

UNIVERSIDAD NACIONAL DEL LITORAL
FACULTAD DE INGENIERÍA QUÍMICA

TESIS PRESENTADA COMO PARTE DE LOS REQUISITOS DE LA UNIVERSIDAD NACIONAL DEL
LITORAL PARA LA OBTENCIÓN DEL GRADO ACADÉMICO DE

Master en Matemática

EN EL CAMPO DE: **Estadística**

TÍTULO DE LA TESIS:

**Big Data para Quimiometría: Distribución asintótica
del estimador PLS en alta dimensión**

INSTITUCIÓN DONDE SE REALIZÓ:

Facultad de Ingeniería Química
(Universidad Nacional del Litoral)

AUTOR:

Lic. Jerónimo Basa

DIRECTORES DE TESIS:

Dr. Miguel Andrés Marcos y Dra. Liliana Forzani

DEFENDIDA ANTE EL JURADO COMPUESTO POR:

Dra. Raquel Mariela Sued

Dra. Ana María Bianco

Dra. Daniela Andrea Rodriguez

AÑO DE PRESENTACIÓN: 2022

A Erika y Florencia, los dos pilares de mi vida

ÍNDICE GENERAL

Resumen	III
Introducción	V
Notaciones	VII
Capítulo 1. Regresión lineal y el mundo de la alta dimensión	1
1.1 Modelo clásico de regresión lineal múltiple	2
1.2 Alta dimensionalidad	4
1.3 Estimación de la matriz de covarianza.....	5
Capítulo 2. Introducción a Partial Least Squares	9
2.1 Algoritmo NIPALS.....	12
2.2 Envelopes y su relación con PLS	15
2.3 Resultados para p fijo.....	19
Capítulo 3. Resultados asintóticos	21
3.1 Predicción PLS	21
3.2 Cantidades que gobiernan las distribuciones.....	22
3.3 Distribución del estimador	24
3.4 Intervalos de confianza.....	28
3.5 Intervalos para predicción	29
Capítulo 4. Simulaciones y aplicaciones	33
4.1 Distribución asintótica.....	33
4.2 Coverage.....	33
4.3 Datos: mejillones.....	36
4.4 Datos: dopamina.....	39
Bibliografía	43

Apéndice A.	Notación	47
Apéndice B.	Algunos resultados conocidos	49
Apéndice C.	El modelo y los estimadores	51
Apéndice D.	Demostración de la Proposición 3.1.....	57
Apéndice E.	Demostración de la Proposición 3.2 y del Teorema 3.3.....	61
Apéndice F.	Demostración del Teorema 3.5.....	81
Apéndice G.	Demostración del Teorema 3.6	85
Apéndice H.	Demostración del Teorema 3.7	89
Apéndice I.	Demostración del Teorema 3.8	93

RESUMEN

Muchos de los problemas en estadística se basan en estudiar el comportamiento de una o más variables llamadas respuesta a partir de un conjunto de variables llamadas predictoras. Para ello se toman mediciones u observaciones, se hace un análisis exploratorio de estos datos y se propone un modelo matemático que explique su relación. Uno de los modelos más usados y estudiados es la regresión lineal. Cuando el número de predictores p es grande comparado con el tamaño de observaciones n , muchos de los métodos tradicionales en estadística comienzan a dar estimaciones que pueden ser muy pobres. En particular, en el contexto de regresión lineal se vuelve un problema estimar el vector de parámetros con el método de mínimos cuadrados, usado tradicionalmente en el caso $n > p$.

Cuando nos encontramos en una situación donde $p > n$ es usual aplicar alguna herramienta de reducción de dimensiones, es decir reducir la cantidad de predictores sin que se pierda información acerca de la respuesta. Dentro de los métodos de reducción de dimensiones encontramos el llamado *Partial Least Squares* (PLS), frecuentemente encontrado en la literatura de la quimiometría. PLS nació en 1966 como un método iterativo dado por un algoritmo que obtenía buenas predicciones según los quimiométricos, pero que no tenía un modelo matemático asociado y por ello muchas de sus propiedades eran desconocidas. Años más tarde en 1990 I. S. Helland mostró que en un contexto de regresión lineal, el algoritmo PLS en realidad obtiene una expresión cerrada para la estimación de parámetros, obteniendo también una correspondiente fórmula para la versión poblacional. Finalmente en un trabajo de Cook, Helland y Su del 2013 se describe un modelo de reducción conocido como *envelope regression* y se prueba que PLS es un modelo envelope para los predictores. En 2018 Cook y Forzani estudian este modelo en un contexto de alta dimensión con una componente y obtienen resultados sobre la consistencia n, p asintótica de la predicción dada por el estimador generado con PLS.

El objetivo de esta tesis es extender los resultados de Cook y Forzani para hallar la distribución n, p asintótica de la predicción PLS. En un contexto general, encontramos la presencia de un sesgo no aleatorio en esta convergencia y damos condiciones para que sea despreciable.

Proponemos estimadores para la varianza asintótica y para el sesgo, mostrando su consistencia. Utilizando éstos deducimos a su vez intervalos de confianza y predicción para el parámetro dado por el modelo. Finalmente, mostramos ejemplos y simulaciones para ilustrar los resultados expuestos.

INTRODUCCIÓN

El método *Partial Least Squares* (PLS) ha sido ampliamente usado en muchas áreas de la ciencia. Fue planteado inicialmente por [Wold, 1966b] en el contexto de regresión lineal como un algoritmo iterativo que obtenía predicciones muy buenas en el estudio de quimiometría en casos donde el número de observaciones n no era necesariamente mayor que el número de variables predictoras p . Al tratarse de un algoritmo, no había un modelo matemático de fondo y por ello gran parte de la comunidad estadística rechazaba este método. En [Helland, 1988] el autor observa que bajo el modelo de regresión lineal el algoritmo obtiene una expresión cerrada para el estimador del parámetro, lo cual le permite deducir una fórmula poblacional de éste y por lo tanto tener un modelo PLS de regresión lineal. Años más tarde en [Cook et al., 2013] los autores obtienen dos resultados importantes. Primero, el modelo PLS puede verse como un modelo de reducción de dimensiones en regresión lineal conocido como *envelope regression*. En segundo lugar, obtienen la distribución asintótica del estimador PLS cuando $n \rightarrow \infty$ y p es fijo. En [Cook and Forzani, 2018] se plantea el estudio de PLS en un contexto de alta dimensión, es decir cuando $n, p \rightarrow \infty$, obteniendo la consistencia de la predicción dada por el estimador PLS con una componente. En esta tesis mostramos la distribución n, p -asintótica de la predicción usando el algoritmo PLS para una componente y además obtenemos intervalos de confianza asintóticos para la media y valores predichos.

El Capítulo 1 está dedicado a introducirnos en la regresión lineal, la idea básica de este modelo y parte de la notación que usaremos luego. Seguidamente hablamos un poco sobre las limitaciones de algunos métodos cuando se considera el caso particular en que el tamaño de una muestra no es superior al tamaño de variables predictoras. Luego exponemos algunas ideas para enfrentar estas limitaciones. Una de estas ideas motiva lo expuesto en los siguientes capítulos.

Por otro lado, en el Capítulo 2 nos concentramos en introducir las ideas básicas que conforman el método Partial Least Squares (PLS). Comenzamos con una breve reseña histórica de PLS, la aplicabilidad que ha tenido en distintas áreas de investigación así como sus principales usos. Hablamos de los algoritmos que dieron inicio a este método [Wold, 1966b],

[de Jong, 1993] y una importante observación en [Helland, 1988] nos permite obtener una expresión del estimador en regresión lineal. Además introducimos las ideas básicas de *envelope regression* [Cook et al., 2007] para mostrar su relación con PLS. El capítulo concluye con un resultado de [Cook et al., 2013] para un caso particular de este método.

En el Capítulo 3 exponemos nuevamente nuestro contexto de estudio y el enfoque que damos a nuestro desarrollo. Comenzamos hablando sobre el trabajo en [Cook and Forzani, 2018] y exponemos la idea de los autores para usarla como punto de partida. Una serie de resultados parciales conllevan a nuestro teorema principal, el cual usaremos para deducir ciertos corolarios y otros resultados.

Finalmente dedicamos el Capítulo 4 a presentar una serie de simulaciones y ejemplos que puedan ilustrar los resultados teóricos desarrollados. La tesis concluye con una sección de apéndices donde exponemos las demostraciones de los resultados presentados a lo largo del trabajo.

NOTACIONES

Reproduciremos en este apartado las notaciones y definiciones que serán utilizadas a lo largo de esta tesis con el objetivo de que la lectura resulte fluida.

Notación. Utilizaremos la notación $\mathbb{R}^{m \times n}$ para indicar el conjunto de todas las matrices reales de dimensión $m \times n$. La traspuesta de una matriz o de un vector \mathbf{z} se denota como \mathbf{z}^T . Para $\mathbf{M} \in \mathbb{R}^{m \times n}$, $\text{span}(\mathbf{M}) \subset \mathbb{R}^m$ es el subespacio generado por los vectores columnas de \mathbf{M} . Las letras en negrita serán usadas para todos los elementos matemáticos que no sean escalares.

Definición. Sea \mathbf{M} una matriz simétrica de dimensión $n \times n$. Esta matriz \mathbf{M} se dice semi-definida positiva, y lo denotamos como $\mathbf{M} \geq 0$, si cumple que para todos los vectores no nulos $\mathbf{z} \in \mathbb{R}^n$ vale $\mathbf{z}^T \mathbf{M} \mathbf{z} \geq 0$. Si la ecuación anterior satisface la desigualdad de forma estricta, es decir $\mathbf{z}^T \mathbf{M} \mathbf{z} > 0$ para todo vector no nulo $\mathbf{z} \in \mathbb{R}^n$, se dice que \mathbf{M} es definida positiva.

Definición. El rango de las columnas de una matriz $\mathbf{M} \in \mathbb{R}^{m \times n}$ es el número máximo de columnas linealmente independientes que esta contiene. De forma análoga, se define el rango de las filas de \mathbf{M} como el número máximo de filas linealmente independientes. Para cualquier matriz \mathbf{M} el rango de las filas es igual al rango de las columnas y este número es llamado el rango de \mathbf{M} y denotado por $\text{rank}(\mathbf{M})$.

Definición. El producto de Kronecker de las matrices $\mathbf{A} : m \times n$ y $\mathbf{B} : p \times q$, denotado con $\mathbf{A} \otimes \mathbf{B}$, es la matriz $mp \times nq$

$$\begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix}.$$

Definición. La función *vectorización* $\text{vec} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{mn}$ actúa sobre una matriz apilando las columnas como un solo vector. Es decir, si \mathbf{a}_i son las columnas de \mathbf{A} , entonces

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}$$

Definición. Sea $\mathbf{v} \in \mathbb{R}^p$, denotamos por $\mathbf{P}_{\mathbf{v}}$ a la matriz proyección en la dirección de \mathbf{v} , es decir, $\mathbf{P}_{\mathbf{v}} = \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}}$. Además, decimos que $\mathbf{Q}_{\mathbf{v}} = \mathbf{I} - \mathbf{P}_{\mathbf{v}}$ es la matriz de proyección en el complemento ortogonal de $\text{span}(\mathbf{v})$. De manera más general, usaremos $\mathbf{P}_{(\cdot)}$ y $\mathbf{Q}_{(\cdot)}$ para denotar las proyecciones sobre un correspondiente espacio y su complemento ortogonal.

Notación

- Denotamos con $\|\cdot\|$ a la norma L^2 y $\|\cdot\|_F$ a la norma Frobenius de matrices.
- Usaremos $\text{tr}(\cdot)$ para denotar la traza de una matriz.
- El superíndice $^\perp$ denota el complemento ortogonal de un espacio.
- Usamos \rightsquigarrow para indicar convergencia en distribución. Además usaremos \xrightarrow{p} para denotar convergencia en probabilidad.

Si A y B son dos sucesiones

- $A = O(B)$ significa que existe un número positivo K y un natural $N_0 = N_0(K)$ tal que $|A| \leq K|B|$ para todo $n \geq N_0$.
- Decimos que $A \asymp B$ si $A = O(B)$ y $B = O(A)$.

Sea X_n una sucesión de variables aleatorias

- Decimos que $X_{n,p} = O_p(a_{n,p})$ si para todo $\varepsilon > 0$ existe una constante $K = K(\varepsilon)$ y un número natural $N_0 = N_0(\varepsilon)$ tal que

$$P\left(\left|\frac{X_{n,p}}{a_{n,p}}\right| \leq K\right) \geq 1 - \varepsilon, \quad \forall n, p \geq N_0.$$

- Decimos que $X_{n,p} = o_p(1)$ si $X_{n,p} \xrightarrow{p} 0$ cuando $n, p \rightarrow \infty$.

Notación. Dada una muestra $\mathbf{X}_1, \dots, \mathbf{X}_n$ del vector aleatorios $\mathbf{X} \in \mathbb{R}^p$ e $Y \in \mathbb{R}$ denotamos

- $\bar{\mathbf{X}} \doteq \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$: vector promedio para la muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$.
- $\Sigma = \text{var}(\mathbf{X})$ es la matriz de covarianza de \mathbf{X} .
- $\sigma = \text{cov}(Y, \mathbf{X})$ es el vector de covarianza entre Y y \mathbf{X} .

▪

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T, \quad \hat{\sigma} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(Y_i - \bar{Y}) \quad (0.1)$$

son los estimadores clásicos de Σ y σ .

CAPÍTULO 1

REGRESIÓN LINEAL Y EL MUNDO DE LA ALTA DIMENSIÓN

Este capítulo introductorio está dedicado a explicar una de las ideas fundamentales de la estadística: la regresión, que será el soporte para el contenido a desarrollar en el resto de la tesis. Además buscamos echar un poco de luz sobre su importancia, la dificultad que presentan, y la necesidad de su estudio.

El concepto general de regresión es antiguo, de hecho vemos sus primeras apariciones en [Galton, 1886] en un estudio sobre genética donde se estudia cómo se transmite el carácter humano de la altura de una generación a la siguiente. Podríamos decir informalmente que la regresión busca explicar la relación entre una o más variables dependientes y una o muchas variables independientes. Desde luego, existen muchas formas en que podríamos analizar este fenómeno pero aquí nos concentramos en la más común de todas: la llamada regresión lineal.

Investigadores basan sus resultados en observaciones e información que recolectan del mundo real (experimentos). Para relacionar estas observaciones, es necesario establecer el método descriptivo que mejor explica los fenómenos observados. Para ello, se selecciona un modelo (de regresión) que explica alguna característica en función de otras. Todo modelo de regresión se conforma de ciertas piezas fundamentales:

- Un conjunto de variables independientes, que se corresponde con datos observados y que nosotros representaremos por un vector \mathbf{X}_i , donde el subíndice i corresponde con el i -ésimo sujeto observado.
- Una o más variables dependientes, que también observamos y que nosotros indicaremos con \mathbf{Y}_i (en esta tesis trabajamos sólo con una variable dependiente).
- Un término de error que no es observable en los datos y que denotamos como ε_i .

- Un modelo estadístico, que generalmente se caracteriza con parámetros desconocidos.

La familia de métodos de regresión pueden representarse por la expresión

$$Y_i = f(\mathbf{X}_i) + \varepsilon_i.$$

El objetivo de los investigadores es seleccionar y estimar la función f que mejor se ajusta (explica) los datos. Notemos que luego del proceso de recolección de datos y elección del modelo, nuestra siguiente meta es el cálculo de los parámetros desconocidos. Desde luego, no somos capaces de hallar estos valores verdaderos, debido a la presencia del ruido no observable y que en general no hay una ley conocida de relación, por lo tanto debemos estimarlos usando la información que tenemos en los datos. A partir de esto, lo que en realidad estamos haciendo es reduciendo la información de los datos a unas pocas cantidades que, esperamos, representen adecuadamente toda (o al menos la gran mayoría de) la información relevante que hay detrás de la relación entre las variables observadas. Esta idea de *estimación de parámetros* desconocidos fue introducida en [Fisher, 1922], una de las referencias más importantes que contiene la mayoría del lenguaje estadístico que usamos hoy en día. A continuación, vamos a desarrollar el enfoque que a nosotros nos importa, que es la elección de una función f lineal.

1.1. Modelo clásico de regresión lineal múltiple

La regresión es un concepto básico y comúnmente usado en el análisis predictivo de datos. Dentro de la familia de métodos de regresión, el modelo lineal es el más estudiado de forma rigurosa y uno de los más usados por la comunidad científica debido a su sencillez y fácil manipulación, que no por ello minimizan el impacto aplicativo que puede tener este modelo.

Este trabajo de tesis recae sobre el modelo de regresión lineal clásico con una sola respuesta, que se basa en que la variable respuesta $Y \in \mathbb{R}$ se relaciona linealmente con la variable predictora $\mathbf{X} \in \mathbb{R}^p$ a través del siguiente *modelo de regresión lineal*

$$Y = \mu_Y + \boldsymbol{\beta}^T (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) + \varepsilon, \quad (1.1)$$

donde ε es independiente de \mathbf{X} con $E(\varepsilon) = 0$, $\text{var}(\varepsilon) = \tau^2$ y donde $\boldsymbol{\mu}_{\mathbf{X}} = E(\mathbf{X})$, $\mu_Y = E(Y)$. Aquí $\boldsymbol{\beta} \in \mathbb{R}^p$ es un vector desconocido de parámetros y estimar f se reduce a estimar $\boldsymbol{\beta}$. Asumimos que tenemos un conjunto de n datos independientes (Y_i, \mathbf{X}_i) , $i = 1, 2, \dots, n$ tomados

de las variables (Y, \mathbf{X}) siguiendo el modelo (1.1). Si usamos la notación

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbb{X} = \begin{bmatrix} X_{11} - \bar{X}_1 & X_{12} - \bar{X}_2 & \dots & X_{1p} - \bar{X}_p \\ X_{21} - \bar{X}_1 & X_{22} - \bar{X}_2 & \dots & X_{2p} - \bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} - \bar{X}_1 & X_{n2} - \bar{X}_2 & \dots & X_{np} - \bar{X}_p \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

entonces el modelo para la muestra puede escribirse en notación matricial como

$$\mathbf{Y} = \bar{Y}\mathbf{1}_n + \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbf{Y}, \boldsymbol{\varepsilon} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^p, \mathbb{X} \in \mathbb{R}^{n \times p}$$

donde $\mathbf{1}_n$ es el vector de unos de dimensión n y \mathbb{X} se conoce usualmente como la matriz de diseño. De esta forma, el problema se reduce a hallar el estimador $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ que mejor ajusta este modelo. Una forma de lograrlo es usar el método de mínimos cuadrados obteniendo el estimador que denotaremos como $\hat{\boldsymbol{\beta}}_{ols}$. En términos de los estimadores muestrales de $\boldsymbol{\sigma} = \text{cov}(\mathbf{X}, Y)$ y $\boldsymbol{\Sigma} = \text{var}(\mathbf{X})$ se tiene que

$$\hat{\boldsymbol{\beta}}_{ols} = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\sigma}}, \quad (1.2)$$

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} \quad (1.3)$$

donde $\hat{\boldsymbol{\Sigma}}$ y $\hat{\boldsymbol{\sigma}}$ están definidos en (0.1) siempre que $\hat{\boldsymbol{\Sigma}}^{-1}$ exista, y esto ocurre con probabilidad 1 si $n > p + 1$. Notemos que $\hat{\boldsymbol{\Sigma}} = (n - 1)^{-1}\mathbb{X}^T\mathbb{X}$ y $\hat{\boldsymbol{\sigma}} = (n - 1)^{-1}\mathbb{X}^T\mathbf{Y}$. El método de mínimos cuadrados es muy antiguo y fue inicialmente presentado por Gauss en su estudio sobre el movimiento de los cuerpos celestes [Stigler, 1981]. Una vez que el investigador cuenta con la estimación de $\boldsymbol{\beta}$ puede usarla para hallar valores predichos \hat{Y} y sus intervalos de confianza, que pueden ser útiles para analizar la eficiencia del modelo propuesto, incluso puede hacer predicciones sobre la variable respuesta en valores no observados en la base de datos (con la precaución no de extrapolar, es decir, desviarse del rango de variables predictoras observadas). Una advertencia común que suele figurar en la bibliografía es el hecho de que debemos tener una suficiente cantidad de observaciones para que nuestro análisis de los datos (con el modelo propuesto) tenga algún sentido. Por ejemplo, imaginemos que vamos a ajustar un modelo con una variable respuesta y una predictora ($p = 1$) a través de una recta. Debemos estimar dos parámetros: la ordenada al origen y la pendiente. En la Figura 1.1 se ajusta un modelo lineal para 100 observaciones. La recta de mínimos cuadrados es aquella que mejor ajusta los datos de acuerdo a su comportamiento. Pero ¿qué ocurre si sólo disponemos de dos observaciones? Estaríamos intentando hallar la recta que mejor ajusta dos puntos. Dicha recta es ni más ni menos que la recta determinada por esos dos puntos, y por lo tanto el ajuste será perfecto. El

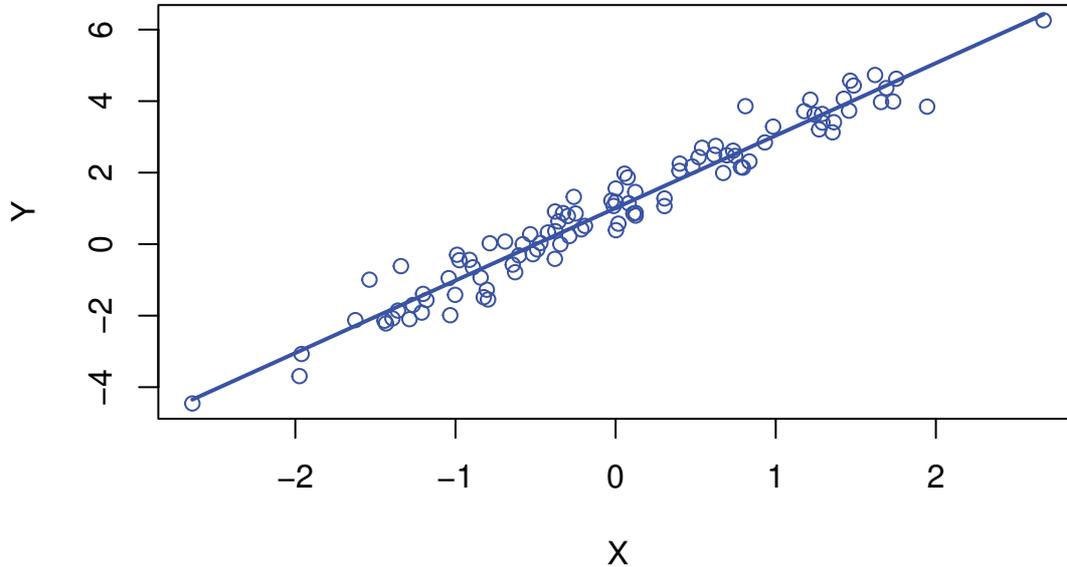


FIGURA 1.1. Representación clásica de una regresión lineal simple.

error, es decir la diferencia entre el valor verdadero y el valor predicho, es 0 para los datos de la muestra. Lo mismo ocurre si tuviéramos que ajustar un modelo con dos variables predictoras ($p = 2$) y sólo disponemos de dos observaciones. En este caso existen infinitos planos que contienen a dos puntos y nuevamente la diferencia entre el valor verdadero y el predicho es 0 pues el ajuste es perfecto. Todo esto nos lleva a concluir que si estamos ajustando un modelo con p parámetros, uno debe usar $n > p + 1$ observaciones para tener unicidad de la expresión (1.2). Esta es una situación que se ha manifestado en casi todos los experimentos desarrollados en el siglo 20.

1.2. Alta dimensionalidad

Si observamos la expresión (1.2), claramente una de las mayores dificultades es obtener la inversa de $\hat{\Sigma}$ (o $\mathbb{X}^T \mathbb{X}$). Para la existencia de dicha inversa es necesario que la matriz sea de rango completo, y esta suposición se cumple con probabilidad 1 si $n > p$ ¿Qué nos dice esta desigualdad? Simplemente que tenemos más observaciones que variables predictoras. Podemos interpretar que el número p representa la cantidad de información (o *features*) que tenemos sobre cada observación. Hasta hace unos años éste era el contexto clásico en la estadística: tener

poca información sobre un número grande de sujetos de estudio. Sin embargo, ocurrió un cambio de paradigma que supone ahora estudiar casos donde sucede lo opuesto, es decir $p \gg n$. Desde luego, si $p > n$ entonces la matriz dada por \mathbb{X} satisface $\text{rank}(\mathbb{X}) \leq n - 1$, y como el rango resulta a lo sumo n , $\mathbb{X}^T \mathbb{X}$ (y por lo tanto $\widehat{\Sigma}$) tendrá rango incompleto. De esta manera, el estimador $\widehat{\beta}_{ols}$ ya no es único, y debemos pensar en buscar alternativas para este método de estimación. Este problema en realidad es uno de los tantos que ocurre debido a la configuración $p \gg n$.

El desarrollo de nuevas tecnologías para la recolección de datos ha hecho posible que se puedan tomar una gran cantidad de medidas de manera simultánea. Estos datos producidos a tan gran escala se llaman de *alta dimensionalidad*, y están presentes hoy en día en casi todas las ramas de la humanidad, incluyendo medicina, genética, astronomía, finanzas, ciencias sociales, etcétera. El hecho de tener acceso a un gran número de información parece, en principio, una bendición. Ya que mientras más datos tengamos, mejor será nuestro análisis estadístico. Desafortunadamente, el análisis de datos en alta dimensión es muy desafiante, ya que la separación entre información útil del ruido es muy difícil de lograr. Este problema se conoce como *la maldición de la dimensionalidad*. La mayor parte de la estadística clásica desarrollada en el siglo 20 se enfocó en el estudio de datos donde el número n de unidades experimentales era significativamente mayor comparado con el número de características p de cada individuo. Más aún, gran parte de la teoría estadística desarrollada eran resultados asintóticos bajo una configuración donde p es fijo y n tiende a infinito. Desde luego, esta teoría es muy útil para analizar datos donde n es grande, y p es chico, pero puede presentar serios problemas para datos modernos de alta dimensionalidad. Algunos ejemplos claros donde esta configuración se hace presente pueden verse en [Giraud, 2014]. Por ello se ha dado lugar a un aumento en el esfuerzo de la comunidad estadística y de analistas de datos por lograr un desarrollo de nuevas herramientas que permitan afrontar esta dificultad.

1.3. Estimación de la matriz de covarianza

Una interesante observación es el hecho de que la matriz de covarianza empírica, además de no ser invertible, no es representativa en el contexto de alta dimensión cuando p es de la escala de n o mayor. Por ejemplo, supongamos que $\mathbf{X} \in N_p(\mathbf{0}, \mathbf{I})$ y que tenemos una muestra de n observaciones i.i.d. de este vector. La matriz de covarianza muestral para las observaciones $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$ se denota por $\widehat{\Sigma}$. Una propiedad conocida es que como $E(\mathbf{X}\mathbf{X}^T) = \mathbf{I}$, el estimador $\widehat{\Sigma}$ converge a la matriz identidad cuando n tiende a infinito y p es fijo por la ley de los grandes

números. Por lo tanto, es de esperar que el espectro (conjunto de autovalores) de $\hat{\Sigma}$ se concentre alrededor de 1 cuando n es significativamente mayor que p . Sin embargo, esta propiedad se pierde cuando p aumenta proporcionalmente con el tamaño de muestra. En la Figura 1.2 se observa el comportamiento del espectro de $\hat{\Sigma}$ para distintas relaciones de n, p .

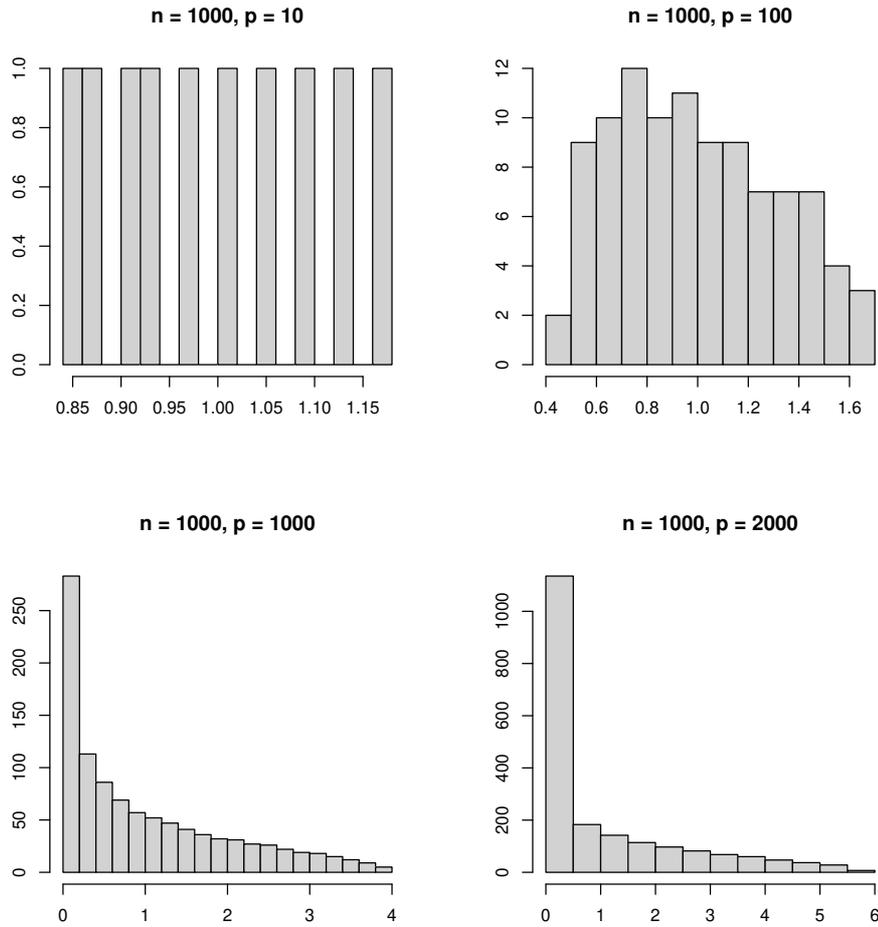


FIGURA 1.2. Histogramas para la distribución de los autovalores de $\hat{\Sigma}$ para distintos valores de p cuando n es fijo.

TABLA 1.1. Norma Frobenius de las matrices $\Sigma = \mathbf{I}$ y $\hat{\Sigma}$ para los casos presentados en la Figura 1.2.

p	$\ \Sigma\ _F = \sqrt{p}$	$\ \hat{\Sigma}\ _F$
10	3.162278	3.167374
100	10	10.45743
1000	31.62278	44.81538
2000	44.72136	77.56161

En el histograma superior izquierdo de la Figura 1.2 vemos el caso para $n \gg p$ donde se observa un comportamiento uniforme de los autovalores de $\hat{\Sigma}$. En los demás histogramas vemos que no se sigue este patrón de distribución uniforme. Además en la Tabla 1.1 vemos que la norma de $\hat{\Sigma}$ parece alejarse de la norma de Σ . Incluso en este caso aparentemente sencillo, debido a que se trata de una distribución normal estándar, la estimación $\hat{\Sigma}$ es un pobre estimador de Σ en estos casos. Más aún [Marcenko and Pastur, 1967] mostraron que para el caso $\Sigma = \mathbf{I}$ la densidad empírica del espectro (es decir, la distribución de los autovalores) de $\hat{\Sigma}$ no converge a una masa puntual en 1 como esperaríamos.

Cuando p es fijo y no depende de n , el estimador clásico de la matriz de covarianza es un estimador consistente de Σ con orden $n^{-1/2}$. Pero cuando p es más grande que n o depende de n el estimador puede ser muy diferente de Σ . Por ejemplo, como se ve en la Figura 1.2 mientras mayor es el cociente p/n hay una mayor concentración de los autovalores cerca de 0.

De este modo, se abre un debate sobre qué opciones considerar para la estimación de Σ (cuando $p > n$) que nos sea útil para hallar los estimadores de los parámetros del modelo propuesto. Algunas soluciones para este problema consisten en suposiciones sobre la forma de la matriz Σ . Por ejemplo, existen métodos de regularización conocidos como *hard thresholding* que obtienen un estimador consistente [Bickel and Levina, 2008]; descomposición modificada de Cholesky [Huang et al., 2006]; construcción de estimadores sparse basados en penalidad LASSO [Rothman et al., 2008]. Algo que tienen en común la gran mayoría de estos métodos es que dependen de suposiciones que se hacen sobre la estructura (desconocida) de la matriz de covarianza para obtener buenos estimadores o para resolver la dificultad que supone hallar la inversa del estimador. En general esta estructura impuesta para tener consistencia es que la matriz sea sparse (rala) suponiendo así que muchas de las covarianzas son cero y por consiguiente se considera un modelo donde las predictoras están poco correlacionadas marginal o condicionalmente. Otra posibilidad que abre camino a nuevos métodos es cambiar la forma en

que hallamos el estimador en (1.2) de manera tal que no se requiera calcular la inversa. Este enfoque será el que nosotros estudiaremos en los próximos capítulos.

INTRODUCCIÓN A PARTIAL LEAST SQUARES

El nombre *Mínimos Cuadrados Parciales o Partial Least Squares* (PLS) se usa para hacer referencia a una colección muy general de métodos estadísticos [Vinzi et al., 2010]. En efecto PLS puede verse como una familia de procesos que incluyen un gran número de variantes y posibilidades que ayudan a lograr diversas tareas, dependiendo del objetivo del analista. Unos de los tantos métodos que pertenecen a esta familia se usa en regresión, el cual se conoce como PLSR (Partial Least Squares Regression) [Wold et al., 2001]. Éste será el eje central en esta tesis, por lo tanto cada vez que mencionemos PLS nos vamos a referir a PLSR. En este capítulo haremos una breve introducción histórica a este método y sus aplicaciones más recientes (para una referencia más completa ver [Vinzi et al., 2010], [Zotto, 2012]).

PLS es uno de los primeros métodos para predicción en regresión lineal de alta dimensión en el caso en que el número de observaciones n no necesariamente es mayor que el número de predictores p . Su surgimiento fue un largo proceso, en el cual diferentes aspectos de este método fueron en realidad extraídos de otros (en gran parte debido a la insuficiencia de estos) y fueron refinándose con el paso del tiempo. El trabajo pionero en PLS se le acredita mayormente al Prof. Herman Wold, que lo estableció como un algoritmo de predicción en quimiometría [Wold, 1966a], [Geladi, 1988], [Wold, 2001]. Se destaca en esta área pues frente a predictores con alta colinealidad ha mostrado tener una mejor performance con respecto a métodos como mínimos cuadrados ordinarios y otras técnicas similares conocidas en el mundo de la alta dimensión. El algoritmo NIPALS de Wold [Wold, 1966b] es reconocido como el punto de partida en la evolución de PLS. Este método tiene mucho potencial en diversas áreas de la ciencia [Berntsson and Wold, 1986], [Vong et al., 1988], [Saaksjarvi et al., 1989], [Gottfries et al., 1995], [Schwartz et al., 2009].

En los años 1930 la tesis doctoral de H. Wold [Wold, 1938] en el campo de las series temporales, en la Universidad de Estocolmo lo llevó a un teorema sobre la separación de las contribuciones estocásticas y determinísticas en una serie de tiempo. Luego de recibir su doctorado, Wold se aventuró en el estudio de la regresión en modelos econométricos. Esto lo llevó a desarrollar el método de punto fijo en los comienzos de 1960, que se trata de un método para diseñar modelos de caminos (*path models*) que obtienen buenos resultados numéricos con pocas iteraciones. Esta experiencia con modelos iterativos jugó un rol importante ya que motivaron a Wold para sus futuros desarrollos en PLS.

Un punto de inflexión en la historia ocurre cuando H. Wold desarrolla el método NIPALS (*Non-linear Iterative Partial Least Squares*) mencionado anteriormente. Algunos de sus usos fueron una reinención del mismo método propuesto por Fisher en [Fisher, 1922]. Wold mostró [Noonan and Wold, 1977] que NIPALS resolvía los siguientes problemas

- componentes principales
- correlación canónica
- modelos híbridos de componentes principales y regresión múltiple
- componentes principales en el caso de información parcial
- sistemas interdependientes
- estimación en el modelo y en la estructura de residuos
- regresión cociente
- análisis de factor.

El método NIPALS contiene muchas propiedades interesantes que facilitaron el camino para el uso de métodos PLS. Estos toman una posición intermedia entre el análisis de datos y los métodos tradicionales basados en la suposición de que las observaciones siguen una distribución de probabilidad conjunta. El enfoque PLS era libre de distribución y se comenzó utilizando como un algoritmo iterativo.

Se requirieron varios años de experimentación hasta que Wold dio una forma más definitiva de PLS a fines de 1977 [Noonan and Wold, 1977], considerado como el nacimiento de PLS tal como lo conocemos hoy. Una de las propiedades principales de PLS que atrajo a la comunidad científica fue el hecho de que no había una restricción en el cociente entre el número de variables y el número de objetos estudiados. Esto abrió un camino al análisis de datos que permitió incluir muchos conjuntos de datos que previamente no habían sido considerados debido a que contenían muchas variables y pocas observaciones. Además, un aspecto importante del método PLS es que

evita caer en la trampa de la colinealidad que siempre molestó a gran parte de la comunidad científica [Jagpal and Brick, 1982], [S. Wold and et al., 1984].

La relativa escasez de estudios sobre la regresión PLS en las principales revistas de estadística puede deberse, parcialmente, a un principio contemporáneo que gobierna la regresión en alta dimensión. En las últimas dos décadas la comunidad estadística ha adoptado la idea conocida como *sparsity* como una caracterización natural de la regresión en alta dimensión, posiblemente esto fue lo que inhibió la consideración de otros paradigmas igualmente plausibles. Cuando hablamos de *sparsity* nos referimos a un modelo donde, independientemente de la dimensión en la que estemos trabajando, sólo un pequeño número de variables son las que importan. En efecto, muchos han visto la idea de *sparsity* como algo que obedece a una ley natural: si estás enfrentando un problema de regresión en alta dimensión entonces naturalmente éste debe ser un modelo *sparse*. De hecho, algunos ven esta opción como la única posible. En contraste con esto, la regresión estándar con PLS funciona mejor en casos de modelos abundantes (*anti-sparse*) donde muchas predictoras contribuyen información sobre la variable respuesta. Citando dos fragmentos de [Wold et al., 1996] (traducción propia)

En situaciones con muchas variables, digamos más de 50 o 100, hay una fuerte tentación de reducir drásticamente el número de variables en el modelo. Esta tentación está reforzada por la “tradición” que existe en regresión sobre reducir las variables tanto como sea posible para obtener la matriz X bien condicionada. Como se discute a continuación, sin embargo, esta reducción de variables usualmente elimina información, genera interpretaciones erróneas e incrementa el riesgo a obtener modelos espurios. Una mejor alternativa a la reducción de variables es dividir las variables en bloques conceptualmente significativos y luego aplicar modelos jerárquicos multibloque PLS (o PC). Estas ideas fueron presentadas por Wold, Martens y otros autores alrededor de 1986, pero en papers bastante oscuros.

[...] Con modelos de proyección multivariable como PLS y PCA, sin embargo, la situación es diferente. Estos métodos funcionan bien incluso con muchas variables aún si el número de observaciones N es pequeño. De hecho, mientras mayor es el número de variables relevantes, más precisos son los scores t (y u en PLS), porque tienen las características de promedios ponderados de todas las variables \mathbf{X} o Y y un promedio es más preciso mientras mayor es el número de elementos que forman la base de dicho promedio. Por lo tanto no hay una

verdadera necesidad de quedarse con un número pequeño de variables; sólo las variables verdaderamente irrelevantes deberían ser eliminadas para estabilizar el modelo y sus predicciones.

Desde luego, esto genera una dicotomía que distingue a PLS de métodos de regresión *sparse*. Si muchos predictores aportan información relevante de la respuesta en una configuración de alta dimensión entonces los métodos *sparse* generalmente fallarán, debido a que la regresión no es *sparse*, sin embargo métodos que tienen en cuenta que muchas predictores aportan, llamados abundantes como regresión PLS, pueden ser eficientes.

2.1. Algoritmo NIPALS

Como ya fue mencionado, PLS nació como un algoritmo iterativo planteado por H. Wold bajo el nombre de NIPALS el cual presentamos a continuación. Además veremos que tipo de conclusiones podemos extraer. El algoritmo NIPALS requiere seleccionar previamente el número de iteraciones que van a realizarse, que nosotros llamaremos q . Se suele referir a la cantidad q como *el número de componentes* y se determina haciendo uso de la información de los datos. Generalmente esta cantidad se calcula con un método de selección conocido como *cross-validation*.

Sean $\mathbb{X} \in \mathbb{R}^{n \times p}$ la matriz de predictores centrada y sea $\mathbb{Y} \in \mathbb{R}^{n \times 1}$ el vector de respuestas centrado. El algoritmo NIPALS se construye de la forma dada en el Algoritmo 1, donde hemos usado la notación $\ell_1(\cdot)$ que denota la función que devuelve el primer autovector, no necesariamente normalizado. Como en esta tesis nos concentramos en el caso de una componente, veamos la expresión del estimador al realizar un paso del algoritmo NIPALS. Primero tenemos que

$$\begin{aligned}\hat{\sigma}_{X_1, Y_1} &= \hat{\sigma} \\ w_1 &= \hat{\sigma} \\ s_1 &= \mathbb{X} \hat{\sigma} \\ l_1 &= \frac{\mathbb{X}^T \mathbb{X} \hat{\sigma}}{\hat{\sigma}^T \mathbb{X}^T \mathbb{X} \hat{\sigma}} \\ m_1 &= \frac{\mathbb{Y}^T \mathbb{X} \hat{\sigma}}{\hat{\sigma}^T \mathbb{X}^T \mathbb{X} \hat{\sigma}}\end{aligned}$$

y observando que $\mathbb{X}^T \mathbb{Y} = (n-1) \hat{\sigma}$ y $\mathbb{X}^T \mathbb{X} = (n-1) \hat{\Sigma}$ obtenemos

Algoritmo 1 NIPALS

Inicializar: $\mathbb{X}_1 = \mathbb{X}, \mathbb{Y}_1 = \mathbb{Y}$

Elegir: $q \leq \min \{ \text{rank}(\widehat{\Sigma}), n - 1 \}$

para $d = 1, \dots, q$:

$$\widehat{\sigma}_{X_d, Y_d} \leftarrow (n - 1)^{-1} \mathbb{X}_d^T \mathbb{Y}_d$$

$$w_d \leftarrow \ell_1 \left(\widehat{\sigma}_{X_d, Y_d} \widehat{\sigma}_{X_d, Y_d}^T \right)$$

$$s_d \leftarrow \mathbb{X}_d w_d$$

$$l_d \leftarrow \frac{\mathbb{X}_d^T s_d}{s_d^T s_d}$$

$$m_d \leftarrow \frac{\mathbb{Y}_d^T s_d}{s_d^T s_d}$$

$$\mathbb{X}_{d+1} \leftarrow \mathbb{X}_d - s_d l_d^T$$

$$\mathbb{Y}_{d+1} \leftarrow \mathbb{Y}_d - s_d m_d^T$$

$$\mathbf{W}_q \in \mathbb{R}^{p \times q} \leftarrow \{w_1, \dots, w_q\}$$

$$\mathbf{S}_q \in \mathbb{R}^{n \times q} \leftarrow \{s_1, \dots, s_q\}$$

$$\mathbf{L}_q \in \mathbb{R}^{p \times q} \leftarrow \{l_1, \dots, l_q\}$$

$$\mathbf{M}_q \in \mathbb{R}^{r \times q} \leftarrow \{m_1, \dots, m_q\}$$

$$\widehat{\beta}_{npls} = \mathbf{W}_q (\mathbf{L}_q^T \mathbf{W}_q)^{-1} \mathbf{M}_q^T$$

$$\begin{aligned} \widehat{\beta}_{npls} &= \mathbf{W}_1 (\mathbf{L}_1^T \mathbf{W}_1)^{-1} \mathbf{M}_1^T \\ &= w_1 (l_1^T w_1)^{-1} m_1^T \\ &= \widehat{\sigma} \left(\frac{\widehat{\sigma}^T \mathbb{X}^T \mathbb{X} \widehat{\sigma}}{\widehat{\sigma}^T \mathbb{X}^T \mathbb{X} \widehat{\sigma}} \right)^{-1} \frac{\widehat{\sigma}^T \mathbb{X}^T \mathbb{Y}}{\widehat{\sigma}^T \mathbb{X}^T \mathbb{X} \widehat{\sigma}} \\ &= \widehat{\sigma} (\widehat{\sigma}^T \widehat{\Sigma} \widehat{\sigma})^{-1} \widehat{\sigma}^T \widehat{\sigma}. \end{aligned} \tag{2.1}$$

Notar que esta forma es independiente del autovector elegido de $\text{span}(\widehat{\sigma})$. Si bien muchas versiones del algoritmo PLS fueron apareciendo en diversos trabajos, en [de Jong, 1993] se plantea una variante conocida como *Statistically Inspired Modification of the Partial Least Squares* (SIMPLS) que, según el autor, ofrece varias ventajas sobre el método tradicional de NIPALS. Al igual que éste, SIMPLS opera sobre la estructura de las matrices de datos y obtiene un estimador de β . Generalmente los estimadores de β que generan los dos algoritmos son diferentes, sin embargo se relacionan en el siguiente resultado presentado en [de Jong, 1993]

Proposición 2.1. Sean $\widehat{\beta}_{spls}$ y $\widehat{\beta}_{npls}$ los estimadores generados por SIMPLS y NIPALS respectivamente, contruidos con q componentes

1. Si se usa una componente, es decir $q = 1$, entonces $\widehat{\beta}_{spls} = \widehat{\beta}_{npls}$.
2. Si la respuesta es univariada, es decir $r = 1$, entonces nuevamente $\widehat{\beta}_{spls} = \widehat{\beta}_{npls}$.

Hemos introducido el algoritmo más básico de PLS en un contexto de regresión para llegar a un estimador de β , pero no hay que olvidar que inicialmente esto fue propuesto como un método de predicción, sin estimar parámetro alguno y sin un modelo matemático de fondo. En el trabajo [Helland, 1988] el autor notó que bajo el modelo de regresión lineal, el algoritmo NIPALS de PLS con q componentes genera un estimador para β con una fórmula cerrada, la cual satisface

$$\widehat{\beta}_{npls} = \widehat{W}_q \left(\widehat{W}_q^T \widehat{\Sigma} \widehat{W}_q \right)^{-1} \widehat{W}_q^T \widehat{\sigma}, \quad (2.2)$$

donde además la matriz \widehat{W}_q es base del subespacio de Krylov $(\widehat{\sigma}, \widehat{\Sigma}\widehat{\sigma}, \widehat{\Sigma}^2\widehat{\sigma}, \dots, \widehat{\Sigma}^{q-1}\widehat{\sigma})$. Una vez que esta expresión fue deducida, el siguiente paso de Helland [Helland, 1990] fue proponer la correspondiente fórmula poblacional reemplazando los estimadores. Podemos decir entonces que tenemos un modelo PLS para regresión lineal, dado por

$$Y = \mu_Y + \beta^T (\mathbf{X} - \mu_{\mathbf{X}}) + \varepsilon \quad (2.3)$$

$$\beta = \mathbf{W}_q (\mathbf{W}_q^T \Sigma \mathbf{W}_q)^{-1} \mathbf{W}_q^T \sigma, \quad (2.4)$$

donde los vectores de \mathbf{W}_q son una base ortogonal del subespacio de Krylov $(\sigma, \Sigma\sigma, \Sigma^2\sigma, \dots, \Sigma^{q-1}\sigma)$, y los vectores w_1, \dots, w_q se obtienen como un proceso de ortogonalización de Gram-Schmidt del espacio de la sucesión de Krylov $\sigma, \Sigma\sigma, \dots, \Sigma^{q-1}\sigma$. Dicho esto, vemos que $\mathbf{W}_1 = \sigma$ y por (2.2) obtenemos la forma de β para una componente

$$\beta = \sigma (\sigma^T \Sigma \sigma)^{-1} \sigma^T \sigma,$$

y su correspondiente versión muestral

$$\widehat{\beta}_{pls} = \widehat{\sigma} (\widehat{\sigma}^T \widehat{\Sigma} \widehat{\sigma})^{-1} \widehat{\sigma}^T \widehat{\sigma}. \quad (2.5)$$

En esta tesis no nos detenemos mucho a analizar algoritmos pues no es el foco de nuestra investigación. Sin embargo es importante su mención pues da el paso inicial a lo que sigue y permite complementar la historia desarrollada. Lo importante es rescatar la forma del estimador generado por el método PLS para $q = 1$ dado en (2.5).

2.2. Envelopes y su relación con PLS

En [Cook et al., 2007] se introduce la noción de *envelopes*, luego desarrollado para modelos lineales multivariados en [Cook et al., 2010] y más generalmente en [Cook, 2018], un método de reducción de dimensiones para el vector de predictores \mathbf{X} (en un contexto más general, también sirve como método de reducción para las variables respuestas). El objetivo de *envelope regression* es separar claramente la información relevante para el objetivo del estudio de aquella parte que es irrelevante, lo cual recae un poco en la noción de [Fisher, 1922] sobre lo que entendemos por estadísticos suficientes. En esta sección trataremos de introducir brevemente de que se trata esta metodología y luego su relación con PLS.

Definición 2.2. Sea \mathcal{R} un subespacio de \mathbb{R}^r y $\mathbf{M} \in \mathbb{R}^{r \times r}$ simétrica. Decimos que \mathcal{R} *reduce* a \mathbf{M} si $\mathbf{M}\mathcal{R} \subseteq \mathcal{R}$ y $\mathbf{M}\mathcal{R}^\perp \subseteq \mathcal{R}^\perp$. Es decir, \mathbf{M} mapea a \mathcal{R} y \mathcal{R}^\perp en subespacios de ellos mismos.

Así mismo, también se puede establecer una equivalencia que será muy importante

Proposición 2.3. [Cook et al., 2007, Proposición 2.1] \mathcal{R} *reduce* a $\mathbf{M} \in \mathbb{R}^{r \times r}$ simétrica si y sólo si \mathbf{M} puede escribirse de la siguiente forma

$$\mathbf{M} = \mathbf{P}_{\mathcal{R}}\mathbf{M}\mathbf{P}_{\mathcal{R}} + \mathbf{Q}_{\mathcal{R}}\mathbf{M}\mathbf{Q}_{\mathcal{R}} \quad (2.6)$$

donde $\mathbf{P}_{\mathcal{R}}$ es la proyección sobre $\text{span}(\mathcal{R})$ y $\mathbf{Q}_{\mathcal{R}}$ es el complemento ortogonal.

Ahora estamos en condiciones de definir formalmente un envelope en su forma más algebraica.

Definición 2.4. Sea $\mathbf{M} \in \mathbb{R}^{r \times r}$ simétrica y $\mathcal{S} \subseteq \text{span}(\mathbf{M})$. Entonces el \mathbf{M} -envelope de \mathcal{S} se denota como $\mathcal{E}_{\mathbf{M}}(\mathcal{S})$, y es la intersección de todos los subespacios de \mathbb{R}^r que reducen a \mathbf{M} y que contienen a \mathcal{S} .

Esta definición considera sólo los subespacios de reducción que contienen a un subespacio en particular, \mathcal{S} , que usualmente está generado por un vector de parámetros.

Como ya dijimos, el objetivo del método envelope es separar la información relevante que puede haber sobre un vector de respuesta Y , dejando de lado una proporción de información irrelevante para una correcta predicción. Sin embargo, en este desarrollo no hemos hecho mención sobre ningún modelo matemático subyacente que determine la relación entre estas variables. Podemos asumir por un momento que sólo disponemos de un vector de respuestas $Y \in \mathbb{R}$ y un vector de predictores $\mathbf{X} \in \mathbb{R}^p$, de los cuales contamos con una muestra aleatoria (\mathbf{X}_i, Y_i) ,

$i = 1, 2, \dots, n$. En varias oportunidades mencionamos que los enfoques clásicos estadísticos pueden no funcionar cuando $p \gg n$ o cuando hay mucha colinealidad entre predictores. Una solución posible es reducir la dimensión del predictor \mathbf{X} sin perder información importante sobre Y . Denotemos entonces con $\mathbf{P}_S \mathbf{X}$ a la reducción de estos predictores, donde $S \subset \mathbb{R}^p$. Claramente $\mathbf{P}_S \mathbf{X}$ debe portar toda o al menos la gran mayoría de la información que \mathbf{X} tiene sobre Y . Esto se interpreta diciendo que dado $\mathbf{P}_S \mathbf{X}$, Y es independiente de $\mathbf{Q}_S \mathbf{X}$. Podemos establecer esta condición como

$$Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{P}_S \mathbf{X}. \quad (2.7)$$

Esta condición es el motor central detrás de la reducción suficiente de dimensiones (SDR), cuyo objetivo es reducir la dimensión de los predictores sin requerir un modelo para la regresión de Y en función de \mathbf{X} . La intersección de todos los subespacios para los cuales vale (2.7) se llama *subespacio central* y se denota por $\mathcal{S}_{Y|\mathbf{X}}$. Por ejemplo consideremos el modelo de regresión lineal: $Y = \mu_Y + \boldsymbol{\beta}^T (\mathbf{X} - \boldsymbol{\mu}_X) + \varepsilon$ con $\varepsilon \perp\!\!\!\perp \mathbf{X}$. En este caso la respuesta sólo depende del vector de predictores a través de $\boldsymbol{\beta}^T \mathbf{X}$, y por lo tanto $Y \perp\!\!\!\perp \mathbf{X} \mid \boldsymbol{\beta}^T \mathbf{X}$ y $\mathcal{S}_{Y|\mathbf{X}} = \text{span}(\boldsymbol{\beta})$. En la aplicación, problemas como la alta dimensionalidad de los datos o la colinealidad pueden dificultar distinguir la parte de $\mathbf{P}_S \mathbf{X}$ de \mathbf{X} que es relevante para predecir Y , ya que puede existir una porción irrelevante en la parte complementaria $\mathbf{Q}_S \mathbf{X}$. Para compensar esta posible complicación se requiere además que aquella parte que seleccionamos como relevante sea independiente de la parte irrelevante, es decir, $\mathbf{P}_S \mathbf{X} \perp\!\!\!\perp \mathbf{Q}_S \mathbf{X}$. Bajo la suposición de normalidad de los predictores, esta condición es equivalente a $\text{cov}(\mathbf{P}_S \mathbf{X}, \mathbf{Q}_S \mathbf{X}) = \mathbf{P}_S \boldsymbol{\Sigma} \mathbf{Q}_S = 0$ (cuando no podemos asumir normalidad, la condición de independencia suele reemplazarse por la de covarianza 0). Esto a su vez es equivalente a

$$\boldsymbol{\Sigma} = \mathbf{P}_S \boldsymbol{\Sigma} \mathbf{P}_S + \mathbf{Q}_S \boldsymbol{\Sigma} \mathbf{Q}_S, \quad (2.8)$$

y podemos observar que esta condición es en efecto la definición de subespacio reductor que vimos en (2.6). Esta nueva condición junto con (2.7) permite establecer una buena definición de envelopes para contextos de predicción

Definición 2.5. El *envelope de predicción* para la regresión de Y en función de \mathbf{X} es la intersección de todos los subespacios reductores de $\boldsymbol{\Sigma}$ que contienen a $\mathcal{S}_{Y|\mathbf{X}}$, y se denota por $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S}_{Y|\mathbf{X}})$.

Antes de continuar, recordemos que envelope es un método de reducción de dimensiones, es decir, un método donde se busca seleccionar una pequeña parte de las p predictoras \mathbf{X} con la esperanza de retener la información que estas nos aportan a Y . El objetivo de estos métodos es

encontrar, si existe, una matriz $\Phi \in \mathbb{R}^{p \times q}$ donde $q < p$ (y en la muestra se espera que $q < n$) que permita reducir la dimensión bajo la transformación $\mathbf{X} \mapsto \Phi^T \mathbf{X} \in \mathbb{R}^q$ y luego estimar un nuevo vector de parámetros mediante un método como mínimos cuadrados. Generalmente Φ es desconocida y se estima a partir de los datos, pero supongamos por un momento que es conocida y no estocástica. Si partimos de nuestro modelo lineal

$$Y = \mu_Y + \beta^T (\mathbf{X} - \mu_{\mathbf{X}}) + \varepsilon, \quad (2.9)$$

considerando la reducción de \mathbf{X} dada por Φ obtenemos un nuevo modelo

$$Y = \mu_Y + \eta^T \Phi^T (\mathbf{X} - \mu_{\mathbf{X}}) + \varepsilon, \quad (2.10)$$

donde tenemos un vector de parámetros $\eta \in \mathbb{R}^q$ para las nuevas variables predictoras reducidas $\Phi^T \mathbf{X}$. Notar que a partir de (2.10) se tiene la forma para β del modelo (2.9) dada por $\beta = \Phi \eta$. Bajo el modelo (2.10) y como $\Sigma > 0$ el vector η será

$$\begin{aligned} \eta &= (\text{var}(\Phi^T \mathbf{X}))^{-1} \text{cov}(\Phi^T \mathbf{X}, Y) \\ &= (\Phi^T \Sigma \Phi)^{-1} \Phi^T \sigma, \end{aligned}$$

y como $n > q$ se puede hallar el correspondiente estimador de mínimos cuadrados $\hat{\eta}_{ols} = (\hat{\Phi}^T \hat{\Sigma} \hat{\Phi})^{-1} \hat{\Phi}^T \hat{\sigma}$ si tenemos un estimador conocido para Φ . De esta forma podemos hallar el estimador plugin

$$\hat{\beta} = \hat{\Phi} \hat{\eta}_{ols} = \hat{\Phi} (\hat{\Phi}^T \hat{\Sigma} \hat{\Phi})^{-1} \hat{\Phi}^T \hat{\sigma}. \quad (2.11)$$

En primer lugar, notamos que este estimador de β no requiere el cálculo directo de $\hat{\Sigma}^{-1}$ sino que la inversa que se necesita es la de $(\hat{\Phi}^T \hat{\Sigma} \hat{\Phi}) \in \mathbb{R}^{q \times q}$. Segundo, una propiedad del estimador (2.11) es que es invariante bajo transformaciones no singulares a derecha de $\hat{\Phi}$, lo cual implica que $\hat{\beta}$ depende de $\hat{\Phi}$ sólo a través del subespacio $\text{span}(\hat{\Phi})$.

Sea $\mathcal{B} = \text{span}(\beta)$ el subespacio de \mathbb{R}^p generado por β . El modelo lineal muestra que Y depende de \mathbf{X} sólo a través de $\beta^T \mathbf{X}$, por lo tanto bajo la Definición 2.5 se sigue que como $\mathcal{S}_{Y|\mathbf{X}} = \mathcal{B}$ el envelope es $\mathcal{E}_{\Sigma}(\mathcal{S}_{Y|\mathbf{X}}) = \mathcal{E}_{\Sigma}(\mathcal{B})$. De esta forma, tomando $\Phi \in \mathbb{R}^{p \times q}$ como una base semi-ortogonal del envelope $\mathcal{E}_{\Sigma}(\mathcal{B})$ (donde q es la dimensión del envelope) tenemos una reducción de dimensiones y considerando Φ_0 tal que $(\Phi, \Phi_0) \in \mathbb{R}^{p \times p}$ sea una base ortogonal se tendrá la descomposición de Σ dada por (2.8) de la siguiente forma

$$\Sigma = \Phi \Phi^T \Sigma \Phi \Phi^T + \Phi_0 \Phi_0^T \Sigma \Phi_0 \Phi_0^T.$$

De esta forma, la idea de reducción de dimensiones en un contexto de regresión dada por $\beta = \Phi\eta$ combinado con la teoría de envelope permiten definir el modelo envelope de regresión lineal dado por

$$Y = \mu_Y + \eta^T \Phi^T (X - \mu_X) + \varepsilon$$

$$\Sigma = \Phi \Upsilon \Phi^T + \Phi_0 \Upsilon_0 \Phi_0^T,$$

donde $\Upsilon = \Phi^T \Sigma \Phi$ y $\Upsilon_0 = \Phi_0^T \Sigma \Phi_0$.

La relación entre envelope y PLS ha sido descrita en varias oportunidades; [Cook et al., 2013], [Cook, 2018], [Cook and Forzani, 2018]. Mostramos ahora que el modelo PLS poblacional es en realidad un modelo envelope.

Teorema 2.6. *Bajo el modelo de regresión lineal si β satisface (2.4) con W_q una base ortogonal del subespacio de Krylov $(\sigma, \Sigma\sigma, \Sigma^2\sigma, \dots, \Sigma^{q-1}\sigma)$, entonces tomando $\Phi = W_q \in \mathbb{R}^{p \times q}$, $\Phi_0 \in \mathbb{R}^{p \times (p-q)}$ como antes y $\eta = (W_q^T \Sigma W_q)^{-1} W_q^T \sigma$ se tiene que $\beta = \Phi\eta$ y*

$$\Sigma = \Phi \Phi^T \Sigma \Phi \Phi^T + \Phi_0 \Phi_0^T \Sigma \Phi_0 \Phi_0^T. \quad (2.12)$$

Adicionalmente, podemos notar que $\Phi^T \Sigma \Phi = \text{var}(\Phi^T X) \in \mathbb{R}^{q \times q}$ y $\Phi_0^T \Sigma \Phi_0 = \text{var}(\Phi_0^T X) \in \mathbb{R}^{(p-q) \times (p-q)}$.

Veremos el caso $q = 1$ que será el que desarrollamos en el resto de la tesis. En este caso, W_1 es una base para el espacio generado por σ y por lo tanto podemos considerar $W_1 = \ell = \sigma / \sqrt{\sigma^T \sigma}$. Queremos ver que se cumple (2.12). Sea $\ell_0 \in \mathbb{R}^{p \times (p-1)}$ tal que $(\ell, \ell_0) \in \mathbb{R}^{p \times p}$ es una base ortogonal. Definiendo $P_\sigma = \ell \ell^T$, $Q_\sigma = I - P_\sigma = \ell_0 \ell_0^T$ y $\Omega = \ell^T \Sigma \ell$ de (2.4) se tiene que

$$\beta = \ell \Omega^{-1} \ell^T \sigma = \Omega^{-1} P_\sigma \sigma = \Omega^{-1} \sigma.$$

Más aún, como $\beta = \Sigma^{-1} \sigma$ notamos que $\Sigma \sigma = \Omega \sigma$, i.e., Ω es un autovalor de Σ . De aquí podemos concluir que

$$\begin{aligned} \Sigma &= (P_\sigma + Q_\sigma) \Sigma (P_\sigma + Q_\sigma) \\ &= P_\sigma \Sigma P_\sigma + P_\sigma \Sigma Q_\sigma + Q_\sigma \Sigma P_\sigma + Q_\sigma \Sigma Q_\sigma \\ &= P_\sigma \Sigma P_\sigma + Q_\sigma \Sigma Q_\sigma \\ &= \ell \ell^T \Sigma \ell \ell^T + \ell_0 \ell_0^T \Sigma \ell_0 \ell_0^T \end{aligned}$$

ya que $P_\sigma \Sigma Q_\sigma = \ell \ell^T \Sigma \ell_0 \ell_0^T = \Omega \ell \ell^T \ell_0 \ell_0^T = 0$ y lo mismo para $Q_\sigma \Sigma P_\sigma$.

En términos de lo que mencionamos antes, esto significa que podemos pensar en $\ell^T \mathbf{X}$ como la parte relevante de \mathbf{X} (la que contiene información sobre Y) mientras que $\ell_0^T \mathbf{X}$ es la parte irrelevante. Por la observación del Teorema 2.6

$$\text{var}(\ell^T \mathbf{X}) = \ell^T \Sigma \ell = \frac{\sigma^T \Sigma \sigma}{\sigma^T \sigma} = \Omega > 0 \quad (2.13)$$

$$\text{var}(\ell_0^T \mathbf{X}) = \ell_0^T \Sigma \ell_0 = \Omega_0 > 0, \quad (2.14)$$

donde la última desigualdad significa que $\Omega_0 \in \mathbb{R}^{(p-1) \times (p-1)}$ es una matriz definida positiva. Por lo tanto tenemos el modelo lineal en una dimensión para envelope, que a su vez corresponde con el modelo de regresión PLS con una componente

$$\begin{aligned} Y_i &= \mu_Y + \beta^T (\mathbf{X}_i - \mu_{\mathbf{X}}) + \varepsilon_i \\ &= \mu_Y + \left(\Omega^{-1} \sqrt{\sigma^T \sigma} \right) \ell^T (\mathbf{X}_i - \mu_{\mathