 W_2)) \right) = 0,$$

o equivalentemente

$$\mathbb{P}_n \left(\left(\begin{array}{c} 1 \\ W_1 \\ W_2 \\ W_1 W_2 \end{array} \right) (X - E(X|W; \widehat{\alpha}_1)) \right) = 0.$$

En particular, tanto en el escenario 2 como en el 4 resulta que

$$\mathbb{P}_n (b_1(W, \widehat{\tau}_1) \{X - E(X|W; \widehat{\alpha}_1)\}) = 0.$$

como queríamos ver.

Capítulo 3

Supuestos y Definiciones

En este capítulo introducimos los funcionales sobre los cuales obtendremos resultados para su estimación múltiple protegida y damos un marco teórico para presentar los principales conceptos referidos a dicho procedimiento de estimación.

3.1. Definición del funcional de interés.

3.1.1. Supuestos.

Asumimos que las observaciones Z_1, Z_2, \dots son funciones medibles, definidas en un mismo espacio de probabilidad tomando valores en un espacio medible $(\mathcal{Z}, \mathcal{A})$. Supondremos que $(Z_i)_{i \geq 1}$ son independientes e idénticamente distribuidas y denotamos con Z a una observación genérica con la misma distribución que Z_1 . Consideramos la familia \mathcal{P} de medidas P de probabilidad en $(\mathcal{Z}, \mathcal{A})$ dominadas por una cierta medida σ -finita μ y mutuamente dominadas entre si. Llamamos a esta familia, el modelo no paramétrico. Asumimos que cada densidad $p = dP/d\mu$ correspondiente a una medida en \mathcal{P} se puede factorizar como el producto de $K + 1$ funciones medibles, g, h_1, \dots, h_K que varían independientemente, de modo que asociando, en un abuso de notación, a cada medida con su densidad, el conjunto \mathcal{P} se escribe como (donde de ahora en adelante identificamos a una medida de probabilidad con su densidad con respecto a μ)

$$\mathcal{P} \equiv \left\{ p = gh_1h_2\dots h_K : g \in \mathcal{U}^{np}, h_k \in \mathcal{V}_k^{np} \text{ para } k \in [K], p(z) \geq 0, \int pd\mu = 1 \right\}, \quad (3.1)$$

para ciertos conjuntos $\mathcal{U}^{np}, \mathcal{V}_k^{np}$, $k \in [K]$, donde denotamos con $[K]$ al conjunto $\{1, 2, \dots, K\}$.

En adelante, utilizaremos h para denotar al vector (h_1, \dots, h_K) y gh para hacer referencia a $p = gh = gh_1 \dots, h_K$.

La definición de los funcionales sobre los cuales construiremos estimadores múltiple protegidos depende de cierta función M . Consideremos para cada $h \in \otimes_{k=1}^K \mathcal{V}_k^{np}$ y para cada $\beta \in \mathbb{R}^r$ la función $m(-; \beta; h) : \mathcal{Z} \rightarrow \mathbb{R}^r$. Para enfatizar la dependencia en los parámetros, utilizaremos $M(\beta; h) \equiv m(Z; \beta; h)$. Sea $\mathcal{D} \subseteq \mathcal{P}$ un conjunto de distribuciones de forma tal que para todo $p = gh \in \mathcal{D}$ se tiene que

1. $E_{gh}[|M(\beta, h)|] < \infty$, para todo $\beta \in \mathbb{R}^r$.
2. Existe una única solución de la ecuación

$$E_{gh}[M(\beta, h)] = 0. \quad (3.2)$$

Utilizaremos $\beta(p)$ para denotar al funcional que a $p = gh \in \mathcal{D}$ le asigna el único elemento de \mathbb{R}^r que resuelve la ecuación (3.2). Más aún, vamos a suponer que existen \mathcal{U} y \mathcal{V}_k , para $k \in [K]$, subconjuntos de \mathcal{U}^{np} y \mathcal{V}_k^{np} , respectivamente, de forma tal que

$$\mathcal{M} \equiv \left\{ p = gh_1 h_2 \dots h_K : g \in \mathcal{U}, h_k \in \mathcal{V}_k \text{ para } k \in [K], p(z) \geq 0 \int p d\mu = 1 \right\} \quad (3.3)$$

está incluido en \mathcal{D} , y si $p = gh \in \mathcal{M}$, resulta que $\beta(p)$ depende en p solo mediante g . En otras palabras, se verifica que

$$\text{si } p = gh \text{ y } \tilde{p} = g\tilde{h} \text{ pertenecen a } \mathcal{M}, \text{ entonces } \beta(p) = \beta(\tilde{p}). \quad (3.4)$$

Tenemos entonces que el funcional $\beta(\cdot)$ está definido en el espacio \mathcal{D} y depende de $p = gh$ solo mediante g si $p \in \mathcal{M}$.

En el presente contexto nos proponemos estimar $\beta(p)$ suponiendo que la distribución de Z es un elemento de \mathcal{M} . Haciendo un pequeño abuso de notación pondremos $\beta(g)$ en lugar de $\beta(p)$ siempre que $p \in \mathcal{M}$.

3.1.2. Supuestos en los ejemplos.

Veremos ahora que en cada uno de los ejemplos del capítulo anterior se verifican los supuestos pedidos, identificando cada uno de los objetos involucrados.

Ejemplo 1: Datos Longitudinales En este caso tenemos que el vector Z está dado por $Z = (L_1, R_1, L_2, R_2, Y)$, cuya densidad p puede escribirse como $p = gh_1 h_2$ donde

$$\begin{aligned} g(z) &\equiv p_{L_1}(l_1) p_{L_2|R_1, L_1}(l_2|r_1, l_1) p_{Y|\bar{R}_2, \bar{L}_2}(y|r_2, \bar{l}_2), \\ h_1(z) &\equiv p_{R_1|L_1}(r_1|l_1), \\ h_2(z) &\equiv p_{R_2|R_1, \bar{L}_2}(r_2|r_1, \bar{l}_2). \end{aligned} \quad (3.5)$$

Tenemos entonces que $K = 2$, y podemos tomar \mathcal{U}^{np} siendo el modelo no paramétrico asociado a g (el conjunto de todos los valores posibles para una tal g) mientras que \mathcal{V}_j^{np} es el modelo no paramétrico asociado a cada distribución condicional h_j $j = 1, 2$. La función M está dada por

$$M(\beta, h_1, h_2) \equiv \frac{R_2 R_1}{E_{h_1}(R_1|L_1) E_{h_2}(R_2|R_1, \bar{L}_2)} (Y - \beta). \quad (3.6)$$

Podemos tomar $\mathcal{V}_j \subseteq \mathcal{V}_j^{np}$ de forma tal que cada uno de sus elementos se encuentre acotado superior e inferiormente por una constante positiva, o bien podemos requerir con el cociente entre dos elementos cualesquiera se encuentre acotado superiormente. En ambos casos, dada $g \in \mathcal{U}^{np}$, vale que $L_2(gh) = L_2(gh^*)$, para toda $h, h^* \in \otimes \mathcal{V}_j$. Mas generalmente, se tiene el siguiente resultado:

Lema 3.1.1. *Sean $g \in \mathcal{U}^{np}$, $h, h^* \in \otimes \mathcal{V}_j$ donde $\mathcal{V}_j \subseteq \mathcal{V}_j^{np}$, de forma tal que el cociente entre h y h^* está acotado superior e inferiormente por una constante positiva. Luego, existen $C_1 < C_2$ de forma tal que para toda función positiva $r(z)$ se verifica que*

$$C_1 \int_{\mathcal{Z}} r(z)g(z)h(z) d\mu(z) \leq \int_{\mathcal{Z}} r(z)g(z)h^*(z) d\mu(z) \leq C_2 \int_{\mathcal{Z}} r(z)g(z)h(z) d\mu(z).$$

Tomando \mathcal{U} como el conjunto de funciones $g \in \mathcal{U}^{np}$ para las cuales existe $h \in \otimes \mathcal{V}_j$ de forma tal que $E_{gh}(|Y|) < \infty$, del Lema 3.1.1, concluimos que \mathcal{M} definido por estos conjuntos satisface lo requerido.

Ejemplo 2: Regresión semiparamétrica con respuestas faltantes En este ejemplo el vector Z viene dado por $Z = (W, X, V, R, Y)$ y la densidad p de Z puede factorizarse como $p = gh_1h_2$ con

$$\begin{aligned} g(z) &\equiv p_{Y|R,W,X,V}(y|r, v, x, w) p_{V|X,W}(v|x, w) p_W(w), \\ h_1(z) &\equiv p_{X|W}(x|w), \\ h_2(z) &\equiv p_{R|V,X,W}(r|v, x, w), \end{aligned}$$

de donde, como en el ejemplo anterior, $K = 2$, \mathcal{U}^{np} es el modelo no paramétrico asociado a g mientras que \mathcal{V}_j^{np} es el modelo no paramétrico asociado a cada distribución condicional h_j $j = 1, 2$. En este caso la función M está dada por

$$M(\beta, h_1, h_2) \equiv \frac{R}{E_{h_2}(R|X, V, W)} (Y - \beta X) \{X - E_{h_1}(X|W)\} q(W),$$

donde $q(W)$ es una función arbitraria que supondremos acotada (en el ejemplo 2 del capítulo precedente utilizamos $q(W) = 1$).

Tal como hicimos en el ejemplo anterior, podemos elegir \mathcal{V}_j siendo un conjunto donde cada elemento se encuentra acotado superior e inferiormente por una constante positiva, o bien podemos requerir con el cociente entre dos elementos cualesquiera se encuentre acotado superiormente. Consideramos ahora \mathcal{A}_1 como el conjunto de funciones $g \in \mathcal{U}^{np}$ para las cuales existe $h \in \otimes \mathcal{V}_j$ de forma tal que $E_{gh}(|Y|) < \infty$. Apelando nuevamente a la Observación 3.1.1, podemos concluir que en $\mathcal{D} = \mathcal{A}_1 \times \otimes \mathcal{V}_j$ el funcional $\beta(\cdot)$ está bien definido.

Sin embargo, en este ejemplo $\mathcal{U} \subsetneq \mathcal{A}_1$, pues el modelo \mathcal{M} impone en g que exista $\beta(g)$ tal que

$$E_g(E_g(Y - \beta(g)X | R = 1, V, X, W) | X, W)$$

no depende de X , luego

$$\begin{aligned} \mathcal{U} \equiv \{g \in \mathcal{A}_1 : \text{existe } \beta(g) \in \mathbb{R}^r \text{ tal que} \\ E_g(E_g(Y - \beta(g)X | R = 1, V, X, W) | X, W) \text{ no depende de } X\}. \end{aligned} \quad (3.7)$$

Es decir, \mathcal{U} es el conjunto de funciones $g \in \mathcal{A}_1$ para las cuales

$$E_g(E_g(Y | R = 1, V, X, W) | X, W) = \beta(g)X + e_g(W).$$

3.2. Estimadores múltiple protegidos en modelos de intersección y unión.

En los ejemplos presentados en el capítulo anterior, para estimar a $\beta(g)$ asumiendo que $p = gh \in \mathcal{M}$ contemplamos diferentes estrategias de estimación. A partir de caracterizar a $\beta(g)$ como la solución de la ecuación dada en (3.2), una primera estrategia de estimación consiste en considerar un modelo paramétrico para h_k , para $k \in [K]$, de la forma

$$\mathcal{V}_{sub,k} \equiv \{h_{k,\alpha_k} \in \mathcal{V}_k : \alpha_k \in \Xi_k\}, \quad (3.8)$$

con $\Xi_k \subset \mathbb{R}^{\ell_k}$. Luego, estimamos a $\beta(g)$ mediante $\widehat{\beta}_{\mathcal{H}}$, solución de

$$\mathbb{P}_n \left(M(\beta, h_{k,\widehat{\alpha}_k}) \right) = 0, \quad (3.9)$$

donde, para $k \in [K]$, $\widehat{\alpha}_k$ denota al estimador de máxima verosimilitud bajo el modelo \mathcal{H}_k , con

$$\mathcal{H}_k \equiv \{p \in \mathcal{M} : h_k \in \mathcal{V}_{sub,k}\}. \quad (3.10)$$

Bajo condiciones de regularidad, el estimador $\widehat{\beta}_{\mathcal{H}}$ resulta consistente para $\beta(g)$ en el submodelo $\cap_{k=1}^K \mathcal{H}_k$. En los ejemplos 1 y 2, esta construcción da lugar a los estimadores

$\widehat{\beta}_{IPW}$ y $\widetilde{\beta}_2$ definidos en la Sección 2.1.2 y en la Sección 2.2.4, respectivamente.

En los ejemplos estudiados en el capítulo anterior presentamos también otro enfoque para estimar a $\beta(g)$, asumiendo modelos paramétricos para ciertos funcionales que dependen de g . Explícitamente, en el ejemplo 1 modelamos paraméricamente a los funcionales ν_1 y ν_2 dados por

$$\begin{aligned}\nu_1(g) &\equiv E_g(E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot) | R_1 = 1, L_1 = \cdot), \\ \nu_2(g) &\equiv E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot).\end{aligned}$$

Es decir, asumimos que

$$\nu_1(g) \in \mathcal{R}_{sub,1} = \{b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\} \text{ y } \nu_2(g) \in \mathcal{R}_{sub,2} = \{b_2(\cdot, \cdot; \tau_2) : \tau_2 \in \Upsilon_2\},$$

para ciertas funciones $b_1(L_1; \cdot)$ y $b_2(\overline{L}_2; \cdot)$ conocidas, indexadas por parámetros finito dimensionales pertenecientes a conjuntos $\Upsilon_j \subset \mathbb{R}^{r_j}$, para $j = 1, 2$. A partir de asumir como verdaderos estos dos modelos construimos el estimador $\widehat{\beta}_{reg}$, definido en la Sección 2.1.1 del capítulo anterior.

En el ejemplo 2 modelamos paraméricamente los funcionales

$$\begin{aligned}\nu_1(g) &\equiv E_g(E_g(Y - \beta(g)X | R = 1, V = \cdot, X = \cdot, W = \cdot) | X = \cdot, W = \cdot), \quad \text{y} \\ \nu_2(g) &\equiv E_g(Y | R = 1, X = \cdot, V = \cdot, W = \cdot),\end{aligned}$$

asumiendo que

$$\nu_1(g) \in \mathcal{R}_{sub,1} \equiv \{b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\} \text{ y } \nu_2(g) \in \mathcal{R}_{sub,2} \equiv \{b_2(\cdot, \cdot; \tau_2) : \tau_2 \in \Upsilon_2\},$$

donde $b_1(W; \cdot)$ es una función suave conocida salvo por el parámetro $\tau_1 \in \Upsilon_1 \subseteq \mathbb{R}^{r_1}$, y $b_2(W, X, V; \cdot)$ es una función suave conocida salvo por $\tau_2 \in \Upsilon_2 \subseteq \mathbb{R}^{r_2}$. Luego, en la Sección 2.2.3 del capítulo anterior, construimos el estimador $\widehat{\beta}_1$, resolviendo el sistema presentado en (2.51).

Para poder generalizar estas construcciones, para $k \in [K]$, consideraremos cierto funcional

$$\rho_k : \mathcal{U} \rightarrow \mathcal{R}_k \tag{3.11}$$

y asumiremos un modelo paramétrico para $\rho_k(g)$. Es decir, supondremos que $\rho_k(g) \in \mathcal{R}_{sub,k}$ con

$$\mathcal{R}_{sub,k} \equiv \{\rho_{k, \tau_k} \in \mathcal{R}_k^{ext} : \tau_k \in \Upsilon_k\}, \tag{3.12}$$

donde $\Upsilon_k \subseteq \mathbb{R}^{r_k}$ y $\mathcal{R}_k \subset \mathcal{R}_k^{ext}$. A partir de estos modelos, en cada uno de los ejemplos ($K = 2$) se construyó un estimador $\widehat{\beta}_{\mathcal{G}}$ que resultó consistente para $\beta(g)$ bajo cualquier ley $p \in \cap_{k=1}^K \mathcal{G}_k$ donde, para cada $k \in [K]$,

$$\mathcal{G}_k \equiv \{p \in \mathcal{M} : \rho_k(g) \in \mathcal{R}_{sub,k}\}. \tag{3.13}$$

En el ejemplo 1, $\widehat{\beta}_{\mathcal{G}} = \widehat{\beta}_{reg}$, mientras que

$$\rho_1(g) \equiv \nu_1(g) = E_g(E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot) | R_1 = 1, L_1 = \cdot),$$

$$\rho_2(g) \equiv \nu_2(g) = E_g(Y | R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot).$$

En el ejemplo 2 teníamos definidos los funcionales

$$\nu_1(g) \equiv E_g(E_g(Y - \beta(g)X | R = 1, V = \cdot, X = \cdot, W = \cdot) | X = \cdot, W = \cdot) \text{ y}$$

$$\nu_2(g) \equiv E_g(Y | R = 1, X = \cdot, V = \cdot, W = \cdot),$$

luego $\widehat{\beta}_{\mathcal{G}} = \widehat{\beta}_1$ y

$$\rho_1(g) \equiv \nu_1(g) = E_g(E_g(Y - \beta(g)X | R = 1, V = \cdot, X = \cdot, W = \cdot) | X = \cdot, W = \cdot),$$

$$\rho_2(g) \equiv \nu_2(g) = E_g(Y | R = 1, X = \cdot, V = \cdot, W = \cdot).$$

Finalmente en cada ejemplo construimos un estimador múltiple protegido para $\beta(g)$, denotado $\widehat{\beta}_{MP}$, que confiere protección con respecto a apartamientos de los modelos $\cap_{k=1}^2 \mathcal{G}_k$ y $\cap_{k=1}^2 \mathcal{H}_k$ en el siguiente sentido: $\widehat{\beta}_{MP}$ resultó consistente para $\beta(g)$ bajo cualquier ley p en el *modelo de intersección unión* dado por $\cap_{k=1}^2 \{\mathcal{H}_k \cup \mathcal{G}_k\}$.

Al ser $\beta(g)$ la solución de (3.2), es inmediato construir estimadores consistentes para nuestro funcional imponiendo restricciones sobre cada h_k , obteniéndose así el estimador $\widehat{\beta}_{\mathcal{H}}$ definido mediante la ecuación (3.9). Por el contrario, la existencia de estimadores $\widehat{\beta}_{\mathcal{G}}$ que resulten consistentes al considerar modelos para $\rho_k(g)$ no es una consecuencia directa de la definición de $\beta(g)$. Menos aún es la existencia de un estimador $\widehat{\beta}_{MP}$ consistente para $\beta(g)$ bajo cualquier ley $p \in \cap_{k=1}^K \{\mathcal{H}_k \cup \mathcal{G}_k\}$. Uno de los propósitos de este trabajo es formular una teoría que explique la existencia de funciones de estimación insesgadas múltiple protegidas, como las presentadas mediante las ecuaciones (2.18) y (2.61), que permitieron construir los estimadores múltiple protegidos presentados en los ejemplos del capítulo anterior. Veremos que cuando los factores h_k presentan cierta estructura, bajo los supuestos establecidos en la Sección 3.1.1, es posible identificar ciertos funcionales ρ_k y proponer modelos $\mathcal{R}_{sub,k}$ que permitan construir estimadores que se espera resulten consistentes bajo cualquier ley $p \in \cap_{k=1}^K \{\mathcal{H}_k \cup \mathcal{G}_k\}$. Con ese fin, proporcionamos la definición de lo que constituye un estimador múltiple protegido en el modelo de intersección unión. La definición asume que, para cada $k \in [K]$, el analista ya identificó un funcional ρ_k como en (3.11) con imagen en \mathcal{R}_k y consideró modelos paramétricos como en (3.12), dando lugar a los modelos \mathcal{G}_k definidos en (3.13). También presupone que, para cada $k \in [K]$, el analista seleccionó una clase paramétrica $\mathcal{V}_{sub,k}$ para h_k como en (3.8) y consideró los submodelos \mathcal{H}_k definidos como en (3.10). Más adelante discutiremos como construir las funcionales ρ_k adecuadas.

Aclaración: En la formulación de la clase paramétrica $\mathcal{R}_{sub,k}$ nos permitimos la posibilidad de que \mathcal{R}_k^{ext} incluya estrictamente a \mathcal{R}_k . Esto es importante pues queremos que la teoría a desarrollar contemple situaciones en las que uno o varios de los modelos $\mathcal{R}_{sub,k}$ son erróneos, es decir $\mathcal{R}_{sub,k} \cap \mathcal{R}_k = \emptyset$ y por consiguientes \mathcal{G}_k no tiene elementos. Cuando \mathcal{G}_k es vacío decimos que $\mathcal{R}_{sub,k}$ es incompatible con \mathcal{M} . Si pidieramos que \mathcal{R}_k sea igual a \mathcal{R}_k^{ext} entonces no podíamos contemplar la posibilidad de que \mathcal{G}_k sea el conjunto vacío. Por supuesto, sería ilógico que el analista de datos formulara intencionalmente modelos incompatibles. Eso es algo que ocurre debido a la complejidad del problema de modelización. Tal es el caso en la construcción de estimadores doble protegidos en modelos estructurales marginales (Robins (2000), Bang & Robins (2005)). Para fijar el concepto de incompatibilidad, analicemos lo que ocurre en el Ejemplo 2. En dicho escenario si tomamos como modelo para $\rho_2(g) = E_g(Y|R=1, X=\cdot, V=\cdot, W=\cdot)$ a

$$\mathcal{R}_{sub,2} = \{b_2(W, X, V, \tau_2) = \tau_2 W X : \tau_2 \in \mathbb{R}\},$$

como para $p = gh \in \mathcal{M}$ resulta que

$$E_g(E_g(Y|R=1, V, X, W) | X, W) = \beta(g) X + e_g(W),$$

tenemos que \mathcal{G}_2 es vacío.

Definición 3.3. *Un estimador $\widehat{\beta}_{MP}$ se dice múltiple protegido consistente para $\beta(\cdot)$ en el modelo de intersección unión dado por $\cap_{k=1}^K (\mathcal{H}_k \cup \mathcal{G}_k)$ si $\widehat{\beta}_{MP} \xrightarrow{P} \beta(g)$ bajo cualquier ley $p = gh_1 \dots h_K \in \mathcal{M}$ verificando: para cada $k \in [K]$ $h_k \in \mathcal{V}_{sub,k}$ o $\rho_k(g) \in \mathcal{R}_{sub,k}$.*

Es decir, un estimador $\widehat{\beta}_{MP}$ es múltiple protegido consistente para $\beta(\cdot)$ si para cada $c \subseteq [K]$, $\widehat{\beta}_{MP}$ es consistente para $\beta(g)$ bajo el modelo

$$\mathcal{M}_c \equiv \{p = gh_1 \dots h_K \in \mathcal{M} : p \in \mathcal{H}_k \text{ para } k \in c \text{ y } p \in \mathcal{G}_k \text{ para } k \notin c\}. \quad (3.14)$$

En adelante, para hacer más fluida la notación, usaremos solo múltiple protegido para denotar a un estimador múltiple protegido consistente.

En palabras, $\widehat{\beta}_{MP}$ es múltiple protegido para $\beta(g)$ en el modelo $\cap_{k=1}^K (\mathcal{H}_k \cup \mathcal{G}_k)$ si converge en probabilidad a $\beta(g)$ cuando para cada k o bien se verifica \mathcal{H}_k o se verifica \mathcal{G}_k . Equivalentemente, un estimador resulta múltiple protegido si es consistente para $\beta(g)$ cuando se verifica alguno de los submodelos \mathcal{M}_c , $c \subseteq [K]$. Si ningún \mathcal{M}_c está incluido en otro $\mathcal{M}_{c'}$ y todos los modelos $\mathcal{R}_{sub,k}$ con $1 \leq k \leq K$ son compatibles, un estimador múltiple protegido en $\cap_{k=1}^K (\mathcal{H}_k \cup \mathcal{G}_k)$ confiere 2^K opciones de modelado bajo las cuales se tiene garantizada la consistencia a $\beta(g)$. Por ese motivo se lo llama 2^K múltiple protegido. Si algún \mathcal{M}_c es vacío, debido a que para algún $k \in \bar{c}$ $\mathcal{G}_k = \emptyset$ o debido a que $\mathcal{G}_k \cap \mathcal{G}_{k'} = \emptyset$, $k, k' \in \bar{c}$, en ese caso el estimador múltiple protegido confiere menos de 2^k escenarios de consistencia.

En el Ejemplo 1 tanto \mathcal{G}_1 como \mathcal{G}_2 son distintos de vacío debido a que \mathcal{U} es el modelo no paramétrico asociado a g . Supongamos que $\rho_{2,\tau_2} = b_2(\bar{L}_2, \tau_2)$ y que $\rho_{1,\tau_1} = b_1(L_1, \tau_1)$, debido a que $E(b_2(\bar{L}_2, \tau_2)|R_1 = 1, L_1) = b_1(L_1, \tau_1)$, puede que $\mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$. En efecto, si consideramos un escenario en donde L_1 es una variable con rango distinto a $\{0, 1\}$ basta tomar por ejemplo $b_2(\bar{L}_2, \tau_2) = \tau_2 L_1^2 : \tau_2 > 0$ y $b_1(L_1, \tau_1) = \tau_1 L_1 : \tau_1 > 0$. Una condición que garantiza que la intersección de dichos conjuntos sea no vacía es que $\mathcal{R}_{sub,1}$ y $\mathcal{R}_{sub,2}$ incluyan a las funciones constantes.

En el Ejemplo 2, \mathcal{U} es un modelo semiparamétrico, \mathcal{G}_1 es no vacío y, como vimos anteriormente, \mathcal{G}_2 puede ser vacío. De hecho \mathcal{G}_1 es no vacío debido a que coincide con el modelo semiparamétrico formado por todas las distribuciones $p = gh \in \mathcal{M}$ que se corresponden con una ley p^f en el sentido de la proposición 2.2.1, tal que

$$E_{p^f}(Y^f|X, W) = \beta(p)X + b_1(W, \tau_1).$$

En este ejemplo también puede ocurrir que $\mathcal{G}_1 \cap \mathcal{G}_2$ sea el conjunto vacío sin que \mathcal{G}_2 lo sea. En efecto, si por ejemplo tomamos $\rho_{2,\tau_2} = b_2(W, X, V, \tau_2)$ y $\rho_{1,\tau_1} = b_1(W, \tau_1)$, con $b_2(W, X, V, \tau_2) = 3X + \tau_2 W^2 : \tau_2 > 0$ y $b_1(W, \tau_1) = \tau_1 W : \tau_1 > 0$, donde W es una variable continua, debido a que $E(b_2(W, X, V, \tau_2)|W, X) = \beta(g)X + b_1(W, \tau_1)$, resulta que $\mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$.

En los ejemplos, salvo que $\mathcal{G}_1 \cap \mathcal{G}_2$ o que algún \mathcal{G}_i sea vacío, tenemos $2^2 = 4$ modelos, siendo que hay 4 opciones para elegir al conjunto c , dadas por $c = \emptyset, \{1\}, \{2\}, \{1, 2\}$. Cada conjunto c da origen a un submodelo \mathcal{M}_c diferente, a saber:

$$\begin{aligned} \mathcal{M}_\emptyset &= \{p = gh_1h_2 \in \mathcal{M} : p \in \{\mathcal{G}_1 \cap \mathcal{G}_2\}\}, \\ \mathcal{M}_{\{1\}} &= \{p = gh_1h_2 \in \mathcal{M} : p \in \{\mathcal{H}_1 \cap \mathcal{G}_2\}\}, \\ \mathcal{M}_{\{2\}} &= \{p = gh_1h_2 \in \mathcal{M} : p \in \{\mathcal{G}_1 \cap \mathcal{H}_2\}\}, \\ \mathcal{M}_{\{1,2\}} &= \{p = gh_1h_2 \in \mathcal{M} : p \in \{\mathcal{H}_1 \cap \mathcal{H}_2\}\}, \end{aligned}$$

En los ejemplos presentados, la construcción de cada estimador múltiple protegido se hizo usando una función de estimación con ciertas propiedades. A continuación definimos el concepto de función de estimación múltiple protegida.

3.3.1. Función de estimación insesgada múltiple protegida en modelos de unión intersección.

En cada uno de los ejemplos presentados en el capítulo anterior, hemos introducido una función $U(\beta, \tau, \alpha)$ mediante la cual construimos estimadores múltiple protegidos. Dicha función satisface las condiciones necesarias para ser una función de estimación múltiple protegida insesgada para $\beta(\cdot)$, concepto que definimos a continuación,

luego de establecer algunas cuestiones notacionales. En adelante, si γ denota el vector $(\gamma_1, \dots, \gamma_K)$ y c es un subconjunto de $[K]$, utilizaremos γ_c para referirnos a las coordenadas de γ con índice en c : $\gamma_c \equiv \left(\gamma_j \right)_{j \in c}$. Más aún, si $c = [K]$ eliminamos los subíndices, es decir, pondremos $\gamma \equiv \gamma_{[K]}$. Por otra parte, si $p = gh \in \mathcal{M}_c$, definimos $\alpha_c(h_c) \equiv \left(\alpha_j(h_j) \right)_{j \in c}$ y $\tau_{\bar{c}}(g) \equiv \left(\tau_j(g) \right)_{j \in \bar{c}}$, donde $\alpha_j(h_j)$ y $\tau_j(g)$ son los valores para los cuales se verifica que

$$h_j = h_{j, \alpha_j(h_j)}, \text{ para } j \in c \text{ y } \rho_j(g) = \rho_{j, \tau_j(g)} \text{ para } j \in \bar{c}. \quad (3.15)$$

Introducir el concepto de función de estimación múltiple protegida asume, como en la Definición 3.3, que para cada $k \in [K]$ contamos con un funcional ρ_k , como en (3.11), con imagen en \mathcal{R}_k y consideramos modelos paramétricos $\mathcal{R}_{sub,k}$, como en (3.12), a partir de los cuales se proponen los modelos \mathcal{G}_k , definidos en (3.13). También contamos con una clase paramétrica $\mathcal{V}_{sub,k}$ para h_k como en (3.8) y consideramos los submodelos \mathcal{H}_k definidos en (3.10).

Definición 3.4. Una función $U(\beta, \tau, \alpha) = u(Z, \beta, \tau, \alpha)$, siendo

$$u(Z, \cdot, \cdot, \cdot) : \mathbb{R}^r \times \otimes_{k=1}^K \Upsilon_k \times \otimes_{k=1}^K \Xi_k \rightarrow \mathbb{R}^r,$$

se dice función de estimación insesgada múltiple protegida para $\beta(\cdot)$ en el modelo $\cap_{k=1}^K (\mathcal{H}_k \cup \mathcal{G}_k)$ si para cada $c \subseteq [K]$, cada $p = gh \in \mathcal{M}_c$, para cada $\tau_c \in \otimes_{k \in c} \Upsilon_k$ y $\alpha_{\bar{c}} \in \otimes_{k \in \bar{c}} \Xi_k$, se verifica que

$$E_p \left(U \left(\beta, \tau, \alpha \right) \right) = 0 \quad \text{en } \beta = \beta(g), \tau_{\bar{c}} = \tau_{\bar{c}}(g) \text{ y } \alpha_c = \alpha_c(h_c).$$

En palabras, U es una función de estimación insesgada múltiple protegida para $\beta(\cdot)$ si bajo $p = gh$, h_K tiene esperanza cero en $\beta = \beta(g)$ siempre que para cada $k \in [K]$ o bien $h_k \in \mathcal{V}_{sub,k}$ o $\rho_k(g) \in \mathcal{R}_{sub,k}$ y U esté evaluada en los valores verdaderos de los parámetros que indexan los modelos correctos.

El estimador $\hat{\beta}_{MP}$ presentado en ejemplo 1, en la Sección 2.1.4, se obtuvo utilizando la función de estimación múltiple protegida insesgada para $\beta(\cdot)$ dada por

$$\begin{aligned} U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) &= \frac{R_2 R_1}{E(R_1 | L_1; \alpha_1) E(R_2 | R_1, \bar{L}_2; \alpha_2)} \{Y - b_2(\bar{L}_2; \tau_2)\} \\ &+ \frac{R_1}{E(R_1 | L_1; \alpha_1)} \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} + \{b_1(L_1; \tau_1) - \beta\}, \end{aligned}$$

mientras que la propuesta múltiple protegida introducida en el ejemplo 2 utiliza

$$U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) = H(\beta, \tau_1, \tau_2, \alpha_2) \{X - E(X|W; \alpha_1)\}$$

con

$$H(\beta, \tau_1, \tau_2, \alpha_2) = \frac{R\{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1).$$

En adelante usaremos indistintamente la expresión función de estimación insesgada múltiple protegida o la expresión función de estimación múltiple protegida.

3.4.1. Función de estimación insesgada múltiple protegida para los parámetros de ruido en modelos de unión intersección.

Supongamos que contamos con una función de estimación $U(\beta, \tau, \alpha)$ múltiple protegida para $\beta(\cdot)$ y que estamos bajos los supuestos enunciados en la Sección 3.1.1, recordando que $p = gh$ denota la distribución de los datos observados. Asumamos también que contamos con $(\hat{\alpha}_{MP}, \hat{\tau}_{MP})$ verificando que $(\hat{\alpha}_{MP}, \hat{\tau}_{MP}) \xrightarrow{P} (\alpha^*(p), \tau^*(p))$, de forma tal que

$$\text{si } p \in \mathcal{M}_c \text{ entonces, } \alpha_c^*(p) = \alpha_c(h_c) \quad \text{y} \quad \tau_c^*(g) = \tau_c(g). \quad (3.16)$$

Bajo las presentes circunstancias, resolviendo en β la ecuación

$$\mathbb{P}_n(U(\beta, \hat{\tau}_{MP}, \hat{\alpha}_{MP})) = 0, \quad (3.17)$$

es de esperar que obtengamos un estimador múltiple protegido para $\beta(g)$. Luego, si contamos con una función de estimación insesgada múltiple protegida $U(\beta, \tau, \alpha)$, para estimar en forma múltiple protegida a $\beta(g)$ basta presentar estimadores $(\hat{\alpha}_{MP}, \hat{\tau}_{MP})$ de forma tal que los valores límite satisfagan la condición (3.16). Cuando los modelos para h_k son paramétricos, los estimadores de máxima verosimilitud bajo los modelos \mathcal{H}_k generalmente cumplirán con el requisito deseado. Es decir, bajo condiciones de regularidad, el estimador de máxima verosimilitud $\hat{\alpha}_k$ puede ser definido como la solución de $\mathbb{P}_n(S_k(\alpha_k)) = 0$, con $S_k(\alpha_k) = \partial \log h_{k, \alpha_k} / \partial \alpha_k$. En tal caso, $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_K)$ converge a $(\alpha_1^*, \dots, \alpha_K^*)$, verificando la primer condición requerida en (3.16). Con el parámetro τ la situación es diferente. Para poder construir un estimador $\hat{\tau}_{MP}$ cuyo límite satisfaga la segunda condición enunciada en (3.16), es necesario poder estimar dichos parámetros usando exclusivamente los modelos postulados en \mathcal{H}_k y \mathcal{G}_k , para $k \in [K]$. En los ejemplos presentados se construyeron estimadores $\hat{\tau}$ para $\tau = (\tau_1, \tau_2)$ verificando (3.16). Es decir, para cada $c \subseteq [2]$, $\hat{\tau}_c$ resultó consistente para $\tau_c(g)$ bajo \mathcal{M}_c . El estimador $\hat{\tau}$ se obtuvo resolviendo conjuntamente en (β, τ, α) , el sistema de ecuaciones

$$\mathbb{P}_n(G(\beta, \tau, \alpha)) = 0 \quad (3.18)$$

donde $G = (U', T', S)'$ con $S \equiv (S_1, \dots, S_K)$, $S_k(\alpha_k) = \partial \log h_{k, \alpha_k} / \partial \alpha_k$, $T \equiv (T_1, \dots, T_K)$ y $K = 2$.

★ En el ejemplo 1

$$h_{1, \alpha_1}(l_1) = P(R_1 = 1 | L_1 = l_1, \alpha_1) .$$

$$h_{2, \alpha_2}(\bar{l}_2) = P(R_2 = 1 | \bar{L}_2 = \bar{l}_2, R_1 = 1, \alpha_2) .$$

$$S_1(\alpha_1) = \partial \log h_{1, \alpha_1}(L_1) / \partial \alpha_1,$$

$$S_2(\alpha_2) = \partial \log h_{2, \alpha_2}(\bar{L}_2) / \partial \alpha_2,$$

$$\begin{aligned} U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) &= \frac{R_2 R_1}{E(R_1 | L_1; \alpha_1) E(R_2 | R_1, \bar{L}_2; \alpha_2)} \{Y - b_2(\bar{L}_2; \tau_2)\} \\ &+ \frac{R_1}{E(R_1 | L_1; \alpha_1)} \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} + \{b_1(L_1; \tau_1) - \beta\}, \end{aligned}$$

$$T_1(\beta, \tau, \alpha) = R_1 d(L_1, \tau_1, \tau_2, \alpha_2) \left(\frac{R_2 \{Y - b_2(\bar{L}_2; \tau_2)\}}{E(R_2 | R_1, \bar{L}_2; \alpha_2)} + \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} \right),$$

$$T_2(\tau_2) = R_2 \frac{\partial b_2(\bar{L}_2, \tau_2)}{\partial \tau_2} (Y - b_2(L_1, L_2, \tau_2))$$

con $d(L_1, \tau_1, \tau_2, \alpha_2)$ una función vectorial arbitraria de la misma dimensión que τ_1 .

★ En el ejemplo 2

$$h_{1, \alpha_1}(w) = P(X = 1 | W = w, \alpha_1) .$$

$$h_{2, \alpha_2}(w, x, v) = P(R = 1 | W = w, X = x, V = v, \alpha_2) .$$

$$S_1(\alpha_1) = \partial \log h_{1,\alpha_1}(W) / \partial \alpha_1,$$

$$S_2(\alpha_2) = \partial \log h_{2,\alpha_2}(W, X, V) / \partial \alpha_2,$$

$$U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) \equiv H(\beta, \tau_1, \tau_2, \alpha_2) \{X - E(X|W; \alpha_1)\}$$

$$T_1(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) = H(\beta, \tau_1, \tau_2, \alpha_2) \{X - E(X|W; \alpha_1)\} d(X, W, \beta, \tau_1, \tau_2, \alpha_1, \alpha_2),$$

con $d(X, W, \beta, \tau_1, \tau_2, \alpha_1, \alpha_2)$ una función vectorial con la misma dimensión que τ_1 y

$$H(\beta, \tau_1, \tau_2, \alpha_2) = \frac{R \{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1),$$

mientras que

$$T_2(\tau_2) = R \frac{\partial b_2(V, X, W; \tau_2)}{\partial \tau_2} (Y - b_2(W, X, V; \tau_2)).$$

La función de estimación G presentada en cada uno de los ejemplos satisface que, para cada $c \subseteq [K]$, $\tau_c \in \otimes_{k \in c} \Upsilon_k$ y para cada $\alpha_{\bar{c}} \in \otimes_{k \in \bar{c}} \Xi_k$,

$$E_p(G_c(\beta, \tau, \alpha)) = 0 \text{ en } \beta = \beta(g), \tau_{\bar{c}} = \tau_{\bar{c}}(g) \text{ y } \alpha_c = \alpha_c(h_c), \quad (3.19)$$

para toda $p = gh_1 \dots h_K \in \mathcal{M}_c$, donde $G_c = (U', T'_{\bar{c}}, S'_c)$ con $T'_{\bar{c}} \equiv \{T_k : k \in \bar{c}\}$ y $S'_c = \{S_j : j \in c\}$.

Para poder generalizar el procedimiento presentado en los ejemplos, dado $c \subseteq [K]$, escribamos $\tau = (\tau_c, \tau_{\bar{c}})$ y $\alpha = (\alpha_c, \alpha_{\bar{c}})$. Supongamos que fijados $\tau_c \in \otimes_{k \in c} \Upsilon_k$ y $\alpha_{\bar{c}} \in \otimes_{k \in \bar{c}} \Xi_k$, la ecuación (en la variables $(\beta, \tau_{\bar{c}}, \alpha_c)$)

$$E_p(G_c(\beta, \tau_c, \tau_{\bar{c}}, \alpha_c, \alpha_{\bar{c}})) = 0,$$

tenga por única solución al vector $(\beta(g), \tau_{\bar{c}}(g), \alpha_c(h_c))$ cuando $p \in \mathcal{M}_c$. Supongamos además que si $p \in \mathcal{M}_c$ la ecuación

$$E_p\{G(\beta, \tau, \alpha)\} = 0$$

tiene una única solución, que denotaremos con $(\beta^\dagger(p), \tau^\dagger(p), \alpha^\dagger(p))$. Luego, tenemos que $(\beta^\dagger(p), \tau_{\bar{c}}^\dagger(p), \alpha_c^\dagger(p))$ debe ser igual a $(\beta(g), \tau_{\bar{c}}(g), \alpha_c(h_c))$. En tal caso, bajo

condiciones de regularidad, es de esperar que el vector $(\widehat{\beta}_{MR}, \widehat{\tau}, \widehat{\alpha})$ que resuelve (3.18) sea un estimador consistente para $(\beta^\dagger(p), \tau^\dagger(p), \alpha^\dagger(p))$ y en consecuencia $\widehat{\beta}_{MR}$ convergerá a $\beta(g)$ bajo el modelo \mathcal{M}_c .

Es decir, la clave para proponer un estimador múltiple protegido para $\beta(g)$ a partir de una función de estimación múltiple protegida, radica en poder estimar los parámetros de ruido de forma múltiple protegida. Para ello, queremos construir funciones $(T_j)_{j \in [K]}$ verificando (3.19). Una colección de funciones $T \equiv (T'_j)_{j \in [K]}$ verificando (3.19) se llama función de estimación insesgada múltiple protegida de ruido. Damos a continuación la definición formal de este objeto.

Definición 3.5. *Dada una función de estimación $U(\beta, \tau, \alpha)$ insesgada múltiple protegida para $\beta(\cdot)$ en el modelo $\cap_{k=1}^K (\mathcal{H}_k \cup \mathcal{G}_k)$, diremos que una aplicación $T(\beta, \tau, \alpha) = t(Z, \beta, \tau, \alpha)$, con*

$$t(Z, \cdot, \cdot, \cdot) : \mathbb{R}^r \times \otimes_{k=1}^K \Upsilon_k \times \otimes_{k=1}^K \Xi_k \rightarrow \mathbb{R}^{r^*},$$

donde $r^* = \sum_{k=1}^K \dim(\tau_k)$, es una función de estimación insesgada múltiple protegida de ruidos en el modelo $\cap_{k=1}^K (\mathcal{H}_k \cup \mathcal{G}_k)$ si para cada $c \subseteq [K]$, para cada $p \in \mathcal{M}_c$, para cada $\tau_c \in \otimes_{k \in c} \Upsilon_k$ y $\alpha_{\bar{c}} \in \otimes_{k \in \bar{c}} \Xi_k$ se verifica que

$$E_p \left(T_{\bar{c}}(\beta, \tau, \alpha) \right) = 0 \quad \text{en } \beta = \beta(g), \tau_{\bar{c}} = \tau_{\bar{c}}(g) \text{ y } \alpha_c = \alpha_c(h_c).$$

Capítulo 4

Construcción de funciones de estimación insesgadas múltiple protegidas en modelos de unión-intersección.

El principal objetivo de este capítulo es presentar una forma de construir funciones de estimación insesgadas múltiple protegidas. La propuesta surge de generalizar un resultado presentado en Robins et al. (2000) y Robins et al (2001). El mismo requiere que ciertos submodelos del modelo no paramétrico \mathcal{P} sean convexos. Comenzaremos entonces definiendo modelos convexos. Veremos que cuando los factores h_k son determinadas densidades condicionales, como ocurre en los ejemplos estudiados, podemos garantizar las condiciones requeridas para generalizar el resultado de Robins et al. (2000) y Robins et al (2001). Luego, enunciaremos una serie de lemas que serán claves para la construcción de funciones de estimación insesgadas múltiple protegidas. Las demostraciones de los mismos están íntimamente relacionadas con algunos conceptos relativos a la estadística semiparamétrica. Finalmente, en la última sección de este capítulo presentamos las funciones buscadas.

Al igual que en capítulos anteriores, con el objeto de hacer más fluida la lectura de este trabajo, presentamos las demostraciones en el Apéndice A y en el Apéndice B.

4.1. Modelos convexos y condiciones.

Los modelos convexos son introducidos por Bickel (1982) durante el desarrollo de procedimientos de estimación adaptativos. Estos procedimientos se usan para estimar

la parte paramétrica, denotada por β , en modelos semiparamétricos como los definidos en el Apéndice A. Podemos pensar que estos métodos son capaces de estimar a β tan bien como si el parámetro de ruido fuera conocido. Bickel enfatiza el rol fundamental que tiene la convexidad del modelo en facilitar la estimación adaptiva.

Los modelos convexos también son usados en la construcción de estimadores eficientes en modelos semiparamétricos. Específicamente, en Bickel J, Wellner J, Klaassen C, Ritov Y, (1993), en el Capítulo 7, aparecen los modelos convexos a la hora de construir M- estimadores basados en scores eficientes.

Estos modelos aparecen también en los procedimientos de estimación doble protegidos. van der Laan M y Robins J (2003), en el Lema 1.8 construyen una función mediante la cual se puede definir estimadores doble protegidos. Una de las hipótesis requeridas para dicha construcción es la de convexidad.

Un modelo $\widetilde{\mathcal{M}}$, se dice convexo, si dados dos elementos p y p^* en él, toda combinación convexa entre ellos también pertenece al modelo: $\{p, p^*\} \subset \widetilde{\mathcal{M}} \Rightarrow \theta p + (1 - \theta)p^* \in \widetilde{\mathcal{M}}$ para cada $\theta \in [0, 1]$.

Recordamos que

$$\mathcal{P} \equiv \left\{ p = gh_1 h_2 \dots h_K : g \in \mathcal{U}^{np}, h_k \in \mathcal{V}_k^{np}, p(z) \geq 0 \text{ para todo } z \text{ y } \int p d\mu = 1 \right\}.$$

Fijado un subconjunto $c \subseteq [K]$, funciones $g \in \mathcal{U}^{np}$ y $h_c \in \otimes_{j \in c} \mathcal{V}_j^{np}$, definimos el submodelo de \mathcal{P}

$$\mathcal{P}_{g, h_c} \equiv \{p^* = gh^* \in \mathcal{P} \text{ tales que } h_k^* = h_k \text{ para } k \in c\}. \quad (4.1)$$

A lo largo del presente capítulo requeriremos la convexidad de los submodelos \mathcal{P}_{g, h_c} . Esta condición se verifica en cada uno de los ejemplos presentados y es consecuencia del siguiente hecho.

Condición 1: Factorización en Densidades Condicionales. El vector Z puede representarse como $Z = (L_1, A_1, L_2, A_2, \dots, A_K, L_{K+1})$ y para $k \in [K]$

$$h_k(z) = p_{A_k | \bar{L}_k, \bar{A}_{k-1}}(a_k | \bar{l}_k, \bar{a}_{k-1}),$$

donde \bar{A}_0 y \bar{a}_0 indican los vectores vacío.

Para facilitar la notación, para $1 \leq j \leq K$, de ahora en mas pondremos

$$(L_1, A_1, L_2, A_2, \dots, A_j, L_{j+1}) = (\bar{A}_j, \bar{L}_{j+1}),$$

en cuyo caso tenemos que $Z = (\bar{A}_K, \bar{L}_{K+1})$. Presentamos en el Apéndice B la demostración del siguiente resultado.

Lema 4.1.1. *Si se verifica la Factorización en Densidades Condicionales, entonces para todo $c \subsetneq [K]$, $g \in \mathcal{U}^{np}$ y $h_c \in \otimes_{j \in c} \mathcal{V}_j^{np}$, el submodelo \mathcal{P}_{g, h_c} es convexo.*

Veremos ahora que los ejemplos presentados en el primer capítulo satisfacen la Condición 1. De hecho, en el ejemplo 1 tenemos $K = 2$ y $Z = (L_1, R_1, L_2, R_2, Y)$. Luego, definiendo $A_i = R_i$ para $i = 1, 2$ y $L_3 = Y$, tenemos

$$\begin{aligned} h_1(z) &\equiv p_{R_1|L_1}(r_1|l_1), \\ h_2(z) &\equiv p_{R_2|R_1, \bar{L}_2}(r_2|r_1, \bar{l}_2). \end{aligned}$$

En el ejemplo 2 tenemos $K = 2$ y $Z = (W, X, V, R, Y)$; luego $L_1 = W$, $A_1 = X$, $L_2 = V$, $A_2 = R$, $L_3 = Y$ y

$$h_1(z) = p_{X|W}(x|w), \quad h_2(z) = p_{R|V, X, W}(r|v, x, w).$$

4.2. Resultados previos.

En esta sección presentamos algunos resultados preliminares, necesarios para construir funciones de estimación insesgadas múltiple protegidas. Los mismos requieren conceptos pertenecientes a la estadística semiparamétrica y es por ello que decidimos incluir en el Apéndice A una breve introducción a dicha disciplina, presentando las nociones de curva, score y espacio tangente. También brindamos algunos ejemplos e incluimos demostraciones de resultados vinculados a la estadística semiparamétrica enunciados a lo largo de este trabajo.

Sin embargo, para facilitar la lectura del documento, creemos conveniente hacer una primera presentación de dichos conceptos que, si bien puede resultar un tanto informal, utiliza nociones con las que, quienes trabajamos en el campo de la estadística, estamos sumamente familiarizados. Vamos ahora entonces a introducir (informalmente) la noción de score y espacio tangente, recordando que en el Apéndice A se realiza una presentación rigurosa de estos mismos conceptos.

Sea \mathcal{N} un modelo para la distribución de Z y sea p un elemento en \mathcal{N} . Dada una curva diferenciable $\{p_t : t \in [0, \epsilon)\}$ (un submodelo unidimensional) contenida en \mathcal{N} que pasa por p en $t = 0$, llamaremos score de la curva a

$$S(z) = \frac{\partial}{\partial t} \log(p_t(z)) \Big|_{t=0}.$$

Bajo condiciones de regularidad, los scores son elementos de cuadrado integrable y esperanza cero: $E_p(S^2) < \infty$ y $E_p(S) = 0$, siendo $S = S(Z)$. Dada $\mathcal{C}(p)$, una familia de curvas diferenciables en el modelo \mathcal{N} que pasan por p en $t = 0$, el conjunto de scores

asociados a dichas curvas conforma el llamado conjunto tangente. Además, llamamos espacio tangente a la clausura (en $L_2(p)$) del subespacio lineal generado por el conjunto tangente. Por último, llamamos espacio tangente maximal al obtenido con todas las curvas diferenciables que pasan por p y lo denotamos mediante $\overline{\mathcal{N}}_p$.

Notemos que los conjuntos tangentes son subespacios vectoriales cerrados en $L_2^0(p)$, el subconjunto de $L_2(p)$ formado por funciones con esperanza cero bajo p .

En lo sucesivo, para cada $B \equiv b(Z) \in L_2(p)$ y para cada Λ subespacio cerrado de $L_2(p)$, denotamos con $\Pi_p(B|\Lambda)$ a la $L_2(p)$ – proyección de B en Λ ; es decir, $\Pi_p(B|\Lambda)$ es el único elemento de Λ que satisface $E_p(\{B - \Pi_p(B|\Lambda)\} B^\dagger) = 0$ para todo $B^\dagger \in \Lambda$. Denotamos con $\Lambda_{\bar{c}}(p)$ al espacio tangente maximal del submodelo \mathcal{P}_{g, h_c} en el punto $p = gh \in \mathcal{P}_{g, h_c}$. Es decir, $\Lambda_{\bar{c}}(p)$ es el menor subespacio vectorial cerrado de $L_2(p)$ que contiene a todos los scores asociados a modelos paramétricos unidimensionales de \mathcal{P}_{g, h_c} que pasan por p . Utilizando la notación clásica en el tema, tenemos que

$$\Lambda_{\bar{c}}(p) = \overline{\mathcal{P}_{g, h_c}(p)}.$$

De ahora en más utilizaremos $\Pi_{gh}(\cdot|\cdot)$ y $\Lambda_j(gh)$ para denotar a $\Pi_p(\cdot|\cdot)$ y $\Lambda_j(p)$, respectivamente, donde $p = gh_1 \dots h_K$.

Las construcciones que haremos a lo largo de este capítulo dependen del resultado presentado en el Lema 4.2.1, razón por la cual, en adelante lo llamaremos Lema Fundamental. La demostración del mismo se basa en que la esperanza de una variables N coincide bajo diferentes distribuciones, cuando N pertenece a cierto espacio y el modelo es convexo. Este hecho subyace en la esencia de la posibilidad de construir bajo ciertos modelos, estimadores de parámetros finito dimensionales que dependen de estimadores de parámetros infinito dimensionales, pero cuya distribución asintótica no se ve afectada por la estimación de los parámetros infinito dimensionales. Ver por ejemplo el ejemplo 25.61 de van der Vaart (1999) *Asymptotic Statistics*). En el Teorema 4.2.1 enunciamos formalmente este resultado. Para facilitar su enunciado y el enunciado de varios resultados posteriores, definimos separadamente el siguiente supuesto sobre un par cualquiera de medidas p y p^* .

Supuesto 1: Dadas dos densidades p y p^* , se satisfacen las siguientes condiciones:

1. $\{z : p^*(z) = 0\} \subseteq \{z : p(z) = 0\}$.
2. $\int \left(\frac{p(z)}{p^*(z)} - 1 \right)^2 p^*(z) d\mu(z) < \infty$.

Teorema 4.2.1. Sean \mathcal{N} un modelo, p y p^* dos elementos en \mathcal{N} . Supongamos que $tp + (1 - t)p^*$ pertenece a \mathcal{N} , para todo t en el intervalo $[0, 1]$. Supongamos además

que p y p^* verifican el Supuesto 1. Entonces, si N pertenece a $L_2(p^*)$ y es ortogonal al espacio tangente maximal del modelo \mathcal{N} en el punto p^* ($N \in \overline{\mathcal{N}_{p^*}}^\perp$), tenemos que N es integrable según p y se verifica que

$$E_p(N) = E_{p^*}(N). \quad (4.2)$$

Demostración. Daremos ahora una prueba heurística del resultado, siguiendo el espíritu de Newey (1990), postergando al Apéndice B, sección (8.3.2) la demostración rigurosa del mismo. Pongamos $N = N(Z)$, y notemos que

$$\begin{aligned} E_p(N) &= \int_{\mathcal{Z}} N(z) p(z) d\mu(z) = \int_{\mathcal{Z}} N(z) \frac{p(z)}{p^*(z)} p^*(z) d\mu(z) = \\ &= \int_{\mathcal{Z}} N(z) \left\{ \frac{p(z)}{p^*(z)} - 1 \right\} p^*(z) d\mu(z) + E_{p^*}(N) = \\ &= E_{p^*}(NS) + E_{p^*}(N), \end{aligned} \quad (4.3)$$

siendo que $S = S(Z)$, con $S(z) = p(z)/p^*(z) - 1$ el score de la curva $tp + (1-t)p^*$. Como N es ortogonal al espacio tangente maximal del modelo \mathcal{N} en el punto p^* , tenemos que $E_{p^*}(NS) = 0$ y por consiguiente, $E_p(N) = E_{p^*}(N)$, tal como queríamos demostrar. \square

Tenemos el siguiente corolario.

Corolario 4.2.1. Sean $p = gh$ y $p^* = gh^*$ dos elementos de \mathcal{P} satisfaciendo el Supuesto 1 y sea $c \subseteq [K]$ de forma tal que $h_i^* = h_i$ para todo $i \in c$. Supongamos que \mathcal{P}_{g, h_c} es convexo. Si $N \in \Lambda_{\bar{c}}(p^*)^\perp$, entonces

$$E_p(N) = E_{p^*}(N).$$

El próximo resultado que enunciaremos es el Lema Fundamental, que generaliza el Lema 1 de Robins, van der Laan Rotnitzky (2000) de $K = 1$ a $K > 1$, pero antes unas palabras sobre el contexto en el que se desarrolla la teoría presentada. Tanto la bibliografía mencionada como muchos de los trabajos existentes en el área, asumen que cuando la distribución de los datos admite una factorización con parámetros que varían de manera independiente, los espacios tangentes correspondiente a modelos que dejan libre solo una coordenada dependen exclusivamente de la misma. Es decir, se asume el siguiente hecho.

Supuesto 2: Sean $p = gh$ y $p^* = g^*h^*$ dos elementos en \mathcal{P} , de forma tal que para algún $j \in [K]$ $h_j = h_j^*$. Entonces

$$\Lambda_j(p) = \Lambda_j(p^*) = \Lambda_j(h_j). \quad (4.4)$$

La razón heurística que da origen al Supuesto 2 es la siguiente: sea $p_t = gh_{j,t} \prod_{k \neq j} h_k$ una curva que pasa por p en $t = 0$, con score (informal) dado por

$$B(z) = \frac{\partial}{\partial t} \log(p_t(z)) \Big|_{t=0} = \frac{\dot{h}_{j,t}}{h_{j,t}} \Big|_{t=0}.$$

Luego $B(z)$ depende de la curva p_t solo mediante $h_{j,t}$. Esto sugiere que si $p^* = g^*h^*$, con $h_j^* = h_j$, la curva $p_t^* = g^*h_{j,t} \prod_{k \neq j} h_k^*$ pasa por p^* en $t = 0$ y también tiene a $B(z)$ por score (informal).

Observación 4.2.1. *En el Apéndice A veremos que el Supuesto 2 se verifica si, por ejemplo, asumimos que todos los posibles cocientes entre funciones en el conjunto \mathcal{U}^{np} y todos los posibles cocientes entre funciones en \mathcal{V}_i^{np} , para $i \neq k$ están acotados superior e inferiormente por constantes positivas.*

A lo largo de la demostración del Lema Fundamental, quedará de manifiesto que el Supuesto 2 puede ser sustituido por una condición mas débil. Basta con que se verifique el siguiente hecho:

Supuesto 2 (débil): Sean $p = gh$ y $p^* = g^*h^*$ dos elementos en \mathcal{P} , de forma tal $h_j = h_j^*$. Entonces $E_p(N) = 0$ para todo $N \in \Lambda_j(p^*)$.

Por otra parte, cuando los factores h_k representan distribuciones condicionales se verifica el siguiente supuesto (ver Tsiatis (2006), Capítulo 4.):

Supuesto 3:

- a) Sea $p = gh \in \mathcal{P}$, si $i, j \in [K], i \neq j$ entonces $\Lambda_i(p)$ y $\Lambda_j(p)$ son espacios ortogonales en $L_2(p)$.
- b) Sea $p \in \mathcal{P}_{g, h_c}$. Se verifica

$$\Lambda_{\bar{c}}(p) = \bigoplus_{j \in \bar{c}} \Lambda_j(p).$$

Damos a continuación una razón heurística que sugiere la validez de este último supuesto, observando antes el siguiente hecho:

Observación 4.2.2. *Sean \mathcal{F}_1 y \mathcal{F}_2 modelos tales que $\mathcal{F}_1 \subseteq \mathcal{F}_2$. Entonces para $P \in \mathcal{F}_1$ resulta que $\overline{\mathcal{F}_{1P}} \subseteq \overline{\mathcal{F}_{2P}}$.*

Para entender la parte a) consideremos $S_i = S_i(Z)$ y $S_j = S_j(Z)$ scores en $\in \Lambda_i(p)$ y $\in \Lambda_j(p)$, respectivamente, asociados a las curvas $p_{i,t} = g \prod_{k \neq i} h_k h_{i,t}$ y $p_{j,r} =$

$g \prod_{k \neq j} h_k h_{j,r}$ en \mathcal{P} que en $t = 0$ pasan por p y veamos que $E_p(S_i S_j) = 0$. Dado el carácter heurístico del presente análisis, supondremos que

$$S_i(z) = \frac{\partial}{\partial t} \log(p_{i,t}(z)) \Big|_{t=0} = \frac{\dot{h}_{i,t}}{h_{i,t}} \Big|_{t=0}, \quad S_j(z) = \frac{\partial}{\partial t} \log(p_{j,r}(z)) \Big|_{t=0} = \frac{\dot{h}_{j,r}}{h_{j,r}} \Big|_{r=0},$$

donde, siendo $p_{i,t}$ un submodelo en \mathcal{P} que pasa por p tenemos que $h_{i,t} \in \mathcal{V}_i^{np}$ para t en un semientorno del origen y $h_{i,0} = h_i$; asimismo, $h_{j,r} \in \mathcal{V}_j^{np}$ para r en un semientorno del origen y $h_{j,0} = h_j$. Además $h_k \in \mathcal{V}_k^{np}$ para $k \neq i, k \neq j$.

Manteniendo el espíritu heurístico, supondremos que están dadas las condiciones necesarias para poder intercambiar derivadas con integrales, en cuyo caso tenemos que

$$\begin{aligned} E_p(S_i S_j) &= \int S_i(z) S_j(z) p(z) d\mu(z) = \\ &= \int \frac{\dot{h}_{i,t}|_{t=0}}{h_i} (z) \frac{\dot{h}_{j,r}|_{r=0}}{h_j} (z) g(z) h_i(z) h_j(z) \prod_{k \in \{1,2..K\} k \neq i, k \neq j} h_k(z) d\mu(z) = \\ &= \frac{\partial}{\partial r} \frac{\partial}{\partial t} \left\{ \int g(x) h_{i,t}(z) h_{j,r}(z) \prod_{k \in \{1,2..K\} k \neq i, k \neq j} h_k(x) d\mu(x) \right\} \Big|_{t=0 r=0} = 0, \end{aligned}$$

pues, para todo t y r , la función $p_{t,r} \equiv h_{i,t}(z) h_{j,r}(z) g(z) \prod_{k \in \{1,2..K\} k \neq i, k \neq j} h_k(x)$ es una densidad en el modelo \mathcal{P} , por consiguiente

$$\int g(z) h_{i,t}(z) h_{j,r}(z) \prod_{k \in \{1,2..K\} k \neq i, k \neq j} h_k(z) d\mu(z) = 1.$$

Parte b) Recordemos que $p = g h_1 \cdots h_K$. Veamos primero que $\Lambda_{\bar{c}}(p) \subseteq \bigoplus_{j \in \bar{c}} \Lambda_j(p)$. Basta ver que si $S = S(Z)$ es un score en $\Lambda_{\bar{c}}(p)$ resulta que $S \in \bigoplus_{j \in \bar{c}} \Lambda_j(p)$. Similarmenente al ítem anterior, la parte heurística radica en suponer que están dadas las condiciones de suavidad para que exista una curva $p_t = g \prod_{j \in \bar{c}} h_{j,t} \prod_{k \in c} h_k$ en \mathcal{P}_{g, h_c} que en $t = 0$ pasa por p , cumpliendo

$$S(z) = \frac{\partial}{\partial t} \log(p_t(z)) \Big|_{t=0} = \sum_{j \in \bar{c}} \frac{\dot{h}_{j,t}}{h_{j,t}} \Big|_{t=0} (z).$$

Definimos para cada $j \in \bar{c}$ la curva

$$p_{j,t} \equiv g h_{j,t} \prod_{k \neq j} h_k.$$

Resulta entonces que $p_{j,t}$ es una curva en el modelo $\mathcal{P}_{g,[K]-j}$ que pasa por p en $t = 0$, por consiguiente su derivada pertenece al espacio $\Lambda_j(p)$ y

$$S(z) = \sum_{j \in \bar{c}} \frac{\dot{h}_{j,t}|_{t=0}(z)}{h_{j,t}} = \sum_{j \in \bar{c}} \frac{\partial}{\partial t} \log(p_{j,t}(x)) |_{t=0} \in \oplus_{j \in \bar{c}} \Lambda_j(p).$$

Veamos la otra inclusión. Para $j \in \bar{c}$, tenemos que $c \subseteq [K] - j$, luego, $\mathcal{P}_{g,h_{[K]-j}} \subseteq \mathcal{P}_{g,h_c}$. De donde usando la Observación 4.2.2, resulta que

$$\Lambda_j(p) \subseteq \Lambda_{\bar{c}}(p)$$

y por lo tanto

$$\oplus_{j \in \bar{c}} \Lambda_j(p) \subseteq \Lambda_{\bar{c}}(p).$$

A lo largo de la demostración del Teorema 4.2.1, quedará de manifiesto que el Supuesto 3 puede ser sustituido por una condición mas débil. Basta con que se verifique el siguiente hecho:

Supuesto 3 (débil): Se $c \subseteq [K]$ y sean p y $p^* \in \mathcal{P}_{g,h_c}$. Si S es el score de la curva $pt + p^*(1-t)$ en $t = 0$, entonces

$$S \in \oplus_{j \in \bar{c}} \Lambda_j(p^*).$$

En el Apéndice A, discutimos condiciones que garantizan la validez de los supuestos hechos.

Vamos finalmente a presentar el Lema Fundamental. Para preservar la tradición en el área, asumiremos los Supuesto 2 y 3. En caso de trabajar considerando apenas el Supuesto 2 (débil), debe escribirse $\Lambda_k(p)$ en lugar de $\Lambda_k(h_k)$.

Lema 4.2.1. Lema Fundamental. Sean $p = gh$ y $p^* = gh^*$ en el modelo \mathcal{M} , satisfaciendo el Supuesto 1. Supongamos que se verifica el Supuesto 2, el Supuesto 3 y que $M(\beta(g), h^*) \in L_2(gh^*)$. Para $k \in [K]$, sean $g_k \in \mathcal{U}^{np}$ y $\beta_k \in \mathbb{R}^r$ verificando $M(\beta_k, h^*) \in L_2(g_k h^*)$. Si además existe $c \subseteq [K]$ cumpliendo

1. \mathcal{P}_{g,h_c} es convexo,
2. $h_k^* = h_k$, para todo $k \in c$,
3. para todo $k \in \bar{c}$

$$\Pi_{g_k h^*} (M(\beta_k, h^*) | \Lambda_k(h_k^*)) = \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*)),$$

entonces

$$E_{gh} \left(M(\beta(g), h^*) - \sum_{k=1}^K \Pi_{g_k h^*} (M(\beta_k, h^*) | \Lambda_k(h_k^*)) \right) = 0. \quad (4.5)$$

Demostración. Bajo las condiciones del presente Lema tenemos que

$$\begin{aligned} & M(\beta(g), h^*) - \sum_{k=1}^K \Pi_{g_k h^*} (M(\beta_k, h^*) | \Lambda_k(h_k^*)) = \\ & M(\beta(g), h^*) - \sum_{k \in \bar{c}} \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*)) - \sum_{k \in c} \Pi_{g_k h^*} (M(\beta_k, h^*) | \Lambda_k(h_k^*)). \end{aligned}$$

Asumiendo el Supuesto 3 resulta que $\Lambda_{\bar{c}}(p^*) = \bigoplus_{j \in \bar{c}} \Lambda_j(p^*)$ y por consiguiente $M(\beta(g), h^*) - \sum_{k \in \bar{c}} \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*))$ pertenece a $\Lambda_{\bar{c}}(p^*)^\perp$. Luego, apelando al Corolario 4.2.1 con $p = gh$ y $p^* = gh^*$, tenemos que

$$\begin{aligned} & E_{gh} \left(M(\beta(g), h^*) - \sum_{k \in \bar{c}} \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*)) \right) = \\ & E_{gh^*} \left(M(\beta(g), h^*) - \sum_{k \in \bar{c}} \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*)) \right). \end{aligned} \quad (4.6)$$

Además, tenemos que esta última esperanza vale cero, utilizando la definición de $\beta(g)$ y notando que $\Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*)) \in L_2^0(gh^*)$. Por otra parte, para $k \in c$, tenemos que $h_k = h_k^*$. En tal caso, el Supuesto 2 (basta con que se verifique la versión débil) garantiza que

$$E_{gh} (\Pi_{g_k h^*} (M(\beta_k, h^*) | \Lambda_k(h_k^*))) = 0.$$

En caso de que se verifique el Supuesto 3 débil, veremos ahora que la igualdad (4.6) también se cumple. Basta adaptar la demostración del Teorema 4.2.1 al presente contexto, de la siguiente manera. Sean ahora $N = M(\beta(g), h^*) - \sum_{k \in \bar{c}} \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*))$, $p = gh$ y $p^* = gh^*$. Vimos en la ecuación (4.3) que

$$E_p(N) = E_{p^*}(NS) + E_{p^*}(N).$$

Como $N = M(\beta(g), h^*) - \sum_{k \in \bar{c}} \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*)) \in \bigoplus_{j \in \bar{c}} \Lambda_j(p^*)^\perp$ y el Supuesto 3 débil establece que $S \in \bigoplus_{j \in \bar{c}} \Lambda_j(p^*)$, concluimos que $E_{p^*}(NS) = 0$ y por consiguiente, se satisface la identidad presentada en (4.6). Queda entonces demostrado el resultado. \square

De alguna manera el Lema Fundamental sugiere una familia de funciones de estimación para $\beta(g)$ siempre que podamos codificar la dependencia en p y β de $\Pi_p (M(\beta, h) | \Lambda_k(p))$

mediante funcionales $\rho_k(g)$ como los definidos en (3.11). Por lo tanto, a continuación, calculamos para cada uno de los ejemplos dichas proyecciones. Notemos que en los ejemplos se verifica la Condición 1, y por lo tanto si $B \in L_2(p)$, tal como demostraremos en el Corolario 7.6.1 del Apéndice A, tenemos que

$$\Pi_p(B|\Lambda_k(p)) = E_p(B|\bar{A}_k, \bar{L}_k) - E_p(B|\bar{A}_{k-1}, \bar{L}_k). \quad (4.7)$$

En el ejemplo 1 recordamos que $M(\beta, h_1, h_2) = \frac{R_1 R_2}{E_{h_1}(R_1|L_1)E_{h_2}(R_2|R_1, \bar{L}_2)}(Y - \beta)$ y si $p = gh_1h_2$, para cada $k = 1, 2$ tenemos que

$$E_p(M(\beta, h_1, h_2) | \bar{R}_k \bar{L}_k) = \frac{\prod_{j=1}^k R_j}{\prod_{j=1}^k E_{h_j}(R_j | \bar{R}_{j-1} \bar{L}_j)} \{\nu_k(g) - \beta\},$$

siendo

$$\begin{aligned} \nu_1(g) &\equiv E_g(E_g(Y|R_2 = R_1 = 1, \bar{L}_2 = \cdot) | R_1 = 1, L_1 = \cdot), \\ \nu_2(g) &\equiv E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot). \end{aligned}$$

Por lo tanto, de la propiedad (4.7), podemos concluir que

$$\Pi_p(M(\beta, h_1, h_2) | \Lambda_1(p)) = \left\{ \frac{R_1}{E_{h_1}(R_1|L_1)} - 1 \right\} \{\nu_1(g) - \beta\}, \quad (4.8)$$

$$\Pi_p(M(\beta, h_1, h_2) | \Lambda_2(p)) = \frac{R_1}{E_{h_1}(R_1|L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2|R_1, \bar{L}_1)} - 1 \right\} \{\nu_2(g) - \beta\}. \quad (4.9)$$

En el ejemplo 2 tenemos que $M(\beta, h_1, h_2) = \frac{R}{E_{h_2}(R|X, V, W)}(Y - \beta X)\Delta_{h_1}(X, W)$, donde $\Delta_{h_1}(X, W) \equiv \{X - E_{h_1}(X|W)\}q(W)$. Luego si $p = gh_1h_2$ pertenece a \mathcal{M} , resulta

- 1) $E_p(M(\beta, h_1, h_2) | W, X, V, R) = \frac{R}{E_{h_2}(R|X, V, W)}(\nu_2(g) - \beta X)\Delta_{h_1}(X, W),$
- 2) $E_p(M(\beta, h_1, h_2) | W, X, V) = (\nu_2(g) - \beta X)\Delta_{h_1}(X, W),$
- 3) $E_p(M(\beta, h_1, h_2) | W, X) = (\nu_1(g) - (\beta - \beta(g))X)\Delta_{h_1}(X, W),$
- 4) $E_p(M(\beta, h_1, h_2) | W) = (\beta - \beta(g))E_{h_1}[X|W](1 - E_{h_1}[X|W]),$

con

$$\nu_1(g) \equiv E_g(E_g(Y - \beta(g)X | R = 1, V = \cdot, X = \cdot, W = \cdot) | X = \cdot, W = \cdot) \equiv e_g(W = \cdot).$$

$$\nu_2(g) \equiv E_g(Y | R = 1, X = \cdot, V = \cdot, W = \cdot),$$

Podemos entonces concluir que

$$\Pi_p(M(\beta, h_1, h_2) | \Lambda_1(p)) = \nu_1(g) \Delta_{h_1}(X, W) + \{\beta(g) - \beta\} (1 - E_{h_1}[X|W]) \Delta_{h_1}(X, W) \quad (4.10)$$

$$\Pi_p(M(\beta, h_1, h_2) | \Lambda_2(p)) = \left\{ \frac{R}{E_{h_2}(R|W, X, V)} - 1 \right\} \{\nu_2(g) - \beta X\} \Delta_{h_1}(X, W). \quad (4.11)$$

Finalmente resumimos la información en la Tabla 3 dada a continuación:

Tabla 3

Ejemplo	$\Pi_p(M(\beta, h_1, h_2) \Lambda_1(p))$
1	$\left\{ \frac{R_1}{E_{h_1}(R_1 L_1)} - 1 \right\} \{\nu_1(g) - \beta\}$
2	$(\nu_1(g) + \{\beta(g) - \beta\} E_{h_1}(1 - X W)) \{X - E_{h_1}(X W)\} q(W)$
Ejemplo	$\Pi_p(M(\beta, h_1, h_2) \Lambda_2(p))$
1	$\frac{R_1}{E_{h_1}(R_1 L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2 R_1, \bar{L}_1)} - 1 \right\} \{\nu_2(g) - \beta\}$
2	$\left\{ \frac{R}{E_{h_2}(R W, X, V)} - 1 \right\} \{\nu_2(g) - \beta X\} \{X - E_{h_1}(X W)\} q(W)$

4.3. Construcción de funciones de estimación insesgadas múltiple protegidas para $\beta(g)$ en modelos de unión e intersección.

Como decíamos en la sección anterior, el Lema Fundamental invita a estudiar la dependencia de la proyección $\Pi_{gh}(M(\beta, h) | \Lambda_k(gh))$ en g y β para identificar candidatos $\rho_k(g)$ cuyo modelado paramétrico permita la construcción de ecuaciones de estimación múltiple robustas. Para poder proyectar $M(\beta, h)$ en el espacio $\Lambda_k(gh)$ es necesario garantizar que $M(\beta, h)$ tenga segundo momento finito bajo $p = gh$. Si bien la definición del funcional $\beta(\cdot)$ no requería esta restricción, en adelante tendremos que asumir que en \mathcal{M} se verifica esta propiedad. Mas aún, consideraremos la siguiente condición.

Condición 2: Representación de la Proyección. Para cada $k \in [K]$, existe \mathcal{U}_k de forma tal que se satisfacen las siguientes tres condiciones:

1. $\mathcal{U} \subseteq \mathcal{U}_k \subseteq \mathcal{U}^{np}$.

2. $M(\beta, h) \in L_2(gh)$, para todo $(g, h, \beta) \in \mathcal{U}_k \times \otimes_{k=1}^K \mathcal{V}_k \times \mathbb{R}^r$.
3. La aplicación $(g, h, \beta) \rightarrow \Pi_{gh}(M(\beta, h) | \Lambda_k(gh))$ con dominio $\mathcal{U}_k \times \otimes_{k=1}^K \mathcal{V}_k \times \mathbb{R}^r$ depende en g y β solo mediante $\psi_k\{\eta_k(g), \beta\}$. Es decir

$$\text{si } g' \in \mathcal{U}_k \text{ y } \psi_k\{\eta_k(g), \beta\} = \psi_k\{\eta_k(g'), \beta'\},$$

$$\text{entonces } \Pi_{gh}(M(\beta, h) | \Lambda_k(gh)) = \Pi_{g'h}(M(\beta', h) | \Lambda_k(g'h)), \quad (4.12)$$

donde

$$\eta_k : \mathcal{U}_k \rightarrow \mathcal{N}_k \text{ y } \psi_k : \mathcal{N}_k \times \mathbb{R}^r \rightarrow \mathcal{X}_k, \quad (4.13)$$

son dos funciones que tienen por imagen a los conjuntos \mathcal{N}_k y \mathcal{X}_k formados por funciones medibles de \mathcal{Z} con imágenes en algunos espacios métricos \mathbb{M}_{η_k} y \mathbb{M}_{ψ_k} , respectivamente.

Bajo este supuesto, para $k \in [K]$ consideremos la función π_k satisfaciendo

$$\Pi_{gh}(M(\beta, h) | \Lambda_k(gh)) = \pi_k(\psi_k\{\eta_k(g), \beta\}, h). \quad (4.14)$$

Los supuestos de la condición 2 se cumplen si tomamos \mathcal{U}_k de forma tal que el segundo momento este bien definido, $\eta_k(g) = g$, $\psi_k\{\eta_k(g), \beta\} = (\eta_k(g), \beta)$ y π_k la función que codifica la dependencia del lado izquierdo de la igualdad dada en (4.14) en (g, β, h) . Estas elecciones nos darían la representación trivial de la proyección. En muchos casos hay representación no trivial. En el ejemplo 1, vamos a redefinir el dominio del funcional, tomando $\mathcal{U} = \mathcal{U}_1 = \mathcal{U}_2$, siendo un conjunto de funciones g para las cuales existe $h \in \otimes_{k=1}^2 \mathcal{V}_k$ con $E_{gh}[Y^2] < \infty$. Apelando al Lema 3.1.1, podemos concluir que con esta elección se satisfacen los primeros dos requisitos de la Condición 2. Además, tenemos que

$$\begin{aligned} \Pi_p(M(\beta, h) | \Lambda_1(p)) &= \left\{ \frac{R_1}{E_{h_1}(R_1 | L_1)} - 1 \right\} \{E_g(E_g(Y | R_2 = 1, \bar{L}_2) | R_1 = 1, L_1) - \beta\}, \\ \Pi_p(M(\beta, h) | \Lambda_2(p)) &= \frac{R_1}{E_{h_1}(R_1 | L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2 | R_1, \bar{L}_1)} - 1 \right\} \{E_g(Y | R_2 = 1, \bar{L}_2) - \beta\}. \end{aligned}$$

Podemos entonces tomar

$$\begin{aligned} \eta_1(g) = \nu_1(g) &\equiv E_g(E_g(Y | R_2 = R_1 = 1, \bar{L}_2 = \cdot) | R_1 = 1, L_1 = \cdot), \\ \eta_2(g) = \nu_2(g) &\equiv E_g(Y | R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot), \\ \psi_k(\eta_k(g), \beta) &= \eta_k(g) - \beta. \quad k = 1, 2. \end{aligned}$$

De esta manera, las proyecciones pueden ser escritas como

$$\begin{aligned}\pi_1(\psi_1(\eta_1(g), \beta), h) &= \left\{ \frac{R_1}{E_{h_1}(R_1|L_1)} - 1 \right\} \psi_1(\eta_1(g), \beta), \\ \pi_2(\psi_2(\eta_2(g), \beta), h) &= \frac{R_1}{E_{h_1}(R_1|L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2|R_1, \bar{L}_1)} - 1 \right\} \psi_2(\eta_2(g), \beta).\end{aligned}$$

En el ejemplo 2, para garantizar que $M(\beta, h) \in L_2(gh)$, vamos a empezar considerando que g pertenece a \mathcal{A}_2 , el conjunto de funciones para las cuales existe $h \in \otimes_{k=1}^2 \mathcal{V}_k$ con $E_{gh}[Y^2] < \infty$. Tomaremos entonces $\mathcal{U}_2 = \mathcal{A}_2$ mientras que redefiniremos \mathcal{U} , siendo $\mathcal{U} = \mathcal{U}_1$, un subconjunto de \mathcal{A}_2 para los cuales se satisface el modelo :

$$\begin{aligned}\mathcal{U}_1 \equiv \mathcal{U} &\equiv \{g \in \mathcal{A}_2 : \text{existe } \beta(g) \in \mathbb{R}^r \text{ tal que} \\ &E_g(E_g(Y - \beta(g)X|R=1, V, X, W)|X, W) \text{ no depende de } X\}.\end{aligned}$$

Notemos que

$$\begin{aligned}\Pi_p(M(\beta, h)|\Lambda_1(p)) &= \nu_1(g) \Delta_{h_1}(X, W) + \{\beta(g) - \beta\} (1 - E_{h_1}[X|W]) \Delta_{h_1}(X, W), \\ \Pi_p(M(\beta, h)|\Lambda_2(p)) &= \left\{ \frac{R}{E_{h_2}(R|W, X, V)} - 1 \right\} \{\nu_2(g) - \beta X\} \Delta_{h_1}(X, W),\end{aligned}$$

siendo $\Delta_{h_1}(X, W) \equiv \{X - E_{h_1}(X|W)\} q(W)$,

$$\begin{aligned}\nu_1(g) &\equiv E_g(E_g(Y - \beta(g)X|R=1, V = \cdot, X = \cdot, W = \cdot)|X = \cdot, W = \cdot) \equiv e_g(W), \\ \nu_2(g) &\equiv E_g(Y|R=1, X = \cdot, V = \cdot, W = \cdot).\end{aligned}$$

Podemos entonces considerar

- ★ $\eta_1(g) = (\eta_{1,1}, \eta_{1,2})$ con $\eta_{1,1}(g) = \nu_1(g)$ y $\eta_{1,2}(g) = \beta(g)$,
- ★ $\psi_1(\eta_1(g), \beta) = (\psi_{1,1}, \psi_{1,2})$ con $\psi_{1,1}(\eta_1(g), \beta) = \eta_{1,1}(g)$ y $\psi_{1,2}(\eta_1(g), \beta) = \eta_{1,2}(g) - \beta$,
- ★ $\eta_2(g) = \nu_2(g)$,
- ★ $\psi_2(\eta_2(g), \beta) = \eta_2(g) - \beta X$,

en cuyo caso las proyecciones pueden ser representadas de la siguiente manera:

$$\begin{aligned}\pi_1(\psi_1(\eta_1(g), \beta), h) &= (\psi_{1,1}(\eta_1(g), \beta) + \psi_{1,2}(\eta_1(g), \beta) E_{h_1}(1 - X|W)) \Delta_{h_1}(X, W), \\ \pi_2(\psi_2(\eta_2(g), \beta), h) &= \left\{ \frac{R}{E_{h_2}(R|W, X, V)} - 1 \right\} \psi_2(\eta_2(g), \beta) \Delta_{h_1}(X, W).\end{aligned}$$

A modo de resumen, en la Tabla 4 se listan las elecciones de $\eta_k \psi_k$ recién presentadas, mientras que en la Tabla 5 listamos π_k para cada uno de los ejemplos.

Tabla 4

Ejemplo	$\eta_1(g)$	$\psi_1(\eta_1(g), \beta)$
1	$\nu_1(g)$	$\eta_1(g) - \beta$
2	$\eta_1 = (\eta_{1,1}, \eta_{1,2})$ with $\eta_{1,1}(g) = \nu_1(g)$ $\eta_{1,2}(g) = \beta(g)$	$\psi_1 = (\psi_{1,1}, \psi_{1,2})$ with $\psi_{1,1}(\eta_1(g), \beta) = \eta_{1,1}(g)$ $\psi_{1,2}(\eta_1(g), \beta) = \eta_{1,2}(g) - \beta$
Ejemplo	$\eta_2(g)$	$\psi_2(\eta_2(g), \beta)$
1	$\nu_2(g)$	$\eta_2(g) - \beta$
2	$\nu_2(g)$	$\eta_2(g) - \beta X$

Tabla 5

Ejemplo	$\pi_1(\psi_1(\eta_1(g), \beta), h)$
1	$\left\{ \frac{R_1}{E_{h_1}(R_1 L_1)} - 1 \right\} \psi_1(\eta_1(g), \beta)$
2	$(\psi_{1,1}(\eta_1(g), \beta) + \psi_{1,2}(\eta_1(g), \beta) E_{h_1}(1 - X W)) \{X - E_{h_1}(X W)\} q(W)$
Example	$\pi_2(\psi_2(\eta_2(g), \beta), h)$
1	$\frac{R_1}{E_{h_1}(R_1 L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2 R_1, \bar{L}_1)} - 1 \right\} \psi_2(\eta_2(g), \beta)$
2	$\left\{ \frac{R}{E_{h_2}(R W, X, V)} - 1 \right\} \psi_2(\eta_2(g), \beta) \{X - E_{h_1}(X W)\} q(W)$

En el Apéndice B demostramos que la Condición 2 permite reformular el Lema 4.2.1 de la siguiente manera.

Proposición 4.3.1. Sean $p = gh$ y $p^* = gh^*$ en el modelo \mathcal{M} , satisfaciendo el Supuesto 1. Supongamos que se verifica el Supuesto 2 (o el débil), el Supuesto 3 y la Condición 2, para ciertas funciones η_k y ψ_k definidas como en (4.13). Sea π_k la función definida en (4.14). Para cada $k \in [K]$, tomemos $g_k \in \mathcal{U}_k$ y $\beta_k \in \mathbb{R}^r$. Luego, fijado $b \subseteq [K]$, la ecuación en la variable β dada por

$$E_{gh} \left(M(\beta, h^*) - \sum_{j \in b} \pi_j(\psi_j\{\eta_j(g_j), \beta_j\}, h^*) - \sum_{j \in \bar{b}} \pi_j(\psi_j\{\eta_j(g_j), \beta\}, h^*) \right) = 0 \quad (4.15)$$

tiene como solución a $\beta = \beta(g)$, siempre que exista $c \subseteq [K]$ verificando

i) \mathcal{P}_{g, h_c} es convexo,

ii) $h_j^* = h_j$ si $j \in c$,

iii) $\psi_j \{ \eta_j(g_j), \beta_j \} = \psi_j \{ \eta_j(g), \beta(g) \}$ si $j \in \bar{c} \cap b$,

iv) $\eta_j(g_j) = \eta_j(g)$ si $j \in \bar{c} \cap \bar{b}$.

La proposición anterior induce a construir funciones de estimación insesgadas múltiple protegidas considerando modelos paramétricos \mathcal{G}_k para los funcionales $\eta_k(g)$ o bien para $\psi_k(\eta_k(g), \beta(g))$. Es decir, fijado $b \subseteq [K]$, para cada $k \in [K]$ definimos $\rho_k : \mathcal{U} \rightarrow \mathcal{R}_k$ mediante

$$\rho_k(g) \equiv \begin{cases} \psi_k \{ \eta_k(g), \beta(g) \} & \text{si } k \in b \\ \eta_k(g) & \text{si } k \in \bar{b} \end{cases} \quad (4.16)$$

y consideramos cierta clase paramétrica $\mathcal{R}_{sub,k}$ para modelar $\rho_k(g)$. Es decir, supondremos que $\rho_k(g) \in \mathcal{R}_{sub,k}$ con

$$\mathcal{R}_{sub,k} \equiv \{ \rho_{k, \tau_k} \in \mathcal{R}_k^{ext} : \tau_k \in \Upsilon_k \}, \quad (4.17)$$

donde $\Upsilon_k \subseteq \mathbb{R}^{r_k}$ y $\mathcal{R}_k \subset \mathcal{R}_k^{ext}$, tomando $\mathcal{R}_k^{ext} = \mathcal{X}_k$ si $k \in b$, y $\mathcal{R}_k^{ext} = \mathcal{N}_k$ si $k \in \bar{b}$, para \mathcal{X}_k y \mathcal{N}_k introducidos en (4.13). Con esta formulación garantizamos uno de nuestros objetivos: contemplamos la posibilidad de que \mathcal{U} esté incluido estrictamente en \mathcal{U}_k , con lo cual puede que \mathcal{R}_k^{ext} incluya estrictamente a \mathcal{R}_k , posibilitando que $\mathcal{R}_{sub,k} \cap \mathcal{R}_k = \emptyset$. Como ya hemos mencionado, en tal caso, decimos que $\mathcal{R}_{sub,k}$ es incompatible con \mathcal{M} .

Ya estamos en condiciones de presentar una familia de funciones de estimación insesgadas múltiple protegidas. Para cada $k \in [K]$ consideramos la clase paramétrica $\mathcal{V}_{sub,k}$ para h_k definida como en (3.8) y denotamos con h_α a $(h_{1, \alpha_1}, \dots, h_{K, \alpha_K})$. Definimos

$$U^b(\beta, \tau, \alpha) \equiv M(\beta, h_\alpha) - \sum_{j \in b} \pi_j(\rho_{j, \tau_j}, h_\alpha) - \sum_{j \in \bar{b}} \pi_j(\psi_j \{ \rho_{j, \tau_j}, \beta \}, h_\alpha). \quad (4.18)$$

La próxima proposición, demostrada en el Apéndice B, establece que $U^b(\beta, \tau, \alpha)$ es una función de estimación insesgada múltiple protegida para $\beta(g)$ en el modelo de unión intersección dado por $\cup_{k=1}^K \{ \mathcal{H}_k \cap \mathcal{G}_k \}$ y generaliza el Teorema 2 de Robins et. Al. (2001) de $K = 2$ a $K > 2$.

Proposición 4.3.2. *Supongamos que se verifica el Supuesto 2 (o su versión débil), la Condición 2 y que \mathcal{P}_{g, h_c} es convexo para todo $c \subseteq [K]$, $g \in \mathcal{U}$ y $h_c \in \otimes_{j \in c} \mathcal{V}_j$. Consideremos $\mathcal{V}_{sub,k}$, modelos paramétricos para cada h_k , como los definidos en (3.8). Asumamos también que dada $p = gh \in \mathcal{M}_c$ y $p^* = gh_\alpha$, con $\alpha_i = \alpha_i(h_i)$ para $i \in c$, p y p^* satisfacen el Supuesto 1, cualquiera sea $c \subseteq [K]$.*

Fijado un subconjunto b de $[K]$, para cada $k \in [K]$ consideremos $\rho_k(g)$ definida mediante (4.16) y consideremos modelos paramétricos $\mathcal{R}_{sub,k}$ para $\rho_k(g)$ definidos como en (4.17). Para $1 \leq k \leq K$ sean

$$\mathcal{H}_k \equiv \{p \in \mathcal{M} : h_k \in \mathcal{V}_{sub,k}\},$$

$$\mathcal{G}_k \equiv \{p \in \mathcal{M} : \rho_k(g) \in \mathcal{R}_{sub,k}\}.$$

Bajo estos supuestos la función $U^b(\beta, \tau, \alpha)$ definida en (4.18) es una función de estimación insesgada múltiple protegida para $\beta(g)$ en el modelo de unión intersección dado por $\cup_{k=1}^K \{\mathcal{H}_k \cap \mathcal{G}_k\}$. Es decir, se satisface que para cada $c \subseteq [K]$, cada $p = gh \in \mathcal{M}_c$, para cada $\tau_c \in \otimes_{k \in c} \Upsilon_k$ y $\alpha_{\bar{c}} \in \otimes_{k \in \bar{c}} \Xi_k$

$$E_p \left(U^b(\beta, \tau, \alpha) \right) = 0 \quad \text{en } \beta = \beta(g), \tau_{\bar{c}} = \tau_{\bar{c}}(g) \text{ y } \alpha_c = \alpha_c(h_c).$$

Es importante destacar que, fijado un subconjunto b de $[K]$, obtenemos una función de estimación insesgada múltiple protegida $U^b(\beta, \tau, \alpha)$. Para simplificar la notación, en adelante escribiremos $U(\beta, \tau, \alpha)$ en lugar de $U^b(\beta, \tau, \alpha)$, omitiendo la dependencia en b . De todas formas, vale la pena remarcar que la proposición anterior brinda una familia de funciones de estimación múltiple protegidas indexada en todos los posibles subconjuntos b de $[K]$, cada una de ellas induciendo un procedimiento de estimación diferente para el parámetro de interés.

Destaquemos también que, debido a la definición de los funcionales $\rho_k(g)$, si $k \in b$ entonces vamos a modelar a la proyección de la función M evaluada en el parámetro de interés $\beta(g)$, mientras que si $k \in \bar{b}$ modelamos la proyección libre, es decir, evaluada en un β mudo. La flexibilidad que presenta esta construcción en elegir modelar distintos objetos nos permitió recuperar la ecuación U presentada en cada uno de los ejemplos.

Puede ocurrir que evaluar en $\beta(g)$ algunas proyecciones brinde métodos de estimación más parsimoniosos por sobre otros procedimientos. Durante la elaboración del presente trabajo, en una primer instancia, no se consideró la posibilidad de modelar a las proyecciones de la función M evaluadas en $\beta(g)$. Es decir, solo consideramos el caso en que $b = \emptyset$. Recién después de un análisis exhaustivo de diferentes ejemplos y algunas controversias, arribamos a esta formulación.

Para finalizar este capítulo, notemos que tomando $b = \{\emptyset\}$ en el ejemplo 1 obtenemos que $U^b(\beta, \tau, \alpha)$ coincide con la función de estimación cuadruple protegida presentada en el ejemplo 1, mientras que en el ejemplo 2, tomando $b = \{1\}$, encontramos la función de estimación múltiple protegida usada en el procedimiento presentado en el ejemplo 2. Para ser más específicos, recordamos que en el ejemplo 1 tenemos

$$M(\beta, h_1, h_2) = \frac{R_2 R_1}{E_p(R_1 | L_1) E_p(R_2 | R_1, \bar{L}_2)} (Y - \beta),$$

$$\begin{aligned}\Pi_p(M(\beta, h) | \Lambda_1(p)) &= \left\{ \frac{R_1}{E_{h_1}(R_1|L_1)} - 1 \right\} \{E_g \{E_g(Y|R_2 = 1, \bar{L}_2) | R_1 = 1, L_1\} - \beta\}, \\ \Pi_p(M(\beta, h) | \Lambda_2(p)) &= \frac{R_1}{E_{h_1}(R_1|L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2|R_1, \bar{L}_1)} - 1 \right\} \{E_g(Y|R_2 = 1, \bar{L}_2) - \beta\}.\end{aligned}$$

Luego, poniendo

$$\begin{aligned}\eta_1(g) &= E_g(E_g(Y|R_2 = R_1 = 1, \bar{L}_2 = \cdot) | R_1 = 1, L_1 = \cdot), \\ \eta_2(g) &= E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot), \\ \psi_k(\eta_k(g), \beta) &= \eta_k(g) - \beta. \quad k = 1, 2,\end{aligned}$$

tenemos que

$$\begin{aligned}\pi_1(\psi_1(\eta_1(g), \beta), h) &= \left\{ \frac{R_1}{E_{h_1}(R_1|L_1)} - 1 \right\} \psi_1(\eta_1(g), \beta) \\ \pi_2(\psi_2(\eta_2(g), \beta), h) &= \frac{R_1}{E_{h_1}(R_1|L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2|R_1, \bar{L}_1)} - 1 \right\} \psi_2(\eta_2(g), \beta).\end{aligned}$$

Tomando $b = \{\emptyset\}$ para construir $U^b(\beta, \tau, \alpha)$ según la fórmula (4.18), tenemos que

$$\begin{aligned}\star \rho_1(g) &= E_g(E_g(Y|R_2 = R_1 = 1, \bar{L}_2 = \cdot) | R_1 = 1, L_1 = \cdot), \\ \star \mathcal{R}_{sub,1} &= \{\rho_{1,\tau_1} = b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\}, \\ \star \rho_2(g) &= E_g(Y|R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot), \\ \star \mathcal{R}_{sub,2} &= \{\rho_{2,\tau_2} = b_2(\cdot; \tau_2) : \tau_2 \in \Upsilon_2\}.\end{aligned}$$

y luego

$$\begin{aligned}U^b(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) &\equiv \frac{R_2(Y - b_2(\bar{L}_2, \tau_2))}{E(R_1|L_1; \alpha_1) E(R_2|R_1, \bar{L}_2; \alpha_2)} \\ &+ \frac{R_1}{E(R_1|L_1; \alpha_1)} \{b_2(\bar{L}_2, \tau_2) - b_1(L_1, \tau_1)\} + \{b_1(L_1; \tau_1) - \beta\},\end{aligned}$$

coincidiendo con la función de estimación utilizada para construir el estimador múltiple protegido del primer ejemplo.

En el ejemplo 2 tenemos que

$$M(\beta, h) = \frac{R}{E_{h_2}(R|X, V, W)}(Y - \beta X)\Delta_{h_1}(X, W),$$

donde $\Delta_{h_1}(X, W) \equiv \{X - E_{h_1}(X|W; \cdot)\}q(W)$. Luego si $p = gh_1h_2$ resulta que

$$\begin{aligned}\Pi_p(M(\beta, h) | \Lambda_1(p)) &= \nu_1(g) \Delta_{h_1}(X, W) + \{\beta(g) - \beta\} (1 - E[X|W])\Delta_{h_1}(X, W) \\ \Pi_p(M(\beta, h) | \Lambda_2(p)) &= \left\{ \frac{R}{E_{h_2}(R|W, X, V)} - 1 \right\} \{\nu_2(g) - \beta X\} \Delta_{h_1}(X, W)\end{aligned}$$

siendo

$$\begin{aligned}\nu_1(g) &\equiv E_g(E_g(Y - \beta(g)X | R=1, V=\cdot, X=\cdot, W=\cdot) | X, W), \\ \nu_2(g) &\equiv E_g(Y | R=1, X=\cdot, V=\cdot, W=\cdot).\end{aligned}$$

Considerando

- ★ $\eta_1(g) = (\eta_{1,1}, \eta_{1,2})$ con $\eta_{1,1}(g) = \nu_1(g)$ y $\eta_{1,2}(g) = \beta(g)$
- ★ $\psi_1(\eta_1(g), \beta) = (\psi_{1,1}, \psi_{1,2})$ con $\psi_{1,1}(\eta_1(g), \beta) = \eta_{1,1}(g)$ y $\psi_{1,2}(\eta_1(g), \beta) = \eta_{1,2}(g) - \beta$
- ★ $\eta_2(g) = \nu_2(g)$
- ★ $\psi_2(\eta_2(g), \beta) = \eta_2(g) - \beta X$,

tenemos que

$$\begin{aligned}\pi_1(\psi_1(\eta_1(g), \beta), h) &= (\psi_{1,1}(\eta_1(g), \beta) + \psi_{1,2}(\eta_1(g), \beta) E_{h_1}(1 - X|W)) \Delta_{h_1}(X, W), \\ \pi_2(\psi_2(\eta_2(g), \beta), h) &= \left\{ \frac{R}{E_{h_2}(R|W, X, V)} - 1 \right\} \psi_2(\eta_2(g), \beta) \Delta_{h_1}(X, W).\end{aligned}$$

Eligiendo $b = \{1\}$, resulta

- ★ $\rho_1(g) = (\nu_1(g), 0)$, y abusando de la notación, utilizaremos en adelante $\rho_1(g) = \nu_1(g)$,
- ★ $\mathcal{R}_{sub,1} = \{\rho_{1,\tau_1} = b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\}$,
- ★ $\rho_2(g) = \nu_2(g)$,
- ★ $\mathcal{R}_{sub,2} = \{\rho_{2,\tau_2} = b_2(\cdot; \tau_2) : \tau_2 \in \Upsilon_2\}$.

Luego, para $b = \{1\}$ construimos $U^b(\beta, \tau, \alpha)$ según la fórmula (4.18) obteniendo

$$U^b(\beta, \tau, \alpha) = \left\{ \frac{R \{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1) \right\} \Delta_{\alpha_1}(X, W),$$

donde $\Delta_{\alpha_1} \equiv \{X - E(X|W, \alpha_1)\}q(W)$, luego, si $q(W) = 1$, U^b es igual a la función de estimación múltiple protegida utilizada en el ejemplo 2.

Capítulo 5

Construcción de Funciones de estimación insesgadas para los parámetros de ruido

Para poder estimar $\beta(g)$ utilizando de forma efectiva la función de estimación $U(\beta, \tau, \alpha)$ presentada en el capítulo anterior, resta encontrar estimadores múltiple protegidos de los parámetros (τ, α) , tal como planteamos en la Sección 3.4.1. La función S_k deseada será generalmente el score para el parámetro α_k en el modelo paramétrico asumido para h_k y $\hat{\alpha}_k$ el estimador de máxima verosimilitud de α_k , ya que bajo condiciones de regularidad, la media muestral del score S_k se anula en el estimador de máxima verosimilitud de α_k . Luego, resta obtener un sistema $T = (T_1, \dots, T_K)$ que permita estimar en forma múltiple protegida los parámetros τ_k que indexan los modelos $\mathcal{R}_{k,sub}$, definidos como en la ecuación (4.17). Recordemos que $\mathcal{G}_k \equiv \{p \in \mathcal{M} : \rho_k(g) \in \mathcal{R}_{sub,k}\}$, y si $p = gh \in \mathcal{G}_k$, existe un único valor $\tau_k(g)$ para el cual $\rho_k(g) = \rho_{k,\tau_k(g)}$. Luego, si $p \in \mathcal{G}_k$, podemos entonces pensar en estimar en forma múltiple protegida $(\beta(g), \tau_k(g))$ conjuntamente, repitiendo la construcción hecha en el capítulo anterior para estimar $\beta(g)$. Esta propuesta resultará exitosa en la medida en que no surjan nuevos funcionales a ser modelados.

En la primer sección de este capítulo, veremos que cuando se cumple la Condición 1, es decir cuando el vector Z puede representarse como $Z = (L_1, A_1, L_2, A_2, \dots, A_K, L_{K+1})$ y para $k \in [K]$

$$h_k(z) = p_{A_k | \bar{L}_k, \bar{A}_{k-1}}(a_k | \bar{l}_k, \bar{a}_{k-1}),$$

si conseguimos codificar la dependencia de $E_{gh}(M(\beta, h) | \bar{A}_k, \bar{L}_k)$ en (g, β) mediante ψ_k y η_k podemos construir funciones de estimación múltiple protegidas insesgadas $T(\beta, \tau, \alpha) = \{T_k(\beta, \tau, \alpha); k = 1, \dots, K\}$, en el sentido de la Definición 3.5.

En la segunda sección aplicaremos nuestra propuesta para construir ecuaciones de estimación de los parámetros de ruido a cada uno de los ejemplos y veremos que en cada caso, eligiendo adecuadamente algunos de los parámetros del problema, recuperamos los procedimientos descritos en el Capítulo 2.

Al igual que en capítulos anteriores, con el objeto de hacer más fluida la lectura de este trabajo, posponemos las demostraciones hasta el Apéndice B.

5.1. Construcción

Cuando se satisfaga la Condición 1 requeriremos además que para funciones η_k y ψ_k como las definidas en (4.13) se verifique que

$$\begin{aligned} & \text{si } g' g \in \mathcal{U}_k \text{ y } \psi_k \{ \eta_k (g), \beta \} = \psi_k \{ \eta_k (g'), \beta' \}, \\ & \text{entonces } E_{gh} (M (\beta, h) | \bar{A}_k, \bar{L}_k) = E_{g'h} (M (\beta', h) | \bar{A}_k, \bar{L}_k). \end{aligned} \quad (5.1)$$

Es decir, vamos a considerar el siguiente escenario, que será de utilidad en la medida en que se verifique la Condición 1.

Condición 3: Representación de la esperanza condicional. Para cada $k \in [K]$, existe \mathcal{U}_k de forma tal que se satisfacen las siguientes tres condiciones:

1. $\mathcal{U} \subseteq \mathcal{U}_k \subseteq \mathcal{U}^{np}$.
2. $M(\beta, h) \in L_2(gh)$, para todo $(g, h, \beta) \in \mathcal{U}_k \times \otimes_{k=1}^K \mathcal{V}_k \times \mathbb{R}^r$.
3. La aplicación $(g, h, \beta) \rightarrow E_{gh} (M (\beta, h) | \bar{A}_k, \bar{L}_k)$ con dominio $\mathcal{U}_k \times \otimes_{k=1}^K \mathcal{V}_k \times \mathbb{R}^r$ depende en g y β solo mediante $\psi_k \{ \eta_k (g), \beta \}$. Es decir

$$\begin{aligned} & \text{si } g' g \in \mathcal{U}_k \text{ y } \psi_k \{ \eta_k (g), \beta \} = \psi_k \{ \eta_k (g'), \beta' \}, \\ & \text{entonces } E_{gh} (M (\beta, h) | \bar{A}_k, \bar{L}_k) = E_{g'h} (M (\beta', h) | \bar{A}_k, \bar{L}_k), \end{aligned} \quad (5.2)$$

donde

$$\eta_k : \mathcal{U}_k \rightarrow \mathcal{N}_k \text{ y } \psi_k : \mathcal{N}_k \times \mathbb{R}^r \rightarrow \mathcal{X}_k, \quad (5.3)$$

son dos funciones que tienen por codominio a los conjuntos \mathcal{N}_k y \mathcal{X}_k formados por funciones medibles de \mathcal{Z} en ciertos espacios métricos \mathbb{M}_{η_k} y \mathbb{M}_{ψ_k} , respectivamente.

Luego, bajo la Condición 3, para $k \in [K]$ consideramos la función e_k que nos permite escribir

$$E_{gh} (M (\beta, h) | \bar{A}_k, \bar{L}_k) = e_k (\psi_k \{ \eta_k (g), \beta \}, h). \quad (5.4)$$

Veamos en cada uno de los ejemplos considerados, posibles representaciones para las esperanzas condicionales involucradas. Utilizaremos en cada caso el mismo conjunto \mathcal{U}_i considerado en la Sección 4.3 del capítulo anterior. En el ejemplo 1 recordemos que

$$E_p(M(\beta, h) | \bar{R}_k, \bar{L}_k) = \frac{\prod_{j=1}^k R_j}{\prod_{j=1}^k E_{h_j}(R_j | \bar{R}_{j-1}, \bar{L}_j)} \{\nu_k(g) - \beta\}, \quad 1 \leq k \leq 2,$$

donde

$$\begin{aligned} \eta_1(g) = \nu_1(g) &\equiv E_g(E_g(Y | R_2 = R_1 = 1, \bar{L}_2) | R_1 = 1, L_1 = \cdot), \\ \eta_2(g) = \nu_2(g) &\equiv E_g(Y | R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot) \\ \psi_k(\eta_k(g), \beta) &= \eta_k(g) - \beta, \quad k = 1, 2. \end{aligned}$$

En suma, obtenemos la siguiente representación para cada esperanza condicional.

$$e_1(\psi_1(\eta_1(g), \beta), h) = \frac{R_1}{E_{h_1}(R_1 | L_1)} \psi_1(\eta_1(g), \beta),$$

y

$$e_2(\psi_2(\eta_2(g), \beta), h) = \frac{R_1}{E_{h_1}(R_1 | L_1)} \frac{R_2}{E_{h_2}(R_2 | R_1, \bar{L}_1)} \psi_2(\eta_2(g), \beta).$$

En el ejemplo 2 tenemos que

$$E_p(M(\beta, h) | W, X, V, R) = \frac{R}{E_{h_2}(R | X, V, W)} (\nu_2(g) - \beta X) \Delta_{h_1}(X, W),$$

$$E_p(M(\beta, h) | W, X) = (\nu_1(g) - (\beta - \beta(g))X) \Delta_{h_1}(X, W),$$

con

$$\nu_1(g) \equiv E_g(E_g(Y - \beta(g)X | R = 1, V = \cdot, X = \cdot, W = \cdot) | X = \cdot, W = \cdot),$$

$$\nu_2(g) \equiv E_g(Y | R = 1, X = \cdot, V = \cdot, W = \cdot).$$

Luego, poniendo

- ★ $\eta_1(g) = (\eta_{1,1}, \eta_{1,2})$ con $\eta_{1,1}(g) = \nu_1(g)$ y $\eta_{1,2}(g) = \beta(g)$,
- ★ $\psi_1(\eta_1(g), \beta) = (\psi_{1,1}, \psi_{1,2})$ con $\psi_{1,1}(\eta_1(g), \beta) = \eta_{1,1}(g)$ y $\psi_{1,2}(\eta_1(g), \beta) = \eta_{1,2}(g) - \beta$,
- ★ $\eta_2(g) = \nu_2(g)$,

$$\star \psi_2(\eta_2(g), \beta) = \eta_2(g) - \beta X,$$

obtenemos la siguiente representación para cada esperanza condicional

$$e_1(\psi_1(\eta_1(g), \beta), h) = (\psi_{1,1}(\eta_1(g), \beta) + \psi_{1,2}(\eta_1(g), \beta) X) \{X - E_{h_1}(X|W)\} q(W),$$

y

$$e_2(\psi_2(\eta_2(g), \beta), h) = \frac{R}{E_{h_2}(R|W, X, V)} \psi_2(\eta_2(g), \beta) \{X - E_{h_1}(X|W)\} q(W).$$

Resumimos las descomposiciones para las esperanzas condicionales de cada ejemplo en la tabla 8 dada a continuación

Tabla 8.

Example	$e_1(\psi_1(\eta_1(g), \beta), h)$
1	$\frac{R_1}{E_{h_1}(R_1 L_1)} \psi_1(\eta_1(g), \beta)$
2	$(\psi_{1,1}(\eta_1(g), \beta) + \psi_{1,2}(\eta_1(g), \beta) X) \{X - E_{h_1}(X W)\} q(W)$
Example	$e_2(\psi_2(\eta_2(g), \beta), h)$
1	$\frac{R_1}{E_{h_1}(R_1 L_1)} \frac{R_2}{E_{h_2}(R_2 R_1, \bar{L}_1)} \psi_2(\eta_2(g), \beta)$
2	$\frac{R}{E_{h_2}(R W, X, V)} \psi_2(\eta_2(g), \beta) \{X - E_{h_1}(X W)\} q(W)$

El siguiente resultado, cuya demostración presentamos en el Apéndice B, establece que si $Z = (L_1, A_1, L_2, A_2, \dots, A_K, L_{K+1})$ y, para $k \in [K]$, h_k denota la distribución condicional de A_k dado $(\bar{A}_{k-1}, \bar{L}_k)$ (Condición 1), si ψ_k y η_k codifican la dependencia de $E_{gh}(M(\beta, h)|\bar{A}_k, \bar{L}_k)$ en (g, β) , también codifican la dependencia de la proyección de $M(\beta, h)$ en el espacio $\Lambda_k(p)$.

Proposición 5.1.1. *Supongamos que se verifica la Condición 1. Entonces, si se satisface la Condición 3, también se cumple la Condición 2, utilizando los mismos objetos involucrados en cada una de ellas.*

La proposición anterior nos garantiza que, pese a haber impuesto nuevas restricciones en la Condición 3, siguen valiendo las hipótesis necesarias para que valgan los resultados del capítulo anterior.

En lo que resta, asumiremos que se satisface la Condición 1 y la Condición 3, en cuyo caso, como hemos mencionado se verifica que

$$\Pi_p(B|\Lambda_k(p)) = E_p(B|\bar{A}_k, \bar{L}_k) - E_p(B|\bar{A}_{k-1}, \bar{L}_k).$$

Dada la fórmula presentada para las proyecciones, en las construcciones venideras podremos considerar que $\Pi_p(B|\Lambda_k(p))$ está definido para cualquier función B de esperanza

finita. Mas generalmente, operaremos formalmente utilizando las propiedades que satisface la esperanza condicional, asumiendo en cada caso que están dadas las condiciones para hacer uso de las mismas.

Veremos ahora como la Condición 3 nos permite construir ecuaciones de estimación múltiple protegidas para los parámetros de ruido de la función de estimación múltiple protegida $U^b(\beta, \tau, \alpha)$, construida mediante los funcionales involucrados en la propia Condición 3. Con ese fin, recordemos que para construir $U^b(\beta, \tau, \alpha)$, fijamos $b \subseteq [K]$, definimos $\rho_k : \mathcal{U} \rightarrow \mathcal{R}_k$ mediante

$$\rho_k(g) \equiv \begin{cases} \psi_k \{ \eta_k(g), \beta(g) \} & \text{si } k \in b \\ \eta_k(g) & \text{si } k \in \bar{b} \end{cases}, \quad (5.5)$$

para η_k y ψ_k definidas en (5.3). Luego, consideramos la clase paramétrica $\mathcal{R}_{sub,k}$ para modelar $\rho_k(g)$, denotada por

$$\mathcal{R}_{sub,k} \equiv \{ \rho_{k, \tau_k} \in \mathcal{R}_k^{ext} : \tau_k \in \Upsilon_k \}, \quad (5.6)$$

donde $\Upsilon_k \subseteq \mathbb{R}^{r_k}$ y $\mathcal{R}_k \subset \mathcal{R}_k^{ext}$, tomando $\mathcal{R}_k^{ext} = \mathcal{X}_k$ si $k \in b$, y $\mathcal{R}_k^{ext} = \mathcal{N}_k$ si $k \in \bar{b}$, siendo \mathcal{X}_k y \mathcal{N}_k la imagen de ψ_k y η_k , respectivamente, como presentamos en (5.3). También recordemos que

$$\mathcal{G}_k \equiv \{ p \in \mathcal{M} : \rho_k(g) \in \mathcal{R}_{sub,k} \}. \quad (5.7)$$

Por otra parte, recordemos que para cada $k \in [K]$ consideramos la clase paramétrica

$$\mathcal{V}_{sub,k} = \{ h_{k, \alpha_k} \in \mathcal{V}_k : \alpha_k \in \Xi_k \},$$

para modelar h_k . Tras considerar los modelos para h_k y $\rho_k(g)$, vimos en la Proposición 4.3.2 que

$$U^b(\beta, \tau, \alpha) \equiv M(\beta, h_\alpha) - \sum_{j \in b} \pi_j(\rho_{j, \tau_j}, h_\alpha) - \sum_{j \in \bar{b}} \pi_j(\psi_j \{ \rho_{j, \tau_j}, \beta \}, h_\alpha) \quad (5.8)$$

es una función de estimación múltiple protegida para $\beta(g)$. Para poder construir funciones de estimación insesgadas múltiple protegidas para los parámetros de ruido T_k , recordemos que si $p \in \mathcal{G}_k$, definimos a $\tau_k(g)$ siendo el único valor para el que se satisface que

$$\rho_k(g) = \rho_{k, \tau_k(g)}. \quad (5.9)$$

Vamos entonces ahora a pensar en $(\beta(g), \tau(g))$ como un nuevo funcional de interés definido sobre la familia \mathcal{G}_k y procurar construir una función múltiple protegida para este nuevo objeto, inspirados en la construcción presentada en el capítulo anterior. El siguiente resultado, cuya demostración presentamos en el Apéndice B, sugiere un posible candidato a ser utilizado como función de estimación inicial para $(\beta(g), \tau_k(g))$.

Lema 5.1.1. *Supongamos que se verifica la Condición 3. Sea e_k la aplicación definida en (5.4). Para cada k denotamos con $d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k)$ a un vector aleatorio arbitrario, de dimensión $r_k \equiv \dim(\tau_k)$. Fijado $b \subseteq [K]$ con el que construimos U^b en (5.8), utilizando para cada $k \in [K]$ el funcional $\rho_k(g)$ definido en (5.5), la clase paramétrica $\mathcal{R}_{sub,k}$ y la familia \mathcal{G}_k definidas en (5.6) y en (5.7) respectivamente, definimos*

$$M_k(\beta, \tau_k, h) \equiv \begin{cases} d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \{M(\beta, h) - e_k(\rho_{k, \tau_k}, h)\} & \text{si } k \in b, \\ d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) (M(\beta, h) - e_k\{\psi_k(\rho_{k, \tau_k}, \beta), h\}) & \text{si } k \in \bar{b}. \end{cases} \quad (5.10)$$

Entonces, si $p = gh_1 \dots h_K \in \mathcal{G}_k$

$$E_p(M_k(\beta(g), \tau_k(g), h) | \bar{A}_k, \bar{L}_k) = 0,$$

y por consiguiente, tenemos que

$$E_p(M_k(\beta(g), \tau_k(g), h)) = 0.$$

Si $p \in \mathcal{G}_k$, el lema anterior sugiere utilizar M_k definida en (5.10) como función de estimación insesgada para $(\beta(g), \tau_k(g))$, y a partir de ella construir una ecuación múltiple robusta como presentamos en el capítulo anterior. Una pregunta natural que surge es la siguiente: ¿Puedo garantizar que al proyectar no surgirá la necesidad de modelar nuevos funcionales que inducirían a estimar nuevos parámetros? La respuesta es la siguiente: bajo la Condición 1 y la Condición 3 podremos proyectar la función M_k definida en (5.10) (apropiadamente evaluada) y construir funciones de estimación T_k , tal como enunciaremos en el Lema 5.1.2. Como veremos en los próximos resultados, el éxito de la propuesta radica en que la Condición 3 combinada con el hecho de que cada h_k modela la distribución condicional de A_k dado $(\bar{L}_k, \bar{A}_{k-1})$.

Lema 5.1.2. *Supongamos que se verifica la Condición 1 y la Condición 3. Sea e_k la aplicación definida en (5.4). Para cada k denotamos con $d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k)$ a un vector aleatorio arbitrario, de dimensión r_k donde $r_k \equiv \dim(\tau_k)$. Dado $b \subseteq [K]$ con el cual construimos U^b definida en (5.8), dadas las funciones $\rho_k(g)$ definidas en (5.5) y dadas la clase paramétrica $\mathcal{R}_{sub,k}$ y la familia \mathcal{G}_k definidas en (5.6) y en (5.7) respectivamente, para cada $k \in [K]$ consideramos la función $M_k(\beta, \tau_k, h)$ definida en (5.10). Luego si $p \in \mathcal{G}_k$ tenemos que*

a) para $1 \leq j \leq k$

$$\Pi_p(M_k(\beta(g), \tau_k(g), h) | \Lambda_j(p)) = 0,$$

b) para $k+1 \leq j \leq K$

$$\begin{aligned} \Pi_p(M_k(\beta, \tau_k, h) | \Lambda_j(p)) &= d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \Pi_p(M(\beta, h) | \Lambda_j(p)) \\ &= d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \pi_k(\psi_k\{\eta_k(g), \beta\}, h). \end{aligned}$$

En vista de los resultados recientemente presentados, íntimamente inspirados en la construcción de la función U^b presentada en el capítulo anterior, para cada $k \in [K]$ definimos la función

$$T_k^b(\beta, \tau, \alpha) \equiv M_k(\beta, \tau_k, h_\alpha) - d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \left\{ \sum_{j \in \bar{b}, j > k} \pi_j(\rho_{j, \tau_j}, h_\alpha) + \sum_{j \in \bar{b}, j > k} \pi_j(\psi_j\{\rho_{j, \tau_j}, \beta\}, h_\alpha) \right\} \quad (5.11)$$

Notar que si bien T_k^b depende de b y depende del parámetro τ solo mediante (τ_k, \dots, τ_K) , salvo que se aclare lo contrario, para aliviar la notación ponemos $T_k(\beta, \tau, \alpha)$. El siguiente teorema establece que $T = \{T_k\}_{k \in [K]}$ es una familia de funciones de estimación insesgada para los parámetros de ruido. Antes de enunciar dicho teorema, recordemos que para $k \in [K]$,

$$\mathcal{V}_{sub,k} \equiv \{h_{k, \alpha_k} \in \mathcal{V}_k : \alpha_k \in \Xi_k\}, \quad (5.12)$$

$$\mathcal{H}_k \equiv \{p \in \mathcal{M} : h_k \in \mathcal{V}_{sub,k}\}, \quad (5.13)$$

$$\mathcal{G}_k \equiv \{p \in \mathcal{M} : \rho_k(g) \in \mathcal{R}_{sub,k}\}. \quad (5.14)$$

Proposición 5.1.2. *Supongamos que dadas $p = gh \in \mathcal{M}_c$ y $p^* = gh_\alpha$, con $\alpha_i = \alpha_i(h_i)$ para $i \in c$, p y p^* satisfacen el Supuesto 1, cualquiera sea $c \subseteq [K]$. Asumamos también que se satisface el Supuesto 2 y el Supuesto 3, y que se verifica la Condición 2 y la Condición 3. Sea e_k la aplicación definida en (5.4). Consideramos $\mathcal{V}_{sub,k}$ definido en (5.12). Dado b el subconjunto de $[K]$ con el cual construimos U^b definida en (5.8) sea $\mathcal{R}_{sub,k}$ definido como en (5.6). Sean \mathcal{H}_k y \mathcal{G}_k definidos como en (5.13) y (5.14) respectivamente. Supongamos que para $p \in \cup_{c \subseteq [K]} \mathcal{M}_c$, $(\beta, \tau, \alpha) \in \mathbb{R}^r \times \otimes_{k=1}^K \Upsilon_k \times \otimes_{k=1}^K \Xi_k$ y $k \in [K]$ la función $M_k(\beta, \tau_k, h_\alpha)$ definida en (5.10) está en $L_2(p)$.*

Entonces, dada U^b definida en (5.8), la colección $T = \{T_k\}_{1 \leq k \leq K}$ donde T_k está definida en (5.11) es una familia de funciones de estimación insesgada múltiple protegida en el modelo de unión intersección dado por $\cup_{k=1}^K \{\mathcal{H}_k \cap \mathcal{G}_k\}$. Es decir, para cada $c \subseteq [K]$, $p = gh \in \mathcal{M}_c$, para cada $\tau_c \in \otimes_{k \in c} \Upsilon_k$ y $\alpha_{\bar{c}} \in \otimes_{k \in \bar{c}} \Xi_k$, se verifica que

$$E_p \left(T_{\bar{c}} \left(\beta(g), \tau, \alpha \right) \right) = 0 \quad \text{en} \quad \tau_{\bar{c}} = \tau_{\bar{c}}(g) \quad \alpha_c = \alpha_c(h_c).$$

5.2. Reconstrucción de los procedimientos dados en los ejemplos.

5.2.1. Ejemplo 1

En el ejemplo 1 recordamos que para $p = gh_1h_2$, teníamos

$$E_p(M(\beta, h) | \bar{R}_k \bar{L}_k) = \frac{\prod_{j=1}^k R_j}{\prod_{j=1}^k E_{h_j}(R_j | \bar{R}_{j-1} \bar{L}_j)} \{\nu_k(g) - \beta\}, \quad 1 \leq k \leq 2$$

donde

$$\begin{aligned} \eta_1(g) = \nu_1(g) &\equiv E_g \{E_g(Y | R_2 = R_1 = 1, \bar{L}_2) | R_1 = 1, L_1 = \cdot\}, \\ \eta_2(g) = \nu_2(g) &\equiv E_g(Y | R_2 = R_1 = 1, L_1 = \cdot, L_2 = \cdot) \\ \psi_k(\eta_k(g), \beta) &= \eta_k(g) - \beta. \quad k = 1, 2. \end{aligned}$$

En suma, obtenemos la siguiente representación para cada esperanza condicional.

$$e_1(\psi_1(\eta_1(g), \beta), h) = \frac{R_1}{E_{h_1}(R_1 | L_1)} \psi_1(\eta_1(g), \beta),$$

y

$$e_2(\psi_2(\eta_2(g), \beta), h) = \frac{R_1}{E_{h_1}(R_1 | L_1)} \frac{R_2}{E_{h_2}(R_2 | R_1, \bar{L}_1)} \psi_2(\eta_2(g), \beta).$$

y por lo tanto como vale la condición 2 resulta

$$\begin{aligned} \Pi_p(M(\beta, h) | \Lambda_1(p)) &= \left\{ \frac{R_1}{E_{h_1}(R_1 | L_1)} - 1 \right\} \{E_g \{E_g(Y | R_2 = 1, \bar{L}_2) | R_1 = 1, L_1\} - \beta\}, \\ \Pi_p(M(\beta, h) | \Lambda_2(p)) &= \frac{R_1}{E_{h_1}(R_1 | L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2 | R_1, \bar{L}_1)} - 1 \right\} \{E_g(Y | R_2 = 1, \bar{L}_2) - \beta\}, \end{aligned}$$

y luego

$$\begin{aligned} \pi_1(\psi_1(\eta_1(g), \beta), h) &= \left\{ \frac{R_1}{E_{h_1}(R_1 | L_1)} - 1 \right\} \psi_1(\eta_1(g), \beta) \\ \pi_2(\psi_2(\eta_2(g), \beta), h) &= \frac{R_1}{E_{h_1}(R_1 | L_1)} \left\{ \frac{R_2}{E_{h_2}(R_2 | R_1, \bar{L}_1)} - 1 \right\} \{E_g(Y | R_2 = 1, \bar{L}_2) - \beta\}, \end{aligned}$$

Buscamos recobrar el sistema de ecuaciones presentado en el desarrollo del ejemplo 1 en el Capítulo 2. Tomamos $b = \emptyset$. Tenemos que

- * $\rho_1(g) = \nu_1(g)$,
- * $\mathcal{R}_{sub,1} = \{\rho_{1,\tau_1} = b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\}$,
- * $\rho_2(g) = \nu_2(g)$,
- * $\mathcal{R}_{sub,2} = \{\rho_{2,\tau_2} = b_2(\cdot; \tau_2) : \tau_2 \in \Upsilon_2\}$,

luego

$$U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) \equiv \frac{R_2(Y - b_2(\bar{L}_2, \tau_2))}{E(R_1|L_1; \alpha_1) E(R_2|R_1, \bar{L}_2; \alpha_2)} \\ + \frac{R_1}{E(R_1|L_1; \alpha_1)} \{b_2(\bar{L}_2, \tau_2) - b_1(L_1, \tau_1)\} + \{b_1(L_1; \tau_1) - \beta\}.$$

Además

$$M_2(\beta, \tau_2, \alpha) = d_2(\bar{R}_2, \bar{L}_2, \alpha, \tau_2) \left\{ \frac{R_2(Y - \beta)}{E(R_2|R_1, \bar{L}_2, \alpha_2) E(R_1|L_1, \alpha_1)} - \frac{R_2(b_2(\bar{L}_2, \tau_2) - \beta)}{E(R_2|R_1, \bar{L}_2, \alpha_2) E(R_1|L_1, \alpha_1)} \right\}, \\ = d_2(\bar{R}_2, \bar{L}_2, \alpha, \tau_2) \left\{ \frac{R_2(Y - b_2(\bar{L}_2, \tau_2))}{E(R_2|R_1, \bar{L}_2, \alpha_2) E(R_1|L_1, \alpha_1)} \right\},$$

luego

$$T_2(\beta, \tau, \alpha) = d_2(\bar{R}_2, \bar{L}_2, \alpha, \tau_2) \left\{ \frac{R_2}{E(R_2|R_1, \bar{L}_2, \alpha_2) E(R_1|L_1, \alpha_1)} (Y - b_2(\bar{L}_2, \tau_2)) \right\}.$$

Para encontrar a T_1 , tenemos que

$$M_1(\beta, \tau_1, \alpha) = d_1(R_1, L_1, \alpha, \tau_1) \left\{ \frac{R_2}{E(R_2|R_1, \bar{L}_2, \alpha_2) E(R_1|L_1, \alpha_1)} (Y - \beta) - \frac{R_1(b_1(L_1, \tau_1) - \beta)}{E(R_1|L_1, \alpha_1)} \right\}$$

luego

$$T_1(\beta, \tau, \alpha) = \frac{R_1 d_1(R_1, L_1, \alpha, \tau_1)}{E(R_1|L_1, \alpha_1)} \left(\frac{R_2 \{Y - b_2(\bar{L}_2; \tau_2)\}}{E(R_2|R_1, \bar{L}_2; \alpha_2)} + \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} \right)$$

En suma tenemos que para $b = \emptyset$, el estimador $\hat{\beta}_{MP}$ es la solución en (β, τ, α) del sistema $\mathbb{P}_n[U(\beta, \tau, \alpha)] = \mathbb{P}_n[T_2(\beta, \tau_2, \alpha)] = \mathbb{P}_n[T_1(\beta, \tau, \alpha)] = 0$ para

i)

$$U(\beta, \tau_1, \tau_2, \alpha_1, \alpha_2) = \frac{R_2 (Y - b_2(\bar{L}_2, \tau_2))}{E(R_1|L_1; \alpha_1) E(R_2|R_1, \bar{L}_2; \alpha_2)} + \frac{R_1}{E(R_1|L_1; \alpha_1)} \{b_2(\bar{L}_2, \tau_2) - b_1(L_1, \tau_1)\} + \{b_1(L_1, \tau_1) - \beta\}.$$

ii)

$$T_2(\beta, \tau, \alpha) = d_2(\bar{R}_2, \bar{L}_2, \alpha, \tau_2) \left\{ \frac{R_2}{E(R_2|R_1, \bar{L}_2, \alpha_2) E(R_1|L_1, \alpha_1)} (Y - b_2(\bar{L}_2, \tau_2)) \right\}.$$

iii)

$$T_1(\beta, \tau, \alpha) = \frac{R_1 d_1(R_1, L_1, \alpha, \tau_1)}{E(R_1|L_1, \alpha_1)} \left(\frac{R_2 \{Y - b_2(\bar{L}_2; \tau_2)\}}{E(R_2|R_1, \bar{L}_2; \alpha_2)} + \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} \right)$$

por lo tanto, para recobrar el procedimiento de estimación para $\beta(g)$ presentado en el capítulo 2 es suficiente tomar

$$\star d_2(\bar{R}_2, \bar{L}_2, \alpha, \tau_2) = E(R_1|L_1, \alpha_1) E(R_2|R_1, \bar{L}_1, \alpha_2) \frac{\partial}{\partial \tau_2} b_2(\bar{L}_2, \tau_2).$$

$$\star d_1(R_1, L_1, \alpha, \tau_1) = E(R_1|L_1, \alpha_1) d(L_1, \tau, \alpha_2), \text{ con } d(L_1, \tau, \alpha_2) \text{ una función vectorial arbitraria de la misma dimensión que } \tau_1.$$

5.2.2. Ejemplo 2

En el ejemplo 2 para $p = gh_1h_2$ teníamos

$$E_p(M(\beta, h) | W, X, V, R) = \frac{R}{E_{h_2}(R|X, V, W)} (\nu_2(g) - \beta X) \Delta_{h_1}(X, W),$$

$$E_p(M(\beta, h) | W, X) = (\nu_1(g) - (\beta - \beta(g))X) \Delta_{h_1}(X, W),$$

con

$$\nu_1(g) \equiv E_g(E_g(Y - \beta(g)X | R = 1, V, X, W) | X, W) \equiv e_g(W),$$

$$\nu_2(g) \equiv E_g(Y | R = 1, X = \cdot, V = \cdot, W = \cdot),$$

$$\Delta_{h_1}(X, W) \equiv \{X - E_{h_1}(X|W)\} q(W),$$

$$\star \eta_1(g) = (\eta_{1,1}, \eta_{1,2}) \text{ con } \eta_{1,1}(g) = \nu_1(g) \text{ y } \eta_{1,2}(g) = \beta(g)$$

- * $\psi_1(\eta_1(g), \beta) = (\psi_{1,1}, \psi_{1,2})$ con $\psi_{1,1}(\eta_1(g), \beta) = \eta_{11}(g)$ y $\psi_{1,2}(\eta_1(g), \beta) = \eta_{12}(g) - \beta$
- * $\eta_2(g) = \nu_2(g)$
- * $\psi_2(\eta_2(g), \beta) = \eta_2(g) - \beta X$.

de donde obtenemos la siguiente representación para cada esperanza condicional

$$e_1(\psi_1(\eta_1(g), \beta), h) = (\psi_{1,1}(\eta_1(g), \beta) + \psi_{1,2}(\eta_1(g), \beta) X) \{X - E_{h_1}(X|W)\} q(W),$$

y

$$e_2(\psi_2(\eta_2(g), \beta), h) = \frac{R}{E_{h_2}(R|W, X, V)} \psi_2(\eta_2(g), \beta) \{X - E_{h_1}(X|W)\} q(W).$$

con lo cual

1. $\pi_1(\psi_1(\eta_1(g), \beta), h) = (\psi_{1,1}(\eta_1(g), \beta) + \psi_{1,2}(\eta_1(g), \beta) E_{h_1}(1 - X|W)) \Delta_{h_1}(X, W)$
2. $\pi_2(\psi_2(\eta_2(g), \beta), h) = \left\{ \frac{R}{E_{h_2}(R|W, X, V)} - 1 \right\} \psi_2(\eta_2(g), \beta) \Delta_{h_1}(X, W)$

Buscamos recobrar el sistema de ecuaciones presentado en el desarrollo del ejemplo 1 en el Capítulo 2.

Tomamos $b = \{1\}$. Tenemos que

- * $\rho_1(g) = (\nu_1(g), 0)$, abusando de la notación, utilizaremos en adelante $\rho_1(g) = \nu_1(g)$.
- * $\mathcal{R}_{sub,1} = \{\rho_{1,\tau_1} = b_1(\cdot; \tau_1) : \tau_1 \in \Upsilon_1\}$,
- * $\rho_2(g) = \nu_2(g)$,
- * $\mathcal{R}_{sub,2} = \{\rho_{2,\tau_2} = b_2(\cdot; \tau_2) : \tau_2 \in \Upsilon_2\}$,

luego

$$U(\beta, \tau, \alpha) = \left\{ \frac{R \{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1) \right\} \Delta_{\alpha_1}(X, W).$$

Además

$$T_2(\beta, \tau_2, \alpha) = d_2(X, W, R, V, \alpha, \tau_2) \frac{R}{E(R|X, V, W, \alpha_2)} (Y - b_2(W, X, V; \tau_2)) \Delta_{\alpha_1}(X, W)$$

Por otro lado

$$M_1(\beta, \tau_1, \alpha) = \left\{ \frac{R}{E(R|X, V, W, \alpha_2)} (Y - \beta X) - b_1(W, \tau_1) \right\} \Delta_{\alpha_1}(X, W),$$

luego

$$T_1(\beta, \tau_1, \alpha) = d_1(W, X, V, \alpha, \tau) \left\{ \frac{R \{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1) \right\} \Delta_{\alpha_1}(X, W).$$

En suma, para $b = \{1\}$

$$U(\beta, \tau, \alpha) = \left\{ \frac{R \{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1) \right\} \Delta_{\alpha_1}(X, W)$$

$$T_2(\beta, \tau_2, \alpha) = d_2(W, X, V, R, \alpha, \tau) \frac{R}{E(R|X, V, W, \alpha_2)} (Y - b_2(W, X, V; \tau_2)) \Delta_{\alpha_1}(X, W)$$

$$T_1(\beta, \tau_1, \alpha) = d_1(W, X, \alpha, \tau) \left\{ \frac{R \{Y - b_2(W, X, V; \tau_2)\}}{E(R|W, X, V; \alpha_2)} + b_2(W, X, V; \tau_2) - \beta X - b_1(W; \tau_1) \right\} \Delta_{\alpha_1}(X, W).$$

Luego, para recobrar el procedimiento de estimación presentado en el capítulo 2 es suficiente tomar

$$\star q(W) = 1,$$

$$\star d_2(W, X, V, R, \alpha, \tau) = \frac{\partial}{\partial \tau_2} b_2(X, V, W; \tau_2) E(R|X, V, W, \alpha_2) \Delta_{\alpha_1}(X, W)^{-1},$$

$$\star d_1(W, X, \alpha, \tau) = \{X - E(X|W; \alpha_1)\} \frac{\partial b_1}{\partial \tau_1}(W, \tau_1).$$

Capítulo 6

Problemas abiertos

Como anticipamos en la introducción, en esta tesis presentamos una teoría que brinda condiciones suficientes bajo las cuales es posible construir ecuaciones de estimación múltiple protegidas. Recordamos que en el Capítulo 2 presentamos dos modelos inferenciales describiendo en cada uno de ellos diferentes estimadores, incluyendo procedimientos múltiple protegidos. En el primer ejemplo trabajamos sobre un modelo basado en hacer una pequeña modificación del caso estudiado por Tchetgen Tchetgen (2009), mientras que en el segundo ejemplo, se procura estimar la componente paramétrica bajo un modelo de regresión parcialmente lineal, ante la presencia de respuestas faltantes. La propuesta de estimación desarrollada en este trabajo, permite reconstruir en los ejemplos las funciones de estimación que dan lugar al estimador múltiple protegido presentado en cada uno de ellos.

El éxito de este procedimiento radica en que pudimos dar condiciones suficientes bajo las cuales estimar a los parámetros de ruido en forma múltiple protegida.

A pesar de los avances aportados por esta tesis, aún quedan muchos problemas abiertos, que discutiremos a lo largo de este capítulo. Planeamos investigar estos problemas en el futuro cercano.

6.1. Aplicación del procedimiento múltiple protegido propuesto en este trabajo a otros modelos inferenciales.

La propuesta de estimación desarrollada en este trabajo, permite reconstruir en los ejemplos dados en el Capítulo 2 las funciones de estimación que dan lugar al estimador múltiple protegido presentado en cada uno de ellos. En un principio, habíamos incluido un tercer ejemplo en nuestro trabajo. El ejemplo fue el estudiado por Vansteelandt S,

VadderWeele T, Tchetgen Tchetgen E, Robins J. (2008), quienes con el objetivo de decidir si hay un efecto de la interacción entre dos variables sobre una determinada variable respuesta Y , consideran un modelo semiparamétrico para la esperanza condicional de Y , que postula la interacción estadística en términos de un parámetro finito dimensional. Bajo ese modelo, los autores presentan estimadores triple protegidos para dicho parámetro. Al aplicar nuestra metodología de estimación a este ejemplo, específicamente cuando aplicamos la fórmula (5.11) a este modelo, encontramos que las ecuaciones construidas con nuestro método eran distintas a las propuestas por los autores. Es decir, nuestro procedimiento brinda un mecanismo de estimación de los parámetros del modelo estudiado por Vansteelandt et al (2008) diferente al que ellos presentaron. Para ser más precisos, la diferencia entre ambos métodos radica en la manera en la que los parámetros de ruido son estimados en cada caso; las ecuaciones para estimar a los parámetros de ruido que se construyen en este trabajo difieren de las presentadas por los mencionados autores.

Esperamos poder generalizar los resultados obtenidos en esta tesis para así recuperar las ecuaciones de estimación propuestas por Vansteelandt et al (2008) o entender estructuralmente como generalizar el procedimiento por ellos presentado.

6.2. Mejorando el desempeño en términos de eficiencia de los estimadores múltiple protegidos construidos en este trabajo.

Si bien en este trabajo no tratamos cuestiones relativas a la eficiencia de los estimadores estudiados, este tema ha sido abordado por diferentes autores. Mencionaremos algunos resultados existentes en el contexto del ejemplo presentado en la introducción, para explicar de que manera esperamos poder generalizar dichos resultados.

En la introducción presentamos el problema de estimar la esperanza de una variable aleatoria Y en presencia de datos faltantes bajo el supuesto de MAR. En ese marco, recordamos que construimos tres estimadores:

1. $\hat{\beta}_{reg}$ verificando

$$\hat{\beta}_{reg} = \mathbb{P}_n (b(L, \hat{\tau})),$$

donde $\hat{\tau}$ es un estimador de τ_0 bajo el modelo

$$E(Y|L, R = 1) = b(L, \tau_0). \quad (6.1)$$

2. $\hat{\beta}_{IPW}$ siendo la solución en β de la ecuación

$$\mathbb{P}_n \left(\frac{R}{\pi(L, \hat{\alpha})} (Y - \beta) \right) = 0,$$

donde $\hat{\alpha}$ es el estimador de máxima verosimilitud de α_0 bajo el modelo

$$P(R = 1|L) = \pi(L, \alpha_0). \quad (6.2)$$

3. $\hat{\beta}_{DP}$ definido como la solución en β de la ecuación

$$\mathbb{P}_n \left(\frac{R}{\pi(L, \hat{\alpha})} (Y - \beta) - \left(\frac{R}{\pi(L, \hat{\alpha})} - 1 \right) (b(L, \hat{\tau}) - \beta) \right) = 0, \quad (6.3)$$

donde $\hat{\alpha}$ es el estimador de máxima verosimilitud de α_0 bajo el modelo (6.2) y $\hat{\tau}$ es un estimador de τ_0 construido bajo el modelo (6.1).

Recordamos también que un estimador $\hat{\beta}_{DP}$ así construido resulta doble protegido, es decir resulta consistente en la unión de los modelos (6.2) y (6.1).

La construcción del estimador $\hat{\beta}_{DP}$ utiliza siempre el mismo procedimiento para estimar a α_0 : su estimador de máxima verosimilitud. Sin embargo, cualquier estimador $\hat{\tau}$ de τ_0 consistente bajo el modelo (6.1) puede ser utilizado. Para destacar la dependencia en $\hat{\tau}$ del estimador $\hat{\beta}_{DP}$, de ahora más también escribiremos $\hat{\beta}_{DP}(\hat{\tau})$ para denotar a $\hat{\beta}_{DP}$. Tenemos entonces que distintas formas para estimar al parámetro τ_0 dan origen a diferentes estimadores doble protegidos para $E[Y]$.

Cuando el modelo (6.1) es incorrecto, la varianza asintótica de $\hat{\beta}_{DP}(\hat{\tau})$ depende del valor límite $\hat{\tau}$. Es decir, si $\hat{\tau}$ converge a τ^ℓ , utilizaremos $\sigma^2(\tau^\ell)$ para denotar a la varianza asintótica de $\hat{\beta}(\hat{\tau})$. Cabe mencionarse que, bajo condiciones de regularidad, en este mismo escenario el estimador $\hat{\beta}(\hat{\tau})$ es asintóticamente equivalente al estimador $\hat{\beta}(\tau^\ell)$.

Kang y Schafer (2007) notaron que el estimador $\hat{\beta}_{DP}(\hat{\tau})$ puede ser altamente ineficiente si el modelo (6.1) es incorrecto. Tan (2008) y Cao, Tsiatis y Davidian (2009), propusieron de manera independiente un mismo estimador que, de alguna manera, subsana este problema. La propuesta, consiste en utilizar $\tilde{\tau}$ convergiendo al valor τ_{opt}^ℓ , que minimice la varianza asintótica $\sigma^2(\tau^\ell)$. Tal sucesión se obtiene minimizando la versión empírica de la varianza asintótica. El estimador $\hat{\beta}_{DP}(\tilde{\tau})$, es decir, el estimador que se obtiene reemplazando en la ecuación (6.3) a $\hat{\tau}$ por $\tilde{\tau}$ así construido, tiene las siguientes propiedades:

- (a) $\hat{\beta}_{DP}(\tilde{\tau})$ es doble protegido.
- (b) si el modelo para $P(R = 1|L)$, dado por (6.2), es correcto pero el modelo (6.1) para $E(Y|L, R = 1)$ es incorrecto, entonces $\hat{\beta}_{DP}(\tilde{\tau})$ tiene varianza asintótica menor o igual que la varianza asintótica de $\hat{\beta}_{DP}(\tau)$, para todo valor de τ .

Con esta construcción, los autores brindan un estimador doble protegido que, cuando el modelo para $P(R = 1|L)$ (h_1 en la teoría general) esta bien especificado, es el más eficiente en la clase de estimadores de la forma $\hat{\beta}_{DP}(\hat{\tau})$.

Recientemente, como mencionamos en la introducción, Rotnitzky et. al. (2012), propusieron estimadores de $\beta_0 \in \mathbb{R}^d$ en el modelo de regresión dado por

$$E[Y|X] = h(X, \beta_0),$$

donde X es un subvector de L , los datos observados conforman una muestra de (L, R, Y) , siendo que R indica si la variable Y es observada, y asumen que se verifica el supuesto MAR: $Y|R = 1, L \sim Y|L$. El problema en consideración se enmarca en la teoría presentada en esta tesis para el caso $K = 1$. Es decir, la distribución de los datos puede escribirse como $p = gh_1$ y $\beta_0 = \beta(g)$, permitiendo así construir estimadores doble protegidos que denotaremos con $\widehat{\beta}_{DP}(\widehat{\tau})$, siguiendo la notación presentada en la sección anterior. Sin embargo, como el parámetro de interés pertenece a \mathbb{R}^d , cuando el modelo para h_1 está bien especificado, la *varianza asintótica* de $\widehat{\beta}_{DP}(\widehat{\tau})$ resulta una matriz, que denotaremos con $\Sigma(\tau^\ell)$, siendo τ^ℓ el límite de $\widehat{\tau}$. En el mencionado trabajo, los autores construyen un estimador doble protegido de la forma $\widehat{\beta}_{DP}(\widehat{\tau})$, de forma tal que la varianza asintótica de $\phi_j(\widehat{\beta}_{DP}(\widehat{\tau}))$ es menor o igual que la varianza asintótica de $\phi_j(\widehat{\beta}_{DP}(\tau))$, para ciertas funciones ϕ_1, \dots, ϕ_T tomando valores en \mathbb{R} , especificadas por el analista, siempre que el modelo h_1 sea correcto. Más aún, cuando h_1 está bien especificada, $\widehat{\beta}_{DP}(\widehat{\tau})$ es más eficiente que el estimador $\widehat{\beta}_{IPW}$ construido bajo este escenario.

Considerando la teoría desarrollada a lo largo de este trabajo, esperamos poder en una segunda instancia construir estimadores multiple protegidos mejorados, en el sentido descrito en los trabajos mencionados. Más específicamente, cuando $p \in \mathcal{M}_c$, la varianza de los estimadores $\phi_j(\widehat{\beta}_{DP}(\widehat{\tau}))$ dependerá de α_c^ℓ y de τ_c^ℓ , los límites de $\widehat{\alpha}_c$ y $\widehat{\tau}_c$, respectivamente. El valor $\widehat{\alpha}_c$ queda unívocamente determinado por el (único) procedimiento utilizado para estimar los parámetros que indexan los modelos para h_i . Sin embargo, una vez más, el valor de τ_c^ℓ depende de las ecuaciones utilizadas para construir $\widehat{\tau}$ (mas, generalmente, del sistema que determina $(\widehat{\beta}_{MP}, \widehat{\alpha}, \widehat{\tau})$). El problema por estudiar consiste entonces en encontrar un estimador múltiple protegido $\widehat{\beta}_{MP}(\widehat{\tau})$, de forma tal que cuando $p \in \mathcal{M}_c$, la varianza asintótica de $\phi_j(\widehat{\beta}_{DP}(\widehat{\tau}))$ sea menor o igual a la de $\phi_j(\widehat{\beta}_{DP}(\tau_c(g), \tau_c))$, para todo τ_c , cualquiera sea $c \subseteq [K]$ y que cuando todos los modelos para h_i estén bien especificados, sea mas eficiente que $\widehat{\beta}_{IPW}$ (comparando matrices de varianza-covarianza asintótica).

6.3. Estimadores de tipo Plug-In.

En la sección (3.1.1), definimos los funcionales sobre los cuales construimos estimadores múltiple protegidos usando una cierta función M . Específicamente, con el fin de definir el funcional de interés, consideramos para cada $h \in \otimes_{k=1}^K \mathcal{V}_k^{np}$ y para cada $\beta \in \mathbb{R}^r$ la función $m(-; \beta; h) : \mathcal{Z} \rightarrow \mathbb{R}^r$. Para enfatizar la dependencia en los paráme-

tros, utilizamos $M(\beta; h) \equiv m(Z; \beta; h)$. Sea $\mathcal{D} \subseteq \mathcal{P}$ un conjunto de distribuciones de forma tal que para todo $p = gh \in \mathcal{D}$ se tiene que

1. $E_{gh}[|M(\beta, h)|] < \infty$, para todo $\beta \in \mathbb{R}^r$.
2. Existe una única solución de la ecuación

$$E_{gh}[M(\beta, h)] = 0. \quad (6.4)$$

Utilizamos $\beta(p)$ para denotar al funcional que a $p = gh \in \mathcal{D}$ le asigna el único elemento de \mathbb{R}^r que resuelve la ecuación (6.4). Más aún, supusimos que existen \mathcal{U} y \mathcal{V}_k , para $k \in [K]$, subconjuntos de \mathcal{U}^{np} y \mathcal{V}_k^{np} , respectivamente, de forma tal que

$$\mathcal{M} \equiv \left\{ p = gh_1 h_2 \dots h_K : g \in \mathcal{U}, h_k \in \mathcal{V}_k \text{ para } k \in [K], p(z) \geq 0 \int p d\mu = 1 \right\} \quad (6.5)$$

está incluido en \mathcal{D} , y si $p = gh \in \mathcal{M}$, resulta que $\beta(p)$ depende en p solo mediante g . En otras palabras, se verifica que

$$\text{si } p = gh \text{ y } \tilde{p} = g\tilde{h} \text{ pertenecen a } \mathcal{M}, \text{ entonces } \beta(p) = \beta(\tilde{p}). \quad (6.6)$$

Luego el funcional $\beta(\cdot)$ está definido en el espacio \mathcal{D} y depende de $p = gh$ solo mediante g si $p \in \mathcal{M}$.

En el presente contexto estimamos a $\beta(p)$ suponiendo que la distribución de Z es un elemento de \mathcal{M} . He hicimos un pequeño abuso al denotar a $\beta(p)$ mediante $\beta(g)$ siempre que $p \in \mathcal{M}$. Para funcionales de este tipo, dado $p = gh \in \mathcal{M}$, diremos que un estimador $\hat{\beta}$ es de tipo Plug-In si $\hat{\beta} = \beta(\hat{g})$, con $\hat{g} \in \mathcal{U}$. Los estimores así constuidos tienen la propiedad de tomar valores en el rango del funcional de interés.

Lamentablemente, los estimadores $\hat{\beta}_{MP}$ que surgen de aplicar nuestra teoría pueden tomar valores fuera del rango del funcional que queremos estimar. Por ejemplo, si Y es binaria, en el ejemplo 1, $\beta(g) \in [0, 1]$ para toda distribución g mientras que $\hat{\beta}_{MP}$ no tiene por que ser un número entre 0 y 1. Sin embargo, si de alguna manera pudieramos contruir un estimador $\hat{\beta}_{MP}$ multiple protegido que coincida con $\beta(\hat{g})$, para cierta \hat{g} , garantizaríamos que el estimador toma valores en el mismo espacio donde lo hace el funcional $\beta(\cdot)$.

Rotnitzky et al (2012), propusieron un estimador que, además de ser doble protegido y tener las propiedades de eficiencia mencionadas en la sección anterior, toma valores en el rango del funcional estimado.

Un punto a desarrollar sería tratar de adaptar las ideas de Rotnitzky et. al. (2012) para construir un estimador múltiple protegido que a la vez sea plug-in en el marco general donde se desarrolla la teoría de esta tesis.

6.4. Relajando algunas hipótesis.

Como dijimos anteriormente, uno de los principales aportes de este trabajo consistió en presentar un procedimiento de estimación múltiple protegido para los parámetros de ruido asociados a los submodelos de reducción. Dicho procedimiento requirió de ciertas restricciones sobre la distribución que genera los datos. Específicamente, pedimos que las funciones h_k sean distribuciones condicionales. Una pregunta que queda pendiente es si existen otras condiciones más débiles que garanticen los mismos resultados.

6.5. Estimación múltiple protegida en modelos para medidas repetidas con non-ignorable non-response.

Actualmente Vansteelandt S, Rotnitzky A y Robins J, están trabajando en propuestas de estimación múltiple protegidas en modelos para medidas repetidas con non-ignorable non-response. En este problema, la versomilitud se puede escribir como el producto de factores h_k y g_k , $k = 0, 1, \dots, K$ que varían en forma independiente y tales que cada g_k y cada h_k es una distribución condicional. Sin embargo, el parámetro de interés depende no solo de g_k sino también de h_k . Por lo tanto el estimador múltiple protegido en este problema no surge como resultado de la teoría desarrollada en esta tesis. Un problema abierto es desarrollar una teoría que provea condiciones suficientes generales para la existencia de estimadores múltiple protegidos en situaciones como las contempladas en el trabajo de Vansteelandt, Rotnitzky y Robins.

Capítulo 7

Apéndice A: Espacio Tangente

El principal objetivo de este capítulo es introducir los conceptos de estadística semi-paramétrica utilizados a lo largo del trabajo. Fundamentalmente, nos interesa desarrollar la noción de espacio tangente. Para ello, comenzaremos definiendo diferentes modelos para la distribución de un elemento aleatorio. Nos concentraremos luego en los modelos paramétricos, recordando la noción de score, para poder definir conjuntos tangentes. Estudiaremos la estructura de los conjuntos tangentes en modelos que admiten factorizaciones, como las presentadas en los modelos considerados a lo largo de esta tesis y daremos demostraciones rigurosas bajo ciertas hipótesis de los Supuestos introducidos en el Capítulo 4.

7.1. Modelos para la distribución de un elemento aleatorio.

A lo largo de este capítulo, utilizaremos Z para denotar un elemento aleatorio tomando valores en cierto espacio medible $(\mathcal{Z}, \mathcal{A})$, y con P_Z denotaremos la distribución (o probabilidad inducida por) Z .

A partir de un conjunto de observaciones $Z_1, Z_2 \dots Z_n$ independientes e idénticamente distribuidas, (i.i.d.), con la misma distribución que Z , la inferencia estadística procura tomar decisiones que involucren a P_Z , asumiendo que pertenece a una clase \mathcal{F} , formada por ciertas medidas de probabilidad sobre el espacio $(\mathcal{Z}, \mathcal{A})$.

Si la clase \mathcal{F} no impone restricciones sobre el conjunto de medidas de probabilidad que la forman diremos que \mathcal{F} es un modelo no paramétrico para P_Z . También se dice que una clase \mathcal{F} formada por ciertas medidas de probabilidad sobre el espacio $(\mathcal{Z}, \mathcal{A})$ es un modelo no paramétrico para P_Z cuando la misma impone restricciones de suavidad sobre algún funcional infinito dimensional vinculado a P_Z como ser alguna esperanza

condicional o densidad. La razón por la cual estas dos definiciones no son incompatibles en la teoría semiparamétrica es porque restricciones de suavidad no alteran el espacio tangente maximal del modelo que no impone restricciones. Cabe mencionar que abordaremos la noción de espacio tangente maximal en este mismo capítulo. Cuando la clase \mathcal{F} verifica

$$\mathcal{F} = \{P_\theta : \theta \in \Theta\}, \quad (7.1)$$

con Θ subconjunto de un espacio vectorial y la dimensión de Θ es finita diremos que \mathcal{F} es un modelo paramétrico para P_Z . Si un modelo \mathcal{F} no es ni paramétrico ni no paramétrico diremos que es semiparamétrico. En adelante, asumiremos que existe una medida μ definida en $(\mathcal{Z}, \mathcal{A})$ que domina a todos los elementos en la clase \mathcal{F} ; es decir, para todo $\theta \in \Theta$, P_θ es absolutamente continua respecto de μ y utilizaremos p_θ para denotar a la derivada de Radon-Nikodym de P_θ respecto de μ : $p(z; \theta) = (dP_\theta/d\mu)(z)$. Si \mathcal{F} verifica la representación dada en (7.1) con $\theta = (\beta, \eta) \in \Theta$, $\Theta \subseteq \Gamma \times \Delta$, siendo Γ un subconjunto de \mathbb{R}^k , mientras que Δ es un subconjunto de un espacio vectorial de dimensión infinita, entonces, \mathcal{F} es un modelo semiparamétrico y decimos que β es la componente paramétrica del modelo. Por lo general, si un modelo se representa como en (7.1) sea el caso paramétrico o semiparamétrico, el objeto de interés depende de la parte paramétrica del modelo, pudiendo ser el propio vector o alguna de sus coordenadas. Consideremos los siguiente ejemplos:

a) Modelos Paramétricos

Ejemplo 1: distribución normal con varianza conocida. Supongamos que Z es una variable aleatoria con densidad p , perteneciente a la familia normal con $\sigma^2 = 1$. Es decir, asumimos que existe θ_0 de forma tal que $p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-\theta_0)^2}{2}}$. Luego, un modelo paramétrico para p es:

$$\mathcal{F}_1 = \left\{ p(z, \theta) : p(z, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-\theta)^2}{2}} \quad \theta \in \mathbb{R} \right\}.$$

En este caso $\Theta = \mathbb{R}$. Consideramos como objeto de interés a la esperanza de Z . En este ejemplo el parámetro del modelo coincide con el objeto de estudio.

Ejemplo 2: distribución normal con varianza desconocida. Supongamos que Z es una variable aleatoria con densidad p en la familia normal, donde desconocemos tanto el parámetro de posición como el de escala.

Es decir, ahora asumimos que $p(z) = \frac{1}{\sqrt{2\pi\theta_2^2}} e^{-\frac{(z-\theta_1,0)^2}{2\theta_2,0^2}}$. Luego, un modelo paramétrico para p viene dado por

$$\mathcal{F}_2 = \left\{ p(z, \theta) : p(z, \theta) = \frac{1}{\sqrt{2\pi\theta_2^2}} e^{-\frac{(z-\theta_1)^2}{2\theta_2^2}}, \quad \theta = (\theta_1, \theta_2), \text{ donde } \theta_1 \in \mathbb{R}, \theta_2 \in \mathbb{R}^+ \right\},$$

En este caso $\Theta = \mathbb{R} \times \mathbb{R}^+ \subseteq \mathbb{R}^2$. Si nuevamente consideramos como objeto de interés la esperanza de Z , resulta que la misma coincide con la primer coordenada del parámetro del modelo propuesto para la distribución de Z .

b) Modelos Semiparamétricos

Ejemplo 3: distribuciones simétricas. Sea Z es una variable aleatoria con densidad p . Asumamos que $p(z) = h_0(z - \beta_0)$, donde β_0 es un número real desconocido mientras que h_0 es una función de densidad desconocida de la cual solo se sabe que es simétrica alrededor del cero. Luego un modelo para p viene dado por

$$\mathcal{F}_3 = \{h(z - \beta) : h(z) = h(-z), \beta \in \mathbb{R}\}.$$

El parámetro del modelo resulta ser $\theta = (\beta, h) \in \mathbb{R} \times \mathcal{S}$, donde

$$\mathcal{S} = \{h : h : \mathbb{R} \rightarrow \mathbb{R}, h(z) = h(-z), h \text{ densidad}\}.$$

Tenemos entonces que el modelo esta indexado por una componente paramétrica (β) y una componente perteneciente a un espacio de dimensión infinita, razón por la cual \mathcal{F}_3 es un modelo semiparamétrico. Por lo general, bajo este modelo el objeto de interés es el centro de simetría de la distribución que genera los datos, coincidiendo con la componente paramétrica del modelo.

Ejemplo 4: modelo de regresión. Supongamos que Z es un vector aleatorio, $Z = (X, Y)$ con densidad conjunta dada por p . Asumamos el modelo de regresión paramétrico, dado por

$$Y = \sigma(X, \beta_0) + \varepsilon \quad \text{donde} \quad E[\varepsilon|X] = 0$$

donde σ es una función conocida y $\beta_0 \in \mathbb{R}^k$. En tal caso, decimos que σ es la función de regresión mientras que nos referimos a β_0 como el parámetro de la regresión. Notemos que $p(x, y) = p_{(X, \varepsilon)}(x, y - \sigma(x, \beta_0))$. Luego, podemos parametrizar la distribución del vector utilizando la densidad marginal de X , la densidad condicional de $\varepsilon|X$ y el parámetro de la regresión. De esta forma, tenemos un modelo semiparamétrico para p , dado por

$$\mathcal{F}_4 = \{f_{(X, Y)}(x, y) = f_X(x)f_{\varepsilon|X=x}(y - \sigma(x, \beta)) : \beta \in \mathbb{R}^k, f_X \in \Gamma_1, f_{\varepsilon|X} \in \Gamma_2\},$$

siendo

$$\Gamma_1 = \{f_X(x) : f_X \text{ es una función de densidad}\},$$

$$\Gamma_2 = \{f_{\varepsilon|X}(\varepsilon) : \text{para cada } x, f_{\varepsilon|X=x} \text{ es una función de densidad con media cero}\}.$$

Tenemos entonces que el modelo está parametrizado por $(\beta, f_X, f_{\varepsilon|X}) \in \mathbb{R}^k \times \Gamma_1 \times \Gamma_2$, explicando la naturaleza semiparamétrica del mismo. Por lo general, en este caso el objeto de interés es el parámetro de regresión.

c) **Modelos no paramétricos.**

Ejemplo 5: Supongamos que Z es una variable aleatoria de la que solo se sabe que tiene densidad p . Un modelo no paramétrico para p sería considerar

$$\mathcal{F}_5 = \{f : f \text{ es una función de densidad}\}.$$

d) **Modelos en la generalización de los ejemplos.**

En la Sección 3.1.1 del Capítulo 3, consideramos a la familia \mathcal{P} conformada por medidas de probabilidad en $(\mathcal{Z}, \mathcal{A})$ dominadas por una cierta medida μ , cuyas densidades pueden escribirse como el producto de $K + 1$ factores que varían en forma independiente. Es decir

$$\mathcal{P} \equiv \left\{ p = gh_1h_2\dots h_K : g \in \mathcal{U}^{np}, h_k \in \mathcal{V}_k^{np}, p(z) \geq 0 \text{ y } \int pd\mu = 1 \right\},$$

para ciertos conjuntos $\mathcal{U}^{np}, \mathcal{V}_k^{np}, k \in [K]$. Resulta que \mathcal{P} es un modelo no paramétrico para la densidad de los datos observados.

7.1.1. Modelos Paramétricos.

Vamos ahora a restringir nuestra atención a los modelos paramétricos. Es decir, consideraremos una familia $\mathcal{F} = \{p(z; \theta) : \theta \in \Theta\}$, siendo Θ un subconjunto de \mathbb{R}^k , para algún k . Denotemos con $L_2(\mu)$ al conjunto de funciones medibles $f : \mathcal{Z} \rightarrow \mathbb{R}$ de cuadrado integrable respecto de la medida μ . Utilizaremos $E_\theta[g(Z)]$ para denotar la esperanza de la función $g(z)$ respecto de la medida P_θ .

La teoría que desarrollaremos a continuación es local, entendiendo que las definiciones dadas se refieren a un valor θ fijo en el interior del espacio Θ .

Cuando $p(z; \theta)$ es una función derivable en θ , bajo ciertas condiciones, el estimador de máxima verosimilitud satisface la ecuación $\mathbb{P}_n[\dot{\ell}(Z; \hat{\theta}_n)] = 0$, donde $\dot{\ell}(z; \theta)$ es el vector en \mathbb{R}^k , definido por

$$\dot{\ell}(z; \theta) = \frac{\partial \ln(p(z; \theta))}{\partial \theta}.$$

En este mismo contexto, la matriz de información de Fisher está dada por

$$I(\theta) = E_\theta \left(\dot{\ell}(Z, \theta) \dot{\ell}^t(Z; \theta) \right).$$

Tanto el score como la matriz de información de Fisher juegan un papel fundamental en la teoría de eficiencia y pueden definirse en un escenario mas general; basta con que la aplicación $\theta \rightarrow p(-; \theta)$ sea diferenciable en media cuadrática. Es decir, diremos que

la aplicación $\theta \rightarrow p(-; \theta)$ es diferenciable (o derivable) en media cuadrática si existe un vector $S(z; \theta)$ de forma tal que

$$\int_{\mathcal{Z}} \left(\sqrt{p(z; \theta + h)} - \frac{1}{2} h^t S(z; \theta) \sqrt{p(z; \theta)} \right)^2 d\mu(z) = o(\|h\|^2). \quad (7.2)$$

Notemos que, si $p(z; \theta)$ es diferenciable en θ , entonces

$$\frac{\partial p(z; \theta)^{1/2}}{\partial \theta} = \frac{1}{2} \dot{\ell}(z; \theta) \sqrt{p(z; \theta)}.$$

van der Vaart (1998), (Lema 7.6) encuentra condiciones que, sumadas a la diferenciable de $p(z; \theta)$, garantizan la diferenciable en media cuadrática de la aplicación $\theta \rightarrow p(-, \theta)$ en cierto valor θ . Además, bajo esas mismas condiciones, demuestra que $S(z; \theta) = \dot{\ell}(z, \theta)$, hecho que sugiere llamar a $S(z; \theta)$ función de score y denotarlo con $\dot{\ell}_\theta$, cuando resulte conveniente. Cuando el modelo es diferenciable en media cuadrática en cierto valor θ , se demuestra (van der Vaart (1998) Teorema 7.2) que

$$E_\theta \left(S(Z, \theta) \right) = 0 \quad \text{y existe la matriz} \quad E_\theta \left(S^t(Z, \theta) S(Z, \theta) \right). \quad (7.3)$$

En lo que sigue, nos concentraremos en submodelos unidimensionales para poder construir espacios tangentes en el contexto estadístico.

7.2. Espacio Tangente.

7.2.1. Submodelos unidimensionales - Curvas

Consideremos un modelo \mathcal{F} y sea P un elemento en \mathcal{F} . Diremos que \mathcal{F}_t es una curva en \mathcal{F} que pasa por P si es un submodelo paramétrico unidimensional que pasa por P en $t = 0$. Es decir, tenemos que

1. $\mathcal{F}_t = \{P_t : t \in [0, \epsilon]\}$ es un subconjunto de \mathcal{F}
2. $P_0 = P$

A modo de ejemplo, si \mathcal{F} es el conjunto de probabilidades absolutamente continuas respecto de μ , dada P con densidad $p(z)$ y una función $S^*(z)$ acotada con $E_P[S^*(Z)] = 0$, podemos construir la curva

$$\mathcal{F}_t = \{p_t(z) = (1 + tS^*(z))p(z) : t \in [0, \epsilon]\}.$$

Para fijar ideas, haciendo una analogía con \mathbb{R}^m podemos pensar que el modelo \mathcal{F} es una superficie, P un punto de dicha superficie, y el submodelo \mathcal{F}_t es una curva dentro de la superficie \mathcal{F} que pasa por P . Disponiendo de una noción de derivada, el conjunto formado por las derivadas de las curvas que pasan por un punto forman su espacio tangente. Necesitamos entonces dar una noción de derivada para las curvas \mathcal{F}_t en el modelo \mathcal{F} que pasan por P en $t = 0$, para así poder construir el espacio tangente en este nuevo contexto.

Definición 7.3. *Sea \mathcal{F} un modelo y P una probabilidad en \mathcal{F} . Una curva \mathcal{F}_t en \mathcal{F} que pasa por P se dice diferenciable en $t = 0$ si lo es en media cuadrática: existe $S(z) \in L_2(P)$ de forma tal que*

$$\lim_{t \rightarrow 0^+} \int_{\mathcal{Z}} \left(\frac{\sqrt{p_t}(z) - \sqrt{p}(z)}{t} - \frac{1}{2} S(z) \sqrt{p}(z) \right)^2 d\mu(z) \rightarrow 0$$

donde $dP_t = p_t d\mu$ y $dP = p d\mu$.

La función $S(z)$ es llamada score de la curva. Indistintamente, también llamaremos score a $S \equiv S(Z)$ donde $Z \sim P$.

A continuación repasamos el Lema 7.6 de van der Vaart (1998), debidamente adaptado para la noción de curvas diferenciables presentada en la definición anterior ya que el mismo será invocado en las demostraciones de los resultados de esta tesis.

Lema 7.3.1. Adaptación el Lema 7.6 de van der Vaart (1998) a curvas diferenciables. *Sea $\mathcal{F}_t = \{P_t : t \in [0, \varepsilon]\}$ una curva que pasa por P en $t = 0$, de forma tal que $dP_t = p_t d\mu$ y $dP = p d\mu$. Supongamos que la aplicación $t \rightarrow \sqrt{p_t}(z)$ es continuamente diferenciable para cada z y que la aplicación dada por $\rho(t) = \int \left\{ \frac{\dot{p}_t(z)}{p_t(z)} \right\}^2 p_t(z) d\mu(z)$ está bien definida y es continua en $[0, \varepsilon)$. Entonces, la curva \mathcal{F}_t es diferenciable (en el sentido de la Definición 7.3) en $t = 0$ y su score está dado por*

$$S(z) = \frac{\partial}{\partial t} \log(p_t(z))|_{t=0}.$$

Este Lema será de utilidad a la hora de caracterizar derivadas de curvas.

Observación 7.3.1. *Si la curva $\mathcal{F}_t = \{P_t : t \in [0, \varepsilon]\}$, con $dP_t = p_t d\mu$ y $dP = p d\mu$, que pasa por P en $t = 0$, es diferenciable, abusando del lenguaje, diremos que que $p_t(z)$ es diferenciables en $t = 0$.*

Habiendo definido derivadas de curvas, podemos introducir la noción de espacio tangente. La misma es relativa a un conjunto de curvas, como se indica en las próximas definiciones.

7.3.1. Conjunto tangente relativo - Tangente maximal

Definición 7.4. Sea \mathcal{F} un modelo y P un elemento en \mathcal{F} . Dada $\mathcal{C}(P)$, una familia de curvas diferenciables en \mathcal{F} que pasan por P , definimos su conjunto tangente relativo siendo el formado por los scores asociados a dichas curvas.

Como hemos mencionado, los scores tienen media cero y varianza finita, y por consiguiente el conjunto tangente relativo está formado por elementos en $L_2(P)$ con media cero. Este conjunto no tiene por qué ser un subespacio lineal. Resultará de suma utilidad considerar entonces el siguiente espacio.

Definición 7.5. Sea \mathcal{F} un modelo y P un elemento en \mathcal{F} . El espacio tangente asociado a una familia de curvas diferenciables $\mathcal{C}(P)$ en \mathcal{F} que pasan por P , es la clausura en $L_2(P)$ del subespacio vectorial generado por los scores de dichas curvas.

Observación 7.5.1. Notemos que todos los elementos del espacio tangente tienen esperanza cero.

A modo de ejemplo, si \mathcal{F} es un modelo paramétrico diferenciable en media cuadrática, como introdujimos en la Sección 7.1.1, el score asociado a la curva $p(z, \theta + te_i)$, siendo e_i el i -ésimo elemento de la base canónica, es la i -ésima componente del vector de scores $S(z; \theta) = \dot{\ell}(z, \theta)$. Por consiguiente, el espacio tangente asociado a las curvas $\mathcal{C} = \{p_t^i(z) = p(z, \theta + te_i); 1 \leq i \leq k\}$ coincide con el generado por las coordenadas del vector de scores $S(z, \theta)$.

El espacio tangente más grande que podemos considerar es el siguiente

Definición 7.6. Dado un modelo \mathcal{F} y P en el modelo, llamaremos espacio tangente maximal al asociado al conjunto de todas las curvas diferenciables en \mathcal{F} que pasan por P . Notaremos dicho conjunto mediante $\overline{\mathcal{F}}_P$.

Vamos ahora a demostrar que el espacio tangente maximal en una probabilidad en el modelo no paramétrico que solo asume la existencia de una función de densidad respecto a la medida μ , coincide con $L_2^0(P)$, el conjunto de variables aleatorias en $L_2(P)$ con media cero.

Antes de eso, en el próximo resultado veremos como calcular scores para cierta familia de curvas, apelando al Lema 7.3.1.

Lema 7.6.1. Sea Z un vector aleatorio con densidad p respecto de la medida μ . Sea $S^*(z)$ una función acotada. Consideremos la curva

$$p_t(z) = \{(1 + tS^*(z))p(z) : t \in [0, \epsilon]\}$$

y supongamos que pertenece al modelo bajo el cual estamos trabajando. Entonces, p_t es diferenciable en $t = 0$ con score dado por $S^* = S^*(Z)$.

Demostración. Comencemos notando que, como S^* es acotado, podemos tomar ϵ de forma tal que $p_t(z)$ es una densidad para todo $t \in [0, \epsilon)$; más aún, tomamos ϵ de forma tal que $1 + tS^*(z) > a > 0$ para todo $t \in [0, \epsilon)$, en cuyo caso tenemos que $\{z : p_t(z) > 0\} = \{z : p(z) > 0\}$. Para probar la diferenciabilidad de la curva, veamos que p_t verifica las condiciones pedidas en el Lema 7.3.1. Si z es tal que $p(z) = 0$, entonces la aplicación $t \rightarrow \sqrt{p_t(z)}$ resulta continuamente diferenciable pues es la función constantemente cero. Si $p(z)$ es distinto de cero, entonces

$$\frac{d}{dt} \left\{ \sqrt{(1 + tS^*(z))p(z)} \right\} = \frac{1}{2\sqrt{(1 + tS^*(z))p(z)}} S^*(z)p(z).$$

Luego, la aplicación $t \rightarrow \sqrt{p_t(z)}$ es continuamente diferenciable para todo z . Por otro lado

$$\rho(t) = \int \left\{ \frac{\dot{p}_t(z)}{p_t(z)} \right\}^2 p_t(z) d\mu(z) = \int \frac{p(z)S^*(z)^2}{(1 + tS^*(z))} d\mu(z).$$

Ahora bien, la función $\psi_t(z) = \frac{p(z)S^*(z)^2}{(1 + tS^*(z))}$ es continua en el intervalo $[0, \epsilon)$ para todo z . Además, como p es integrable, $S^*(z)$ acotada y $1 + tS^*(z) > a > 0$, usando el Teorema de convergencia mayorada, resulta que $\rho(t)$ está bien definida y es continua. En consecuencia, por el Lema 7.3.1, p_t es diferenciable con score

$$\frac{\partial}{\partial t} \log(p_t(z))|_{t=0} = S^*(z).$$

□

Ahora si, vamos a demostrar que el espacio tangente maximal en una probabilidad en el modelo no paramétrico que solo asume la existencia de una función de densidad respecto a la medida μ , coincide con $L_2^0(P)$, el conjunto de variables aleatorias en $L_2(P)$ con media cero. Para ser mas formales, fijada la medida σ -aditiva μ en $(\mathcal{Z}, \mathcal{A})$, consideremos el modelo formado por las funciones de probabilidad absolutamente continuas respecto de la medida μ

$$\mathcal{F} = \left\{ P : dP = p d\mu, p \geq 0 \int_{\mathcal{Z}} p(z) d\mu(z) = 1 \right\}.$$

Dada P en \mathcal{F} , en adelante la identificaremos con su densidad p , utilizando P y p indistintamente.

Como el modelo no pone restricciones, es de esperar que el espacio tangente maximal en p este formado por todas las funciones $S \equiv S(Z) \in L_2^0(p)$ donde

$$L_2^0(p) = \{ b(Z) : E_p[b(Z)^2] < \infty \text{ y } E_p[b(Z)] = 0 \}.$$

Para verificar esta conjetura, dada $S(Z)$ acotada en $L_2^0(p)$, siguiendo el Ejemplo 1.6 presentado en van der Vaart (2002), consideremos $\varepsilon > 0$ de forma tal que

$$p_t(z) = \{(1 + tS(z))p(z) : t \in [0, \varepsilon]\}$$

sea un submodelo de \mathcal{F} . Por el Lema 7.6.1 la curva p_t es diferenciable y tiene a $S(z)$ por score. Ahora consideramos la familia de curvas

$$\mathcal{C}(p) = \{p_t(z) = (1 + tS(z))p(z) : t \in [0, \varepsilon] \quad S \in L_2^0(p) \text{ acotada}\}.$$

Resulta que la clausura del conjunto tangente asociado a esta familia es $L_2^0(p)$. En efecto, por un lado ya vimos que el conjunto tangente asociado a $\mathcal{C}(p)$ está formado por los elementos acotados de $L_2^0(p)$. Resta ver que este conjunto es denso en $L_2^0(p)$. Para ello, dado $S \equiv S(Z) \in L_2^0(p)$ resulta que S es límite de la sucesión

$$S_n = SI_{(-n,n)}(S) - E[SI_{(-n,n)}(S)],$$

formada por elementos acotados en $L_2^0(p)$, de donde obtenemos el resultado deseado.

Comparación con \mathbb{R}^m .

En el cuadro dado a continuación resumimos la analogía hecha anteriormente entre los objetos de un espacio euclídeo y los objetos de un espacio de funciones.

Objetos	Espacio Euclídeo	Modelos
Espacio	\mathbb{R}^m	$L_2(\mu)$
Subconjunto	M	\mathcal{F}
Curvas	$\alpha(t) \in M$	\mathcal{F}_t
Derivadas	$\dot{\alpha}(0)$	S
Conjuntos Tangentes	conjunto de $\dot{\alpha}(0)$	colección de S

7.6.1. Caracterización de la proyección bajo la Condición 1.

A continuación enunciamos un resultado que caracteriza $\Lambda_k(p)$ cuando se verifica la Condición 1. Es decir cuando el vector Z puede representarse como

$$Z = (L_1, A_1, L_2, A_2, \dots, A_K, L_{K+1})$$

y para $k \in [K]$

$$h_k(z) = p_{A_k|\bar{L}_k, \bar{A}_{k-1}}(a_k|\bar{l}_k, \bar{a}_{k-1}).$$

Lema 7.6.2. *Supongamos que se verifica la Condición 1 y sea $p \in \mathcal{P}_{g, h_{[K]}-k}$. Definimos el conjunto*

$$\mathcal{T}_k \equiv \{\psi(\bar{A}_k, \bar{L}_k) \in L_2(p) \text{ tales que } E_{h_k}[\psi(\bar{A}_k, \bar{L}_k) | \bar{A}_{k-1} \bar{L}_k] = 0\}.$$

Entonces

$$\Lambda_k(p) = \mathcal{T}_k.$$

El Lema anterior es una consecuencia directa del siguiente resultado:

Lema 7.6.3. *Sean μ_1 , μ_2 y μ_3 medidas definidas en los espacios medibles $(\mathcal{X}, \mathcal{S}_x)$, $(\mathcal{Y}, \mathcal{S}_y)$ y $(\mathcal{W} \times, \mathcal{S}_w)$, respectivamente. Consideremos $\mu = \mu_1 \times \mu_2 \times \mu_3$ la medida producto definida en el espacio producto $(\mathcal{Z}, \mathcal{S}_z)$, siendo $\mathcal{Z} = \mathcal{X} \times \mathcal{Y} \times \mathcal{W}$, $\mathcal{S}_z = \mathcal{S}_x \times \mathcal{S}_y \times \mathcal{S}_w$ y $z = (x, y, w)$.*

Dadas P y P_t con $t \in [0, \varepsilon)$ medidas de probabilidad en $(\mathcal{Z}, \mathcal{S}_z)$ dominadas por μ , consideramos $p = dP/d\mu$ y $p_t = dP_t/d\mu$, $t \in [0, \varepsilon)$. Asimismo, cualquiera sea $p^ = dP^*/d\mu$, denotemos con $p_1^*(x)$, $p_2^*(y|x)$ y $p_3^*(w|x, y)$ a alguna versión de las densidades condicionales de p^* . Sea $p_t = p_1 p_{2,t} p_3$ (con $p_{t=0} = p$) una curva diferenciable en $t = 0$ con score $S(x, y, w)$, entonces $S = S(Z)$ satisface las siguientes dos propiedades:*

1. $S(Z) = S^*(X, Y)$ en $L_2(p)$, siendo $S^*(x, y) \equiv \int S(x, y, w) p_3(w|x, y) d\mu_3(w)$.
2. $E_p(S^*|X) = 0$, siendo $S^* = S^*(X, Y)$.

Demostración. Parte 1 Sea

$$\Delta_t \equiv \int \left[\frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} - \frac{1}{2} S(z) \sqrt{p(z)} \right]^2 d\mu(z),$$

por ser S el score de la curva p_t en $t = 0$ resulta que

$$\Delta_t \rightarrow 0.$$

Además,

$$\begin{aligned}
 \Delta_t &= \int \left[\frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} - \frac{1}{2} S^*(x, y) \sqrt{p(z)} - \frac{1}{2} (S(z) - S^*(x, y)) \sqrt{p(z)} \right]^2 d\mu(z) \\
 &= \int \left[\frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} - \frac{1}{2} S^*(x, y) \sqrt{p(z)} \right]^2 d\mu(z) \\
 &\quad + \int \left[\frac{1}{2} (S(z) - S^*(x, y)) \sqrt{p(z)} \right]^2 d\mu(z) \\
 &\quad - 2 \int \left[\left\{ \frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} - \frac{1}{2} S^*(x, y) \sqrt{p(z)} \right\} \frac{1}{2} (S(z) - S^*(x, y)) \sqrt{p(z)} \right] d\mu(z) \\
 &= \Delta_t^* + K
 \end{aligned}$$

donde $\Delta_t^* = \int \left[\frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} - \frac{1}{2} S^*(x, y) \sqrt{p(z)} \right]^2 d\mu(z)$ y $K = \int \left[\frac{1}{2} (S(z) - S^*(x, y)) \sqrt{p(z)} \right]^2 d\mu(z)$ pues

$$\begin{aligned}
 &\int \left[\left\{ \frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} - \frac{1}{2} S^*(x, y) \sqrt{p(z)} \right\} \frac{1}{2} (S(z) - S^*(x, y)) \sqrt{p(z)} \right] d\mu(z) \\
 &= \int d\mu_1(x) d\mu_2(y) \sqrt{p_1(x) p_2(y|x)} \left\{ \frac{\sqrt{p_1(x) p_{2,t}(y|x)} - \sqrt{p_1(x) p_2(y|x)}}{t} - \frac{1}{2} S^*(x, y) \sqrt{p_1(x) p_2(y|x)} \right\} A \\
 &= 0
 \end{aligned}$$

ya que $A \equiv \frac{1}{2} \int p_3(w|x, y) (S(x, y, w) - S^*(x, y)) d\mu_3(w) = 0$.

Luego $\Delta_t = \Delta_t^* + K \rightarrow 0$ con $\Delta_t \geq 0 \forall t$ y $K \geq 0$ de donde concluimos que $\Delta_t^* \rightarrow 0$ y $K = 0$ de donde $S(Z) = S^*(X, Y)$ en $L_2(p)$ como queríamos demostrar.

Parte 2. Basta verificar que

$$E_p(S^*(X, Y)a(X)) = 0 \quad (7.4)$$

para toda función $a(x)$ medible y acotada en el espacio (\mathcal{X}, S_x) .

Al ser S^* el score de la curva p_t en $t = 0$ resulta que

$$\frac{\sqrt{p_t(z)} - \sqrt{p}}{t} \rightarrow \frac{1}{2} S^* \sqrt{p} \text{ en } L_2(\mu). \quad (7.5)$$

Luego, tomando producto interno entre (7.5) y $a\sqrt{p}$ tenemos

$$\int \left(\frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} \right) a(x) \sqrt{p(z)} d\mu(z) \rightarrow \frac{1}{2} \int S^*(x, y) a(x) p(z) d\mu(z). \quad (7.6)$$

El hecho de que la curva p_t sea diferenciable garantiza que $\sqrt{p_t} \rightarrow \sqrt{p}$ en $L_2(\mu)$, como además a es una función acotada resulta

$$a\sqrt{p_t} \rightarrow a\sqrt{p} \text{ en } L_2(\mu) \quad (7.7)$$

luego de (7.5) y de (7.7) tenemos

$$\int \left(\frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} \right) a(x) \sqrt{p_t(z)} d\mu \rightarrow \frac{1}{2} \int S^*(x, y) a(x) p(z) d\mu(z). \quad (7.8)$$

Sumando (7.6) y (7.8) obtenemos

$$\int \left(\frac{\sqrt{p_t(z)} - \sqrt{p(z)}}{t} \right) a(x) (\sqrt{p_t(z)} + \sqrt{p(z)}) d\mu(z) \rightarrow \int S^*(x, y) a(x) p(z) d\mu(z), \quad (7.9)$$

a su vez, haciendo diferencia de cuadrados en el lado izquierdo de (7.9) tenemos que

$$\begin{aligned} & \frac{1}{t} \int (\sqrt{p_t(z)} - \sqrt{p(z)}) a(x) (\sqrt{p_t(z)} + \sqrt{p(z)}) d\mu(z) = \\ & \frac{1}{t} \left\{ \int a(x) p_t(z) d\mu(z) - \int a(x) p(z) d\mu(z) \right\} = \\ & \frac{1}{t} \left\{ \int a(x) p_1(x) d\mu_1(x) \int p_{2,t}(y|x) p_3(w|x, y) d\mu_2(y) d\mu_3(w) - \int a(x) p_1(x) d\mu_1(x) \int p_2(y|x) p_3(w|x, y) d\mu_2(y) d\mu_3(w) \right\} \\ & \frac{1}{t} \left\{ \int a(x) p_1(x) d\mu_1(x) - \int a(x) p_1(x) d\mu_1(x) \right\} = 0 \end{aligned}$$

de donde, para toda $a(x)$ acotada, $\int S^*(x, y) a(x) p(z) d\mu(z) = 0$ como queríamos ver. \square

A continuación enunciamos un corolario del Lema 7.6.2, muy usado a lo largo de este trabajo, que da una fórmula para $\Pi_p(B|\Lambda_k(p))$ cuando $B \in L_2(p)$ y se verifica la Condición 1.

Corolario 7.6.1. *Sea $B \in L_2(p)$ y supongamos que se verifica la Condición 1. Entonces*

$$\Pi_p(B|\Lambda_k(p)) = E_p(B|\bar{A}_k, \bar{L}_k) - E_p(B|\bar{A}_{k-1}, \bar{L}_k).$$

Demostración. Pongamos $\pi_k(B) \equiv E_p(B|\bar{A}_k, \bar{L}_k) - E_p(B|\bar{A}_{k-1}, \bar{L}_k)$. Debido al Lema 7.6.2, basta ver que para toda función $\ell(\bar{A}_k, \bar{L}_k)$ con esperanza condicional a \bar{A}_{k-1}, \bar{L}_k igual a cero, se verifica

$$E_p((B - \pi_k(B)) \ell(\bar{A}_k, \bar{L}_k)) = 0.$$

En efecto

$$\begin{aligned} & E_p((B - \pi_k(B)) \ell(\bar{A}_k, \bar{L}_k)) = E_p((E(B|\bar{A}_k, \bar{L}_k) - \pi_k(B)) \ell(\bar{A}_k, \bar{L}_k)) = \\ & E_p(E_p(B|\bar{A}_{k-1}, \bar{L}_k) \ell(\bar{A}_k, \bar{L}_k)) = E_p(E_p(B|\bar{A}_{k-1}, \bar{L}_k) E_p(\ell(\bar{A}_k, \bar{L}_k)|\bar{A}_{k-1}, \bar{L}_k)) = 0, \end{aligned}$$

como queríamos ver. \square

7.6.2. Revisión de los Supuestos hechos a lo largo de la Sección 4.2.

Vamos ahora a dar condiciones que garanticen la validez de los supuestos enunciados en el Capítulo 4.

El Supuesto 1 se satisface automáticamente si asumimos que p/p^* está acotado superiormente. En cambio, para garantizar la validez del Supuesto 2, requerimos que el cociente entre las dos densidades también este acotado inferiormente, como se prueba en el siguiente resultado.

Lema 7.6.4. *Sean $p = gh$, $p^* = g^*h^*$ de forma tal que $h_j = h_j^*$. Supongamos existen constantes $0 < c_1 < c_2$ de forma tal que*

$$c_1 \leq \frac{p(z)}{p^*(z)} \leq c_2. \quad (7.10)$$

Tenemos entonces que

$$\Lambda_j(p) = \Lambda_j(p^*).$$

Demostración. Comencemos notando que bajo la condición (7.10), tenemos que $L_2(p) = L_2(p^*)$. Mas aún,

$$c_1 \int r^2(z)p^*(z) d\mu(z) \leq \int r^2(z)p(z) d\mu(z) \leq c_2 \int r^2(z)p^*(z) d\mu(z). \quad (7.11)$$

Luego, para ver que los espacios tangentes coinciden, basta demostrar que el conjunto de scores en p y p^* es el mismo. Por simetría, basta ver que los scores en p también son scores en p^* . Para ello, sea $S(z)$ el score asociado a una curva $p_t = gh_1, \dots, h_{jt}, \dots, h_K$, en el modelo $\mathcal{P}_{g, h_{[K]-j}}$ que pasa por p en $t = 0$. Recordemos que $S = S(Z)$. Tenemos entonces que

$$\frac{\sqrt{gh_1, \dots, h_{jt}, \dots, h_K} - \sqrt{gh_1, \dots, h_j, \dots, h_K}}{t} - \frac{1}{2} \sqrt{gh_1, \dots, h_j, \dots, h_K} S \rightarrow 0 \text{ en } L_2(\mu).$$

Es decir,

$$\sqrt{\frac{g \prod_{i \neq j}^K h_i}{g^* \prod_{i \neq j}^K h_i^*}} \left\{ \frac{\sqrt{g^* h_1^*, \dots, h_{jt}, \dots, h_K^*} - \sqrt{g^* h_1^*, \dots, h_j, \dots, h_K^*}}{t} - \frac{1}{2} \sqrt{g^* h_1^*, \dots, h_j, \dots, h_K^*} S \right\} \rightarrow 0 \text{ en } L_2(\mu).$$

Luego, si tomamos el submodelo de $\mathcal{P}_{g^*, h_{[K]-j}}$ que pasa por p^* en $t = 0$, dado por $p_t^* = g^* h_1^* \dots h_{jt} \dots h_K^*$, resulta que

$$\sqrt{\frac{g \prod_{i \neq j}^K h_i}{g^* \prod_{i \neq j}^K h_i^*}} \left\{ \frac{\sqrt{p_t^*} - \sqrt{p^*}}{t} - \frac{1}{2} \sqrt{p^*} S \right\} \rightarrow 0 \text{ en } L_2(\mu),$$

con lo cual, como por hipótesis $\sqrt{\frac{g \prod_{i \neq j}^K h_i}{g^* \prod_{i \neq j}^K h_i^*}} = \sqrt{p/p^*}$ está acotado inferiormente, resulta que

$$\left\{ \frac{\sqrt{p_t^*} - \sqrt{p^*}}{t} - \frac{1}{2} \sqrt{p^*} S \right\} \rightarrow 0 \text{ en } L_2(\mu),$$

de donde $S \in \Lambda_j(p^*)$. □

En lo que respecta al Supuesto 3 parte *a*), para garantizar la ortogonalidad de los espacios $\Lambda_k(p)$, requeriremos que, además de las condiciones necesarias para verificar el Supuesto 2, el espacio tangente esté generado por elementos acotados. Es decir, demostraremos el siguiente resultado

Lema 7.6.5. *Supongamos que se verifica el Supuesto 2. Sean $S_i(z)$ y $S_j(z)$ scores en $\Lambda_i(p)$ y $\Lambda_j(p)$ respectivamente, obtenidos mediante caminos diferenciables en media cuadrática de forma tal que S_j está acotado, entonces $E_p(S_i S_j) = 0$.*

Demostración. Sea $p = gh$. Por hipótesis existen submodelos $p_{i,t} = g \prod_{k \neq i} h_k h_{i,t}$ y $p_{j,t} = g \prod_{k \neq j} h_k h_{j,t}$ de $\mathcal{P}_{g, h_{[K]-i}}$ y $\mathcal{P}_{g, h_{[K]-j}}$ respectivamente tales que

$$(a) \quad \frac{\sqrt{p_{j,t}} - \sqrt{p}}{t} \rightarrow \frac{1}{2} \sqrt{p} S_j \text{ en } L_2(\mu), \quad (7.12)$$

$$(b) \quad \frac{\sqrt{p_{i,t}} - \sqrt{p}}{t} \rightarrow \frac{1}{2} \sqrt{p} S_i \text{ en } L_2(\mu). \quad (7.13)$$

Luego, tomando producto interno entre (b) y $S_j \sqrt{p}$ tenemos que

$$\int \frac{\sqrt{p_{i,t}(z)} - \sqrt{p(z)}}{t} S_j(z) \sqrt{p(z)} d\mu(z) \rightarrow \int \frac{1}{2} S_j(z) S_i(z) p(z) d\mu(z). \quad (7.14)$$

El hecho de que la curva $p_{i,t}$ sea diferenciable garantiza que $\sqrt{p_{i,t}} \rightarrow \sqrt{p}$ en $L_2(\mu)$, como además S_j es una función acotada resulta

$$S_j \sqrt{p_{i,t}} \rightarrow S_j \sqrt{p} \text{ en } L_2(\mu). \quad (7.15)$$

Luego de (8.1) y (7.14) tenemos

$$\int \frac{\sqrt{p_{i,t}(z)} - \sqrt{p(z)}}{t} S_j(z) \sqrt{p_{i,t}} d\mu(z) \rightarrow \int \frac{1}{2} S_j(z) S_i(z) p(z) d\mu(z), \quad (7.16)$$

sumando (7.14) y (7.16) obtenemos

$$\int \frac{\sqrt{p_{i,t}(z)} - \sqrt{p(z)}}{t} S_j(z) \{ \sqrt{p_{i,t}} + \sqrt{p(z)} \} d\mu(z) \rightarrow \int S_j(z) S_i(z) p(z) d\mu(z). \quad (7.17)$$

Ahora bien, usando diferencia de cuadrados, vemos que el lado izquierdo de (7.17) coincide con

$$\frac{\int S_j(z) p_{i,t}(z) d\mu(z) - \int S_j(z) p d\mu}{t},$$

cuyo numerador es cero pues, como vale el Supuesto 2, resulta que $S_j \in \Lambda_j(p) = \Lambda_j(p_{i,t})$ y por lo tanto, $E_{p_j}(S_j) = 0$ y $E_{p_{i,t}}(S_j) = 0$, para todo t . Luego $\int S_j(z) S_i(z) p(z) d\mu(z) = 0$ como queríamos ver. \square

Finalmente el siguiente Lema garantiza la validez del Supuesto 3 débil:

Lema 7.6.6. *Supongamos que se verifica la Condición 1. Sea $c \subseteq [K]$ y sean $p = gh$ y $p^* = gh^* \in \mathcal{P}_{g,h_c}$. Supongamos que*

- a) $\{z : h_j^*(z) = 0\} \subseteq \{z : h_j(z) = 0\}$ para cada $1 \leq j \leq K$,
- b) $\int \left(\frac{p(z)}{p^*(z)} - 1 \right)^2 p^*(z) d\mu(z) < \infty$.

Entonces existe $S(z)$ el score de la curva $p(z)t + p^*(z)(1-t)$ en $t = 0$ y verifica

$$S(z) \in \oplus_{j \in \bar{c}} \Lambda_j(p^*).$$

Demostración. De ahora en más, dada q una función pondremos $sop\{q\} = \{z : q(z) > 0\}$ y denotamos con $I_A(z)$ a la función indicadora o característica del conjunto A la cual vale 1 para los elementos de A y cero fuera de A . En el Apéndice B, específicamente en el corolario (8.3.2), veremos que si se verifica el Supuesto 1 entonces existe el score $S(z)$ de la curva $p(z)t + p^*(z)(1-t)$ en $t = 0$ y está dado por

$$S(z) = \left\{ \frac{p(z)}{p^*(z)} - 1 \right\} I_{sop\{p^*\}}(z) = \left\{ \prod_{j=1}^K \frac{h_j(z)}{h_j^*(z)} - 1 \right\} I_{sop\{p^*\}}(z). \quad (7.18)$$

Como bajo los supuesto *a*) y *b*) del presente Lema se verifica el Supuesto 1, resulta que existe el score $S(z)$ de la curva $p(z)t + p^*(z)(1 - t)$ en $t = 0$ y verifica la expresión (7.18). Notemos que si definimos

$$S_{h^*}(z) = \left\{ \prod_{j=1}^K \frac{h_j(z)}{h_j^*(z)} I_{sop\{h_j^*\}}(z) - 1 \right\},$$

tenemos que para todo z perteneciente a $sop\{p^*\}$ resulta que $S_{h^*}(z) = S(z)$, luego estas funciones coinciden en $L_2(p^*)$, que es el espacio en donde estamos trabajando, en vista de lo cual pondremos

$$S(z) = \left\{ \prod_{j=1}^K \frac{h_j(z)}{h_j^*(z)} I_{sop\{h_j^*\}}(z) - 1 \right\}. \quad (7.19)$$

Por otra parte, la Condición 1 establece que el vector Z puede representarse como

$$Z = (L_1, A_1, L_2, A_2, \dots, A_K, L_{K+1}),$$

y para $k \in [K]$

$$h_k(z) = p_{A_k | \bar{L}_k, \bar{A}_{k-1}}(a_k | \bar{l}_k, \bar{a}_{k-1}).$$

Definimos

$$S_1^*(z) \equiv \left\{ \frac{h_1(z)}{h_1^*(z)} I_{sop\{h_1^*\}}(z) - 1 \right\},$$

y para $j = 2, \dots, K$, consideramos

$$S_j^*(z) \equiv \prod_{k=1}^{j-1} \frac{h_k(z)}{h_k^*(z)} I_{sop\{h_k^*\}} \left\{ \frac{h_j(z)}{h_j^*(z)} I_{sop\{h_j^*\}}(z) - 1 \right\}. \quad (7.20)$$

Luego, para todo k , tenemos que

$$S_1^*(z) + S_2^*(z) \dots S_k^*(z) = \prod_{i=1}^k \frac{h_i(z)}{h_i^*(z)} I_{sop\{h_i^*\}}(z) - 1.$$

En particular, de la expresión presentada en (7.19) para $S(z)$ concluimos que

$$S_1^*(z) + S_2^*(z) \dots S_K^*(z) = \left\{ \prod_{j=1}^K \frac{h_j(z)}{h_j^*(z)} I_{sop\{h_j^*\}}(z) - 1 \right\} = S(z).$$

Notemos ahora que para $j \in c$ tenemos que $h_j = h_j^*$, y por consiguiente si $z \in sop\{p^*\}$ resulta que $S_j^*(z) = 0$. Llamemos

$$S_{h_c}(z) = \sum_{j \in \bar{c}} S_j^*(z),$$

resulta que para $z \in \text{sop}\{p^*\}$ tenemos que $S_{h_c}(z) = S(z)$, luego estas funciones coinciden en $L_2(p^*)$, que es el espacio en donde estamos trabajando, en vista de lo cual pondremos

$$S(z) = \sum_{j \in \bar{c}} S_j^*(z).$$

Para completar la demostración basta ver que para cada $j \in \bar{c}$, $S_j^*(z)$ pertenece a $\Lambda_j(p^*)$. Como estamos bajo la Condición 1 y $S_j^* \equiv S_j^*(Z) = S_j^*(\bar{A}_j, \bar{L}_j)$, apelando al Lema 7.6.2, resta ver que $E_{p^*}(S_j^*|\bar{A}_{j-1}, \bar{L}_j) = 0$. Para ello, notemos que

$$\begin{aligned} E_{p^*}(S_j^*|\bar{A}_{j-1}, \bar{L}_j) &= \prod_{k=1}^{j-1} \frac{h_k(\bar{A}_{j-1}, \bar{L}_j)}{h_k^*(\bar{A}_{j-1}, \bar{L}_j)} I_{\text{sop}\{h_k^*\}}(\bar{A}_{j-1}, \bar{L}_j) E_{p^*} \left(\frac{h_j(Z)}{h_j^*(Z)} I_{\text{sop}\{h_j^*\}}(Z) - 1|\bar{A}_{j-1}, \bar{L}_j \right) = \\ &= \prod_{k=1}^{j-1} \frac{h_k(\bar{A}_{j-1}, \bar{L}_j)}{h_k^*(\bar{A}_{j-1}, \bar{L}_j)} I_{\text{sop}\{h_k^*\}}(\bar{A}_{j-1}, \bar{L}_j) \left\{ \int_{\text{sop}\{h_j^*\}} \frac{h_j(z)}{h_j^*(z)} h_j^*(z) d\mu(z) - 1 \right\} = \\ &= \prod_{k=1}^{j-1} \frac{h_k(\bar{A}_{j-1}, \bar{L}_j)}{h_k^*(\bar{A}_{j-1}, \bar{L}_j)} I_{\text{sop}\{h_k^*\}}(\bar{A}_{j-1}, \bar{L}_j) \left\{ \int_{\text{sop}\{h_j\}} \frac{h_j(z)}{h_j^*(z)} h_j^*(z) d\mu(z) + \int_{\text{sop}\{h_j^*\} \cap \{z: h_j(z)=0\}} \frac{h_j(z)}{h_j^*(z)} h_j^*(z) d\mu(z) - 1 \right\} = 0, \end{aligned}$$

pues $\int_{\{z: h_j(z)=0\} \cap \text{sop}\{h_j^*\}} h_j(z) d\mu(z) = 0$ y como $\{z : h_j^*(z) = 0 \text{ y } h_j(z) > 0\} = \emptyset$, resulta que $\{z : h_j(z) > 0\} = \{z : h_j^*(z) > 0 \text{ y } h_j(z) > 0\}$. \square

Capítulo 8

Apéndice B

8.1. Demostraciones correspondientes al Ejemplo 1.

En esta sección vamos a demostrar algunos resultados enunciados durante el ejemplo 1 del capítulo 3.

8.1.1. Demostración de la Proposición 2.1.1.

Demostración. Bajo

$$P(R_2 = 1 | R_1 = 1, L_1, L_2, Y_f) = P(R_2 = 1 | R_1 = 1, L_1, L_2) \quad (8.1)$$

$$Y_f = Y \text{ si } R_2 = 1 \quad (8.2)$$

tenemos que

$$\begin{aligned} E[Y | R_2 = 1, R_1 = 1, L_2, L_1] &= E[Y^f | R_2 = 1, R_1 = 1, L_2, L_1] \\ &= E[Y^f | R_1 = 1, L_2, L_1]. \end{aligned}$$

Luego,

$$E[E[Y | R_2 = 1, R_1 = 1, L_2, L_1] | R_1 = 1, L_1] = E[Y^f | R_1 = 1, L_1].$$

Si ahora suponemos además que

$$P(R_1 = 1 | L_1, Y_f) = P(R_1 = 1 | L_1) \quad (8.3)$$

arribamos a

$$E[E[Y | R_2 = 1, R_1 = 1, L_2, L_1] | R_1 = 1, L_1] = E[Y^f | L_1]$$

de donde

$$E\{E[E[Y | R_2 = 1, R_1 = 1, L_2, L_1] | R_1 = 1, L_1]\} = E(Y^f),$$

como queríamos ver. □

8.1.2. Demostración del Lema 2.1.1

Demostración. Este resultado se desprende de las propiedades de esperanza condicional, notando que cada factor R_i garantiza que condicionar en R_i coincide con condicionar en $R_i = 1$. Tenemos que para $H(\bar{L}_2, Y)$

$$\begin{aligned}
E_p \left(\frac{R_1 R_2 H(\bar{L}_2, Y)}{E_p(R_1|L_1) E_p(R_2|\bar{L}_2, R_1)} \right) &= E_p \left(E_p \left(\frac{R_1 R_2 H(\bar{L}_2, Y)}{E_p(R_1|L_1) E_p(R_2|\bar{L}_2, R_1 = 1)} \middle| R_1, R_2, \bar{L}_2 \right) \right) \\
&= E_p \left(\frac{R_1 R_2}{E_p(R_1|L_1) E_p(R_2|\bar{L}_2, R_1 = 1)} E_p(H(\bar{L}_2, Y)|R_1, R_2, \bar{L}_2) \right) = \\
&= E_p \left(\frac{R_1 R_2}{E_p(R_1|L_1) E_p(R_2|\bar{L}_2, R_1 = 1)} E_p(H(\bar{L}_2, Y)|R_1 = 1, R_2 = 1, \bar{L}_2) \right) = \\
&= E_p \left(E_p \left(\frac{R_1 R_2 E_p(H(\bar{L}_2, Y)|R_2 = R_1 = 1, \bar{L}_2)}{E_p(R_1|L_1) E_p(R_2|\bar{L}_2, R_1 = 1)} \middle| \bar{L}_2, R_1 \right) \right) = \\
&= E_p \left(\frac{R_1 E_p(H(\bar{L}_2, Y)|R_1 = 1, R_2 = 1, \bar{L}_2)}{E_p(R_1|L_1)} \right).
\end{aligned}$$

Para concluir el resultado, condicionamos primero en L_1, R_1 y por ultimo en L_1 , de forma tal que la última expresión escrita en la ecuación precedente coincide con

$$\begin{aligned}
&E_p \left(E_p \left(\frac{R_1 E_p(E_p(H(\bar{L}_2, Y)|R_1 = 1, R_2 = 1, \bar{L}_2)|L_1, R_1 = 1)}{E_p(R_1|L_1)} \middle| L_1 \right) \right) = \\
&E_p \left(E_p \left(E_p(E_p(H(\bar{L}_2, Y)|R_1 = 1, R_2 = 1, \bar{L}_2)|L_1, R_1 = 1) \right) \right),
\end{aligned}$$

como queríamos ver. □

8.1.3. Demostración del Lema 2.1.3

Demostración. Supongamos primero que $p \in \mathcal{G}_1 \cap \mathcal{G}_2$, luego $\tau_1(g)$ y $\tau_2(g)$ verifican

$$* \nu_2(g) = E_g(Y|R_1 = R_2 = 1, \bar{L}_2) = b_2(\bar{L}_2; \tau_2(g)).$$

$$** \nu_1(g) = E_g(E_g(Y|R_1 = R_2 = 1, \bar{L}_2)|R_1 = 1, L_1) = E_g(b_2(\bar{L}_2; \tau_2(g))|R_1 = 1, L_1) = b_1(L_1; \tau_1(g)).$$

Paralelamente tenemos que para todo τ_2, α_1 y α_2

$$E_p \left(\frac{R_2 R_1 \{Y - b_2(\bar{L}_2; \tau_2)\}}{E(R_1|L_1; \alpha_1) E(R_2|R_1 = 1, \bar{L}_2; \alpha_2)} \right) = E_p \left(\frac{R_2 R_1 \{E_g(Y|R_1 = R_2 = 1, \bar{L}_2) - b_2(\bar{L}_2; \tau_2)\}}{E(R_1|L_1; \alpha_1) E(R_2|R_1 = 1, \bar{L}_2; \alpha_2)} \right), \quad (8.4)$$

y para todo τ_1, τ_2 , y α_1

$$E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1)} \{b_2(\bar{L}_2; \tau_2) - b_1(L_1; \tau_1)\} \right) = E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1)} \left\{ E_g(b_2(\bar{L}_2; \tau_2) | L_1 R_1 = 1) - b_1(L_1; \tau_1) \right\} \right).$$

Luego si en (8.4) tomamos $\tau_2 = \tau_2(g)$, en la ecuación anterior tomamos $\tau_1 = \tau_1(g)$ y $\tau_2 = \tau_2(g)$ y usamos que $E_g(b_1(L_1; \tau_1(g))) = \beta(g)$ obtenemos: para todo α_1 y α_2

$$E_p \left(U(\beta(g), \tau_1(g), \tau_2(g), \alpha_1, \alpha_2) \right) = 0,$$

con lo cual es cierta la afirmación a) del Lema.

Ahora supondremos que $p \in \mathcal{H}_1 \cap \mathcal{H}_2$, luego existen $\alpha_1(h_1)$ y $\alpha_2(h_2)$ verificando

$$* \quad E_{h_1}(R_1|L_1) = E(R_1|L_1, \alpha_1(h_1)),$$

$$** \quad E_{h_2}(R_2|R_1, \bar{L}_2) = E(R_2|R_1, \bar{L}_2, \alpha_2(h_2)),$$

de donde, usando propiedades de la esperanza condicional y el Lema 2.1.1 se cumple

$$E_p \left(\frac{R_2 R_1}{E(R_1|L_1; \alpha_1(h_1)) E(R_2|R_1 = 1, \bar{L}_2; \alpha_2(h_2))} Y \right) = \beta(g),$$

$$E_p \left(\frac{R_2 R_1}{E(R_1|L_1; \alpha_1(h_1)) E(R_2|R_1 = 1, \bar{L}_2; \alpha_2(h_2))} b_2(L_1, L_2; \tau_2) \right) = E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1(h_1))} b_2(L_1, L_2; \tau_2) \right),$$

$$E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1(h_1))} b_1(L_1; \tau_1) \right) = E_p(b_1(L_1; \tau_1)),$$

y luego concluimos que para todo $\tau_2 \in \Upsilon_2$, $\tau_1 \in \Upsilon_1$

$$E_p \left(U(\beta(g), \tau_1, \tau_2, \alpha_1(h_1), \alpha_2(h_2)) \right) = 0,$$

verificandose el item b) del Lema.

Supongamos que los modelos asumidos para h_1 y $\nu_2(g)$ son ambos correctos, es decir supongamos que $p \in \mathcal{H}_1 \cap \mathcal{G}_2$, en ese caso tendremos que existen $\alpha_1(h_1)$ y $\tau_2(g)$ verificando

$$(*) \quad E_{h_1}(R_1|L_1) = E(R_1|L_1, \alpha_1(h_1)),$$

$$(**) \quad E_p(Y|R_1 = R_2 = 1, \bar{L}_2) = b_2(\bar{L}_2; \tau_2(g)),$$

luego si en la igualdad (8.4) tomamos $\tau_2 = \tau_2(g)$, y además usamos las igualdades

$$E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1(h_1))} b_2(\bar{L}_2; \tau_2(g)) \right) = E_g \left(E_g \left(b_2(\bar{L}_2; \tau_2(g)) | L_1, R_1 = 1 \right) \right) = \beta(g),$$

y

$$E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1(h_1))} b_1(L_1; \tau_1) \right) = E_p(b_1(L_1; \tau_1)) \quad \forall \tau_1$$

tenemos que

$$E_p \left(U(\beta(g), \tau_1, \tau_2(g), \alpha_1(h_1), \alpha_2) \right) = 0,$$

para τ_1 y α_2 arbitrarios como queríamos ver.

Finalmente supongamos que los modelos asumidos para h_2 y $\nu_1(g)$ son ambos correctos, o lo que es lo mismo, asumamos que $p \in \mathcal{G}_1 \cap \mathcal{H}_2$. Con lo cual

$$\begin{aligned} \star \quad & E_g(E_g(Y|R_2 = R_1 = 1, L_1, L_2) | R_1 = 1, L_1) = b_1(L_1; \tau_1(g)), \\ \star\star \quad & E_{h_2}(R_2|R_1, \bar{L}_2) = E(R_2|R_1, \bar{L}_2, \alpha_2(h_2)). \end{aligned}$$

Luego se verifica

$$\begin{aligned} E_p \left(\frac{R_2 R_1 (Y - b_2(\bar{L}_2; \tau_2(g)))}{E(R_1|L_1; \alpha_1) E(R_2|R_1 = 1, \bar{L}_2; \alpha_2(h_2))} \right) &= E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1)} \left(E_g(Y|R_1 = R_2 = 1, \bar{L}_2) - b_2(\bar{L}_2; \tau_2) \right) \right) = \\ & E_p \left(\frac{R_1}{E(R_1|L_1; \alpha_1)} \{b_1(L_1; \tau_1(g)) - b_2(\bar{L}_2; \tau_2)\} \right), \end{aligned}$$

y

$$E_p(b_1(L_1; \tau_1(g))) = \beta(g),$$

de donde obtenemos:

$$E_p \left(U(\beta, \tau_1(g), \tau_2, \alpha_1, \alpha_2(h_2)) \right) = 0,$$

para τ_2 y α_1 arbitrarios. □

8.2. Demostraciones correspondientes al Ejemplo 2.

8.2.1. Demostración de la Proposición 2.2.1.

Demostración. Sea $p^f \in \mathcal{F}^f$, y sea p la distribución del vector (W, X, V, R, Y) , con $Y = RY^f + (1 - R)$. Tenemos

$$\begin{aligned}
E\left(E\left(Y|R=1, V, X, W\right)|X, W\right) &= \underbrace{E\left(E\left(Y^f|R=1, V, X, W\right)|X, W\right)}_{\text{por definición de } Y} = \\
& \underbrace{E\left(E\left(Y^f|V, X, W\right)|X, W\right)}_{\text{por (2.40)}} = \\
& \underbrace{E\left(E\left(Y^f|X, W\right)\right)}_{\text{por Torres}} = \\
& \gamma\left(p^f\right) X + t_{p^f}(W), \\
& \text{pues } p^f \in \mathcal{F}^f
\end{aligned}$$

como queríamos ver. □

8.2.2. Demostración del Lema 2.2.1.

Demostración. Usando las propiedades de la esperanza condicional resulta

$$\begin{aligned}
E\left(\psi(W, X, V, Y) \frac{R}{E(R|W, X, V)}\right) &= E\left(E\left(\psi(W, X, V, Y) \frac{R}{E(R|W, X, V)} \middle| R, X, W, V\right)\right) = \\
& E\left(E\left(\psi(W, X, V, Y) \middle| R=1, X, W, V\right) \frac{R}{E(R|W, X, V)}\right) = \\
& E\left(E\left(E\left(\psi(W, X, V, Y) \middle| R=1, X, W, V\right) \frac{R}{E(R|W, X, V)} \middle| X, VW\right)\right) = \\
& E\left(\frac{E\left(\psi(W, X, V, Y) \middle| R=1, X, W, V\right)}{E(R|W, X, V)} E(R|X, VW)\right) = \\
& E\left(E\left(\psi(W, X, V, Y) \middle| R=1, W, X, V\right)\right).
\end{aligned}$$

como queríamos ver. □

8.3. Demostraciones correspondientes al Capítulo 4.

En esta sección demostraremos afirmaciones y proposiciones enunciadas durante el Capítulo 4.

Comenzaremos viendo que si vale la factorización dada en la Condición 1, entonces el submodelo \mathcal{P}_{g, h_c} es convexo para todo $c \subsetneq [K]$, $g \in \mathcal{U}^{np}$ y $h_c \in \otimes_{j \in c} \mathcal{V}_j^{np}$.

8.3.1. Demostración del Lema 4.1.1

Demostración. Como se verifica la Condición 1, si tomamos $Z \equiv (L_1, A_1 \dots L_{K+1}) \sim p$ entonces, si para cada $0 \leq i \leq K$ definimos

$$g_i(z) \equiv p_{L_{i+1} | \bar{L}_i, \bar{A}_i} (l_{i+1} | \bar{l}_i, \bar{a}_i),$$

y para cada $1 \leq i \leq K$ definimos

$$h_i(z) \equiv p_{A_i | \bar{L}_i, \bar{A}_{i-1}} (a_i | \bar{l}_i, \bar{a}_{i-1}),$$

se verifica que $p = gh$ con

$$g(z) = \prod_{i=0}^K g_i(z), \quad h(z) = \prod_{i=1}^K h_i(z).$$

Sea $c \subsetneq [K]$, $g \in \mathcal{U}^{np}$ y $h_c \in \otimes_{j \in c} \mathcal{V}_j^{np}$ y sean $p^1 = gh^1$ y $p^2 = gh^2 \in \mathcal{P}_{g, h_c}$. Luego si $k \in c$, entonces $h_k^1 = h_k^2 = h_k$. Queremos ver que la Condición 1 implica que para todo $\theta \in (0, 1)$, $p_\theta(z) \equiv \theta p^1(z) + (1 - \theta)p^2(z) \in \mathcal{P}_{g, h_c}$. Es decir, queremos ver que si $Z^\theta \equiv (L_1^\theta, A_1^\theta \dots L_{K+1}^\theta) \sim p_\theta$, entonces para $k \in c$

$$A_k^\theta | \bar{L}_k^\theta, \bar{A}_{k-1}^\theta \sim h_k.$$

Por construcción

$$p_\theta(z) \equiv \theta g(z)h^1(z) + (1 - \theta)g(z)h^2(z) = \prod_{i=0}^K g_i(z) \left\{ \theta \prod_{i=1}^K h_i^1(z) + (1 - \theta) \prod_{i=1}^K h_i^2(z) \right\},$$

marginalizando obtenemos

$$(\bar{A}_k^\theta, \bar{L}_k^\theta) \sim \prod_{i=0}^{k-1} g_i(z) \left\{ \theta \prod_{i=1}^k h_i^1(z) + (1 - \theta) \prod_{i=1}^k h_i^2(z) \right\} \quad (8.5)$$

y

$$(\overline{A}_{k-1}^\theta, \overline{L}_k^\theta) \sim \prod_{i=0}^{k-1} g_i(z) \left\{ \theta \prod_{i=1}^{k-1} h_i^1(z) + (1-\theta) \prod_{i=1}^{k-1} h_i^2(z) \right\} \quad (8.6)$$

luego, si $k \in c$ como $h_k^1 = h_k^2 = h_k$, reemplazando este hecho en la ecuación (8.5) nos queda

$$(\overline{A}_k^\theta, \overline{L}_k^\theta) \sim \prod_{i=0}^{k-1} g_i(z) \left\{ \theta \prod_{i=1}^{k-1} h_i^1(z) + (1-\theta) \prod_{i=1}^{k-1} h_i^2(z) \right\} h_k \quad (8.7)$$

de donde

$$A_k^\theta | \overline{L}_k^\theta, \overline{A}_{k-1}^\theta \sim h_k.$$

como queríamos ver. □

8.3.2. Demostración del Teorema 4.2.1.

Para demostrar este teorema, enunciamos y demostramos algunos resultados preliminares.

Lema 8.3.2

Si se verifican el Supuesto 1, es decir si se verifican las siguientes condiciones:

- 1) $\{p^*(z) = 0\} \subseteq \{p(z) = 0\}$.
- 2) $\int \left(\frac{p(z)}{p^*(z)} - 1 \right)^2 p^*(z) d\mu(z) < \infty$.

entonces se cumple:

- a) Para cada z la aplicación $t \rightarrow \sqrt{tp(z) + (1-t)p^*(z)}$ tiene derivada continua en $[0, \epsilon)$, $0 < \epsilon < 1$.
- b) La aplicación

$$t \rightarrow \ell(t) \equiv \int \left[\frac{\frac{d}{dt}(tp(z) + (1-t)p^*(z))}{tp(z) + (1-t)p^*(z)} \right]^2 [tp(z) + (1-t)p^*(z)] d\mu(z) \quad (8.8)$$

está bien definida y es continua.

c) Si $N \in L_2(p^*)$ entonces $N \in L_1(p)$.

Demostración. Veamos primero a). Si z es tal que $p^*(z) = 0$ entonces $p(z) = 0$, de donde resulta que $\sqrt{tp(z) + (1-t)p^*(z)}$ es la función constante 0 que es continua con primer derivada continua.

Si z es tal que $p^*(z) > 0$, entonces en cada $t > 0$ se verifica que

$$\frac{d}{dt} \sqrt{tp(z) + (1-t)p^*(z)} = \frac{p(z) - p^*(z)}{2\sqrt{tp(z) + (1-t)p^*(z)}}.$$

Además el límite cuando t tiende a cero por derecha existe y es igual a $\frac{p-p^*}{2\sqrt{p^*}}$. Con esto probamos la parte a). Veamos que vale b). Comencemos notando que, bajo la primer condición del Supuesto 1, tenemos que $\text{sop}\{tp + (1-t)p^*\} = \text{sop}\{p^*\}$, donde recordamos que si f es una función de densidad ponemos $\text{sop}\{f\} = \{z : f(z) > 0\}$. Tenemos entonces que

$$\begin{aligned} \ell(t) &\equiv \int \left[\frac{\frac{d}{dt}(tp(z) + (1-t)p^*(z))}{tp(z) + (1-t)p^*(z)} \right]^2 [tp(z) + (1-t)p^*(z)] d\mu(z) = \\ &\int_{\text{sop}\{tp+(1-t)p^*\}} \frac{(p(z) - p^*(z))^2}{tp(z) + (1-t)p^*(z)} d\mu(z) = \int_{\text{sop}\{p^*\}} \frac{(p(z) - p^*(z))^2}{tp(z) + (1-t)p^*(z)} d\mu(z). \end{aligned}$$

Ahora bien, definimos

$$f_t(z) = I_{\{\text{sop}\{p^*\}}}(z) \frac{(p(z) - p^*(z))^2}{tp(z) + (1-t)p^*(z)},$$

donde recordamos que denotamos con $I_A(z)$ a la función indicadora o característica del conjunto A la cual vale 1 para los elementos de A y cero fuera de A . Luego, para todo z , $f_t(z)$ es una función continua en t y además, para $t \in [0, \epsilon)$, resulta que

$$f_t(z) \leq \frac{1}{1-\epsilon} I_{\text{sop}\{p^*\}}(z) \left(\frac{p(z)}{p^*(z)} - 1 \right)^2 p^*(z),$$

siendo la última función es integrable, por el item (2) del Supuesto 1. Luego, por el teorema de convergencia mayorada, $\ell(t)$ es una función continua para todo $t \in [0, \epsilon)$.

Demostración del item c) Sea $N = N(Z)$, luego

$$\begin{aligned} E_p(|N|) &= \int_{\text{sop}\{p\}} |N(z)| p(z) d\mu(z) = \int_{\text{sop}\{p\} \cap \text{sop}\{p^*\}} |N(z)| p(z) d\mu(z) \\ &\quad + \int_{\text{sop}\{p\} \cap \{z:p^*(z)=0\}} |N(z)| p(z) d\mu(z), \end{aligned}$$

por el ítem 1) del Supuesto 1, resulta que $\text{sop}\{p\} \cap \{z : p^*(z) = 0\} = \emptyset$, de donde $\int_{\text{sop}\{p\} \cap \{z : p^*(z) = 0\}} |N(z)| p(z) d\mu(z) = 0$, luego

$$\begin{aligned} E_p(|N|) &= \int_{\text{sop}\{p\} \cap \text{sop}\{p^*\}} |N(z)| p(z) d\mu(z) \leq \int_{\text{sop}\{p^*\}} |N(z)| p(z) d\mu(z) \\ &= \int_{\{\text{sop}\{p^*\}\}} |N(z)| \left(\frac{p(z)}{p^*(z)} - 1 \right) p^*(z) d\mu(z) + \int_{\{\text{sop}\{p^*\}\}} |N(z)| p^*(z) d\mu(z) < \infty, \end{aligned}$$

pues :

★ por hipótesis $|N| \in L_2(p^*)$ y por el el Supuesto 1 parte b) $\left(\frac{p(z)}{p^*(z)} - 1\right) \in L_2(p^*)$, entonces, por la desigualdad de Cauchy-Schwarz, el primer término en la última igualdad es finito.

★★ por hipótesis $|N| \in L_2(p^*)$, entonces $|N| \in L_1(p^*)$, luego

$$\int_{\{\text{sop}\{p^*\}\}} |N(z)| p^*(z) d\mu(z) < \infty$$

Finalmente $E_p(|N|) < \infty$ como queríamos ver. \square

Ahora enunciamos y demostramos el siguiente corolario.

Corolario 8.3.2

Bajo el Supuesto 1 la aplicación $t \rightarrow \sqrt{tp(z) + (1-t)p^*(z)}$ es derivable en media cuadrática con score en $t = 0$ igual a

$$I_{\text{sop}\{p^*\}}(z) \left(\frac{p(z)}{p^*(z)} - 1 \right).$$

Demostración. Bajo el Supuesto 1 vale a) y b) del lema anterior, luego, por el Lema 7.6 de van der Vaart (1998) Capítulo 7, vale la conclusión del corolario. \square

Ahora estamos en condiciones de demostrar el Teorema 4.2.1.

Demostración del Teorema 4.2.1.

Demostración. Vimos en el ítem c) del Lema 8.3.2 que N pertenece a $L_1(p)$, siendo que pertenece a $L_2(p^*)$, quedando así demostrada la integrabilidad de N respecto de p . Además, tenemos por hipótesis que $p_t(z) \equiv tp(z) + (1-t)p^*(z) \in \mathcal{N}$ y por el corolario anterior, p_t es diferenciable en media cuadrática en $t = 0$ con score dado por

$I_{sop\{p^*\}}(z)\left(\frac{p(z)}{p(z)^*} - 1\right)$. Luego, siendo que $N = N(Z)$ es ortogonal al espacio tangente maximal de \mathcal{N} en p^* , tenemos

$$\begin{aligned} 0 &= E_{p^*} \left(N I_{sop\{p^*\}} \left(\frac{p}{p^*} - 1 \right) \right) = \int_{z:p(z)^* > 0} N(z) \left(\frac{p(z)}{p(z)^*} - 1 \right) p^*(z) d\mu(z) = \\ & \int_{z:p(z)^* > 0} N(z) p(z) d\mu(z) - E_{p^*}(N) = \int_{z:p(z) > 0} N(z) p(z) d\mu(z) + \\ & \int_{z:p(z)=0 \text{ } p^*(z) > 0} N(z) p(z) d\mu(z) - E_{p^*}(N) = E_p(N) - E_{p^*}(N), \end{aligned}$$

pues $\int_{z:p(z)=0 \text{ } p^*(z) > 0} N(z) p(z) d\mu(z) = 0$ y como $\{z : p^*(z) = 0 \text{ } p(z) > 0\} = \emptyset$, resulta que $\{z : p(z) > 0\} = \{x : p^*(z) > 0 \text{ y } p(z) > 0\}$.

Finalmente $E_p(N) = E_{p^*}(N)$, como queríamos ver. □

8.3.3. Demostración del Corolario 4.2.1.

Demostración. Por como tomamos c , resulta que p y p^* son elementos de \mathcal{P}_{g, h_c} convexo, luego $p_t = (1-t)p^* + tp \in \mathcal{P}_{g, h_c}$ para todo $t \in [0, 1]$. Como p y p^* verifican el Supuesto 1, $N \in L_2(p^*)$ y también pertenece al espacio ortogonal al tangente maximal del modelo \mathcal{P}_{g, h_c} en el punto p^* , entonces por el Teorema 4.2.1, tenemos que $E_p(N) = E_{p^*}(N)$. □

8.3.4. Demostración de la Proposición 4.3.1.

Demostración. Tomemos $c \subseteq [K]$ tal que

- i) $h_j^* = h_j$ si $j \in c$,
- ii) $\psi_j \{\eta_j(g_j), \beta_j\} = \psi_j \{\eta_j(g), \beta(g)\}$ si $j \in \bar{c} \cap b$,
- iii) $\eta_j(g_j) = \eta_j(g)$ si $j \in \bar{c} \cap \bar{b}$.

Siendo π_k el operador de proyección presentado en la identidad (4.14), tenemos que

$$E_{gh} \left(M(\beta(g), h^*) - \sum_{j \in b} \pi_j(\psi_j \{\eta_j(g_j), \beta_j\}, h^*) - \sum_{j \in \bar{b}} \pi_j(\psi_j \{\eta_j(g_j), \beta(g)\}, h^*) \right) =$$

$$E_{gh} \left(M(\beta(g), h^*) - \sum_{k=1}^K \Pi_{g_k h^*} (M(\beta_k, h^*) | \Lambda_k(h_k^*)) \right),$$

siendo que para $k \in \bar{c}$

$$\Pi_{g_k h^*} (M(\beta_k, h^*) | \Lambda_k(h_k^*)) = \Pi_{gh^*} (M(\beta(g), h^*) | \Lambda_k(h_k^*)).$$

El resultado se deduce del Lema 4.2.1. \square

8.3.5. Demostración de la Proposición 4.3.2.

Demostración. En adelante, utilizaremos $U(\beta, \tau, \alpha)$ para denotar a la función $U^b(\beta, \tau, \alpha)$, definida en (4.18). Sea $c \subseteq [K]$, y $p = gh \in \mathcal{M}_c$. Tomamos $(\tau, \alpha) \in \otimes_{k=1}^K \Upsilon_k \times \otimes_{k=1}^K \Xi_k$ verificando $(\alpha_c, \tau_{\bar{c}}) = (\alpha_c(h_c), \tau_{\bar{c}}(g))$. Queremos ver que

$$E_p(U(\beta(b), \tau, \alpha)) = 0.$$

Recordemos que

$$\Pi_{gh} (M(\beta, h) | \Lambda_k(gh)) = \pi_k(\psi_k\{\eta_k(g), \beta\}, h) \quad (8.9)$$

donde

- * $\eta_k : \mathcal{U}_k \rightarrow \mathcal{N}_k$
- * $\psi_k : \mathcal{N}_k \times \mathbb{R}^r \rightarrow \mathcal{X}_k$
- * $\pi_k : \mathcal{X}_k \times \otimes_{k=1}^K \mathcal{V}_j^{np} \rightarrow \Delta_k$ con Δ_k un espacio de funciones.

Recordemos también que para construir $U = U^b$, hemos fijado un conjunto $b \subset \{1, \dots, K\}$, considerado

$$\rho_k(g) = \begin{cases} \psi_k\{\eta_k(g), \beta(g)\} & \text{si } k \in b \\ \eta_k(g) & \text{si } k \in \bar{b} \end{cases}$$

y hemos definido a la función U siendo

$$U(\beta, \tau, \alpha) = M(\beta, h_\alpha) - \sum_{j \in b} \pi_j(\rho_{j, \tau_j}, h_\alpha) - \sum_{j \in \bar{b}} \pi_j(\psi_j\{\rho_{j, \tau_j}, \beta\}, h_\alpha). \quad (8.10)$$

Si $p \in \mathcal{M}_c$, para $j \in \bar{c}$ resulta que $p \in \mathcal{G}_j$, y por consiguiente $\rho_j(g) \in \mathcal{R}_{sub, j}$. Recordemos que, en tal caso, estamos utilizando $\tau_j(g)$ para denotar al único valor para el cual $\rho_j(g) = \rho_{j, \tau_j(g)}$. Tenemos entonces que si $j \in \bar{c} \cap \bar{b}$

$$\pi_j(\psi_j\{\rho_{j, \tau_j}, \beta\}, h_\alpha) = \pi_j(\psi_j\{\eta_j(g), \beta\}, h_\alpha) = \Pi_{gh_\alpha} (M(\beta, h_\alpha) | \Lambda_j(gh_\alpha)), \quad (8.11)$$

mientras que si $j \in \bar{c} \cap b$, tenemos que

$$\pi_j (\rho_{j,\tau_j}, h_\alpha) = \pi_j (\psi_j \{ \eta_j (g), \beta (g) \}, h_\alpha) = \Pi_{gh_\alpha} (M (\beta (g), h_\alpha) | \Lambda_j (gh_\alpha)). \quad (8.12)$$

Cuando $j \in c$, debemos explotar el hecho que $h_j = h_{j,\alpha_j(h_j)}$. Vamos ahora a considerar los siguientes casos: (i) $j \in c \cap b$ y (ii) $j \in c \cap \bar{b}$. En el primer caso, si $j \in c \cap b$, tenemos que $\mathcal{R}_{j,sub} \subseteq \chi_j$. Luego, dado τ_j , existen $g_{\tau_j} \in \mathcal{U}_j$ y $\beta_j \in \mathbb{R}^r$ de forma tal que

$$\rho_{j,\tau_j} = \psi_j \{ \eta_j (g_{\tau_j}), \beta_j \} \quad (8.13)$$

y por consiguiente

$$\pi_j (\rho_{j,\tau_j}, h_\alpha) = \pi_j (\psi_j \{ \eta_j (g_{\tau_j}), \beta_j \}, h_\alpha). \quad (8.14)$$

Por último, si $j \in c \cap \bar{b}$, tenemos que dado τ_j existe g_{τ_j} de forma tal que $\nu_j(g_{\tau_j}) = \rho_{j,\tau_j}$ y por consiguiente

$$\pi_j (\psi_j \{ \rho_{j,\tau_j}, \beta \}, h_\alpha) = \pi_j (\psi_j \{ \eta_j (g_{\tau_j}), \beta \}, h_\alpha). \quad (8.15)$$

Tenemos entonces: si $\tau_{\bar{c}} = \tau_{\bar{c}}(g)$,

$$\begin{aligned} U(\beta(g), \tau, \alpha) &= M(\beta(g), h_\alpha) - \sum_{j \in b \cap c} \pi_j (\psi_j \{ \eta_j (g_{\tau_j}), \beta_j \}, h_\alpha) - \sum_{j \in b \cap \bar{c}} \pi_j (\psi_j \{ \eta_j (g), \beta (g) \}, h_\alpha) \\ &- \sum_{j \in \bar{b} \cap c} \pi_j (\psi_j \{ \eta_j (g_{\tau_j}), \beta (g) \}, h_\alpha) - \sum_{j \in \bar{b} \cap \bar{c}} \pi_j (\psi_j \{ \eta_j (g), \beta (g) \}, h_\alpha). \end{aligned}$$

Por último, como $p = gh$ y $p^* = gh_\alpha \in \mathcal{M}$ satisfacen el Supuesto 1, se satisface el supuesto 2 (o el débil) y el supuesto 3 y como para el conjunto c dado se verifican los ítem i) ii) iii) iv) de la Proposición 4.3.1 podemos concluir que

$$E_{gh} (U(\beta(g), \tau, \alpha)) = 0,$$

para $p \in \mathcal{M}_c$, $(\alpha_c, \tau_{\bar{c}}) = (\alpha_c(h_c), \tau_{\bar{c}}(g))$, tal como queríamos demostrar. \square

8.4. Demostraciones correspondientes al Capítulo 5.

En esta sección demostraremos los resultados enunciados en el capítulo 5.

8.4.1. Demostración de la Proposición 5.1.1.

Demostración. Como vale la Condición 1 se verifica

$$\Pi_{gh}(M(\beta, h) | \Lambda_k(gh_k)) = E_{gh}(M(\beta, h) | \bar{A}_k \bar{L}_k) - E_{gh}(M(\beta, h) | \bar{A}_{k-1} \bar{L}_k).$$

Recordemos que

$$E_{gh}(M(\beta, h) | \bar{A}_k, \bar{L}_k) = e_k(\psi_k\{\eta_k(g), \beta\}, h),$$

y por consiguiente $e_k(\psi_k\{\eta_k(g), \beta\}, h)$ depende de Z solo mediante (\bar{A}_k, \bar{L}_k) . Ahora bien,

$$\begin{aligned} E_{gh}(M(\beta, h) | \bar{A}_{k-1} \bar{L}_k) &= E_{gh}(E_{gh}(M(\beta, h) | \bar{A}_k \bar{L}_k) | \bar{A}_{k-1} \bar{L}_k) = \\ E_{gh}(e_k(\psi_k\{\eta_k(g), \beta\}, h) | \bar{A}_{k-1} \bar{L}_k) &= E_{h_k}(e_k(\psi_k\{\eta_k(g), \beta\}, h) | \bar{A}_{k-1} \bar{L}_k), \end{aligned}$$

de donde concluimos que si $E_{gh}(M(\beta, h) | \bar{A}_k \bar{L}_k)$ depende en g y β solo mediante $\psi_k\{\eta_k(g), \beta\}$, lo mismo ocurre con $\Pi_{gh}(M(\beta, h) | \Lambda_k(p))$ y por consiguiente, se satisface la Condición 2. \square

8.4.2. Demostración del Lema 5.1.1.

Demostración. Tomemos $p = gh$ perteneciente al modelo semiparamétrico dado por \mathcal{G}_k y recordemos que $\tau_k(g)$ denota al único valor en el conjunto Υ_k para el cual $\rho_k(g) = \rho_{k, \tau_k(g)}$. Para demostrar el resultado, estudiaremos por separado los siguientes casos: (i) $k \in b$ y (ii) $k \in \bar{b}$. Comencemos considerando que $k \in b$. Tenemos entonces que

$$e_k(\rho_{k, \tau_k(g)}, h_\alpha) = e_k(\psi_k\{\eta_k(g), \beta(g)\}, h_\alpha) = E_{gh_\alpha}(M(\beta(g), h_\alpha) | \bar{A}_k, \bar{L}_k).$$

Utilizando la definición de M_k dada en (5.10), deducimos que

$$M_k(\beta, \tau_k(g), h_\alpha) = d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k(g)) \{M(\beta, h_\alpha) - E_{gh_\alpha}(M(\beta(g), h_\alpha) | \bar{A}_k, \bar{L}_k)\},$$

y por consiguiente

$$E_{gh_\alpha}(M_k(\beta(g), \tau_k(g), h_\alpha) | \bar{A}_k, \bar{L}_k) = 0.$$

Por otra parte, si consideramos $k \in \bar{b}$, tenemos que

$$e_k(\rho_{k, \tau_k(g)}, h_\alpha) = e_k(\psi_k\{\eta_k(g), \beta\}, h_\alpha) = E_{gh_\alpha}(M(\beta, h_\alpha) | \bar{A}_k, \bar{L}_k)$$

con lo cual

$$M_k(\beta, \tau_k(g), h_\alpha) = d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k(g)) \{M(\beta, h_\alpha) - E_{gh_\alpha}(M(\beta, h_\alpha) | \bar{A}_k, \bar{L}_k)\}$$

de donde resulta que, cualquiera sea β ,

$$E_{gh_\alpha} \left(M_k(\beta, \tau_k(g), h_\alpha) | \bar{A}_k \bar{L}_k \right) = 0.$$

En suma, hemos demostrado que si $p = gh \in \mathcal{G}_k$,

$$E_{gh_\alpha} \left(M_k(\beta(g), \tau_k(g), h_\alpha) | \bar{A}_k \bar{L}_k \right) = 0,$$

como queríamos ver, de donde se deduce

$$E_{gh_\alpha} \left(M_k(\beta(g), \tau_k(g), h_\alpha) \right) = 0.$$

□

8.4.3. Demostración del Lema 5.1.2.

Demostración. Sea $p = gh \in \mathcal{G}_k$. En la parte a) del Lema 5.1.1 vimos que

$$E_p \left(M_k(\beta(g), \tau_k(g), h) | \bar{A}_k \bar{L}_k \right) = 0.$$

Por otra parte, bajo la Condición 1 tenemos que h_k representa la distribución condicional de $A_k | (\bar{L}_k, \bar{A}_{k-1})$, y por consiguiente resulta que

$$\Pi_p (B | \Lambda_k(p)) = E_p (B | \bar{A}_k, \bar{L}_k) - E_p (B | \bar{A}_{k-1}, \bar{L}_k).$$

Luego, haciendo uso de la propiedad de torres para la esperanza condicional, tenemos que

$$\Pi_p \left(M_k(\beta(g), \tau_k(g), h) | \Lambda_j(p) \right) = 0, \quad \forall j \leq k.$$

Resta estudiar la proyección de $M_k(\beta, \tau, \alpha)$ en $\Lambda_j(p)$, cuando $j > k$. Para ello notemos que, tanto para el caso en el que k pertenece a b como cuando k pertenece a \bar{b} , tenemos que $e_k(\rho_{k, \tau_k}, h)$ es una variable aleatoria que solo depende de (\bar{A}_k, \bar{L}_k) y por consiguiente

$$E(e_k(\rho_{k, \tau_k}, h) | \bar{A}_m, \bar{L}_n) = e_k(\rho_{k, \tau_k}, h),$$

para $m, n \geq k$. Tenemos entonces que para $m, n \geq k$

$$E_{gh} \left(M_k(\beta, \tau_k, h) | \bar{A}_m, \bar{L}_n \right) = d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \left\{ E(M(\beta, h) | \bar{A}_m, \bar{L}_n) - e_k(\rho_{k, \tau_k}, h) \right\},$$

y por consiguiente, para $k+1 \leq j \leq K$, tenemos que

$$\begin{aligned} \Pi_p (M_k(\beta, \tau_k, h) | \Lambda_j(p)) &= E_{gh} \left(M_k(\beta, \tau_k, h) | \bar{A}_j, \bar{L}_j \right) - E_{gh} \left(M_k(\beta, \tau_k, h) | \bar{A}_{j-1}, \bar{L}_j \right) \\ &= d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \left\{ E_{gh} \left(M(\beta, h) | \bar{A}_j, \bar{L}_j \right) - E_{gh} \left(M(\beta, h) | \bar{A}_{j-1}, \bar{L}_j \right) \right\} \\ &= d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \Pi_p (M(\beta, h) | \Lambda_j(p)) \\ &= d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \pi_k \left(\psi_k \{ \eta_k(g), \beta \}, h \right), \end{aligned}$$

□

8.4.4. Demostración de la Proposición 5.1.2.

Demostración. Sea $c \subset [K]$ y $p = gh \in \mathcal{M}_c$. Tomemos $\tau_c \in \otimes_{k \in c} \Upsilon_k$ y $\alpha_{\bar{c}} \in \otimes_{k \in \bar{c}} \Xi_k$. Queremos ver que

$$E_p \left(T_{\bar{c}} \left(\beta(g), \tau, \alpha \right) \right) = 0 \quad \text{en} \quad \tau_{\bar{c}} = \tau_{\bar{c}}(g) \quad \alpha_c = \alpha_c(h_c).$$

Basta ver que para cada $k \in \bar{c}$

$$E_p \left(T_k \left(\beta(g), \tau, \alpha \right) \right) = 0 \quad \text{en} \quad \tau_{\bar{c}} = \tau_{\bar{c}}(g) \quad \alpha_c = \alpha_c(h_c).$$

En forma análoga a lo hecho al demostrar que $U(\beta, \tau, \alpha)$ es una función de estimación insesgada múltiple protegida, descomponemos a T_k de la siguiente manera

$$\begin{aligned} T_k(\beta(g), \tau_k(g), \tau_{k+1} \dots \tau_K, \alpha) &= M_k(\beta(g), \tau_k(g), h_\alpha) - d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k(g)) \times \\ &\left\{ \sum_{j>k, j \in \bar{c}} \Pi_{gh_\alpha}(M(\beta(g), h_\alpha) | \Lambda_j(gh_\alpha)) - \sum_{j \in b \cap c, j>k} \Pi_{g_{\tau_j} h_\alpha}(M(\beta_j, h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha)) \right. \\ &\left. - \sum_{j \in \bar{b} \cap c, j>k} \Pi_{g_{\tau_j} h_\alpha}(M(\beta(g), h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha)) \right\} \end{aligned}$$

con g_{τ_j} en \mathcal{U}_j , $\beta_j \in \mathbb{R}^r$.

Si $j \in c$ entonces $p = gh$ y $p_{\tau_j, \alpha} \equiv g_{\tau_j} h_\alpha$ comparten la componente h_j , de donde si $j > k$ y $j \in c \cap \bar{b}$, bajo el Supuesto 2 tenemos que

$$0 = E_{g_{\tau_j} h_\alpha} \left(\Pi_{g_{\tau_j} h_\alpha} \left(M_k(\beta(g), \tau_k, h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha) \right) \right) = E_p \left(\Pi_{g_{\tau_j} h_\alpha} \left(M_k(\beta(g), \tau_k, h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha) \right) \right),$$

luego, usando el Lema 5.1.2, resulta que

$$d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \Pi_{g_{\tau_j} h_\alpha} \left(M(\beta(g), h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha) \right) = \Pi_{g_{\tau_j} h_\alpha} \left(M_k(\beta(g), \tau_k, h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha) \right)$$

y finalmente combinando estas dos últimas igualdades obtenemos: si $j > k$ y $j \in c \cap \bar{b}$,

$$E_p \left(d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \Pi_{g_{\tau_j} h_\alpha} \left(M(\beta(g), h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha) \right) \right) = 0.$$

En forma análoga si $j > k$ y $j \in c \cap b$,

$$E_{gh_\alpha} \left(d_k(\bar{A}_k, \bar{L}_k, \alpha, \tau_k) \Pi_{g_{\tau_j} h_\alpha} \left(M(\beta_j, h_\alpha) | \Lambda_j(g_{\tau_j} h_\alpha) \right) \right) = 0.$$

Por ahora, tenemos que

$$E_p \left(T_k(\beta(g), \tau_k(g), \tau_{k+1} \dots \tau_K, \alpha) \right) = E_p \left(M_k(\beta(g), \tau_k(g), h_\alpha) - \sum_{j>k, j \in \bar{c}} \Pi_{gh_\alpha} \left(M_k(\beta(g), \tau_k(g), h_\alpha) | \Lambda_j(gh_\alpha) \right) \right)$$

Ahora bien, como

$$\Pi_p \left(M_k(\beta(g), \tau_k(g), h) | \Lambda_j(p) \right) = 0, \quad \forall j \leq k,$$

entonces podemos escribir

$$E_p \left(T_k(\beta(g), \tau_k(g), \tau_{k+1} \dots \tau_K, \alpha) \right) = E_p \left(M_k(\beta(g), \tau_k(g), h_\alpha) - \sum_{j \in \bar{c}} \Pi_{gh_\alpha} (M_k(\beta(g), \tau_k(g), h_\alpha) | \Lambda_j(gh_\alpha)) \right).$$

Recordemos que p y $p^* = gh_\alpha$ pertenecen al submodelo convexo \mathcal{P}_{gh_α} y satisfacen el Supuesto 1. Luego, combinando las ideas desarrolladas en la demostración del Lema 4.2.1 y el Corolario 4.2.1, tenemos que

$$\begin{aligned} & E_p \left(M_k(\beta(g), \tau_k(g), h_\alpha) - \sum_{j \in \bar{c}} \Pi_{gh_\alpha} (M_k(\beta(g), \tau_k(g), h_\alpha) | \Lambda_j(gh_\alpha)) \right) = \\ & E_{gh_\alpha} \left(M_k(\beta(g), \tau_k(g), h_\alpha) - \sum_{j \in \bar{c}} \Pi_{gh_\alpha} (M_k(\beta(g), \tau_k(g), h_\alpha) | \Lambda_j(gh_\alpha)) \right) \end{aligned}$$

Ahora bien, como $k \in \bar{c}$ resulta que $p \in \mathcal{G}_k$ y entonces $E_{gh_\alpha} \left(M_k(\beta(g), \tau_k(g), h_\alpha) \right) = 0$, además $E_{gh_\alpha} \left(\Pi_{gh_\alpha} (M_k(\beta(g), \tau_k(g), h_\alpha) | \Lambda_j(gh_\alpha)) \right) = 0$ de donde obtenemos el resultado. \square

Bibliografía

- [1] Bang H, Robins J. (2005). Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*. **61**, 962-972.
- [2] Berkson, J. (1946). Limitations of the Application of Fourfold Table Analysis to Hospital Data. *Biometrics Bulletin* 2 (3): 47–53.
- [3] Bianco A, Boente G, González-Manteiga W, Pérez-González A. (2009). Estimation of the marginal location under a partially linear model with missing responses. *Computational Statistics and Data Analysis*. **54**, 546-564.
- [4] Bianco A, Boente G.(2004). Robust Estimators in Semiparametric Partly Linear Regression Models. *Journal of Statistical Planning and Inference*. **122**, 229-252.
- [5] Bickel P. (1982). On Adaptive Estimation. *Annals of Statistics*. **10**, 647-671.
- [6] Cao W, Tsiatis A, Davidian M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96**, 723-34.
- [7] Carpenter J, Kenward, M, Vansteelandt, S.(2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *European Journal of Research Methods for the Behavioral and Social Sciences*. **6**, 37-48.
- [8] Carpenter J, Kenward, M, Vansteelandt, S.(2006). A comparison of múltiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society.)***169**, 571-584.
- [9] Goetgeluk S, Vansteelandt S, Goetghebeur E. (2009). Estimation of controlled direct effects. *Jornal of the Royal Statistical Society*. **70**, 1049-1066.
- [10] Hardle W, Liang H, Gao J. (2000). *Partially Linear Models*. Springer, Physica-Verlang.

- [11] Kang H, Schafer J. (2007). Demystifying Double Robustness: A Comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*. **22**, 523-539.
- [12] Lunceford J, Davidian M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*. **23**, 2937-2960.
- [13] Murphy, S, van der Laan M, Robins, J. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*. **96**, 1410-1423.
- [14] Newey W. (1990). Semiparametric Efficiency Bounds. *Journal of Applied Econometrics*. **5** 99-135.
- [15] Orellana L, Rotnitzky A, Robins J M. (2010). Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part I: Main Content. *The International Journal of Biostatistics*: **6** Iss. 2.
- [16] Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82 (4): 669-688
- [17] Robins J, Rotnitzky A. (2001). Comment on the Bickel and Kwon article “Inference for semiparametric models: Some questions and an answer”. *Statistica Sinica*. **11** 920-936.
- [18] Robins J, Rotnitzky A, van der Laan M. (2000). Comment on “On Profile Likelihood” by S.A. Murphy and A.W. van der Vaart. *Journal of the American Statistical Association, Theory and Methods*. **95**, 431-435.
- [19] Robins J. (2000) Robust Estimation in Sequentially Ignorable Missing Data and Causal Inference Models. *Proc Am Stat Assoc. Section on Bayesian Statistical Science*. **95** 6-10.
- [20] Robins J. (1997). Causal Inference from Complex Longitudinal Data, in *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics*. **120**, 69-117.
- [21] Robins J, Rotnitzky A, Zhao L-P. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*. **90**, 106-121.
- [22] Robins J, Mark SD, Newey WK. (1992). Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders. *Biometrics*, **48**, 479-495.

- [23] Robinson P. (1988). Root-n-Consistent Semiparametric Regression. *Econometrica*. **56**, 931-954.
- [24] Rotnitzky A, Lei Q, Sued M, Robins J. (2012). Improved double-robust estimation in missing data and causal inference models. *Biometrika*. **99**, 439-456.
- [25] Rubin, B. (1976). Inference and Missing Data. *Biometrika*. **62**,581-592.
- [26] Scharfstein D, Rotnitzky a, Robins J.(1999). Adjusting for nonignorable drop-Out using semiparametric nonresponse Models. *Journal of the American Statistical Association* **94**, 448-499.
- [27] Severini T, Staniswalis J.(1994). Quasi-likelihood Estimation in Semiparametric Models.*Journal of the American Statistical Association*. **89**, 501-511 .
- [28] Speckman P. (1988). Kernel Smoothing in Partial Linear Models. *Journal of the Royal Statistical Society. Series B (Methodological)*. **50**, 413-436.
- [29] Tan Z. (2008). Comment: Improved Local Efficiency and Double Robustness. *Interntl J. of Biostatist.* **4**, article 10.
- [30] Tan Z. (2006). Regression and Weighting Methods for Causal Inference Using Instrumental Variables Zhiqiang TAN. *Journal of the American Statistical Association* **101**, nro. 476.
- [31] Tchetgen Tchetgen E, Robins J, Rotnitzky A. (2010). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika*. **97**, 171-180.
- [32] Tchetgen -Tchetgen, E. (2009). A Commentary on "A Commentary on G. Molenaar's Review of Missing Data Methods". *Drug Information Journal*. **43**, 433-435.
- [33] Tsiatis A. (2006). *Semiparametric Theory and Missing Data*. Springer Series in Statistics.
- [34] van der Laan M, Robins J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer Verlag.
- [35] Vansteelandt, S., Rotnitzky, A. and Robins, J. (2013). Multiple robust estimation in models for repeated measures with non-ignorable non-response. Working paper.
- [36] Vansteelandt, S, Rotnitzky A, Robins J. (2008). Estimation of regression models for the mean of repeated outcomes under non-ignorable non-monotone non-response. *Biometrika*. **94**, 841-860.

- [37] Vansteelandt S, VadderWeele T, Tchetgen T, Robins J. (2008). Multiply Robust Inference for Statistical Interactions. *Journal of the American Statistical Association*. **103**, 1693-1704.
- [38] van der Vaart, A. W., (2002). Semiparametric statistics. Ecole d'ete de St Flour 1999 (eds: P. Bernard). *Lecture Notes in Mathematics* **1781**, 331-457. Springer Verlag, Berlin.
- [39] van der Vaart, A. W.,(1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.
- [40] Wang, Q. and Sun, Z. (2007). Estimation in partially linear models with missing responses at random. *Journal of Multivariate Analysis* **98**, 1470-1493.
- [41] Wang Q, Linton O, Hardle W.. (2004) . Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*. **99**, 334-345.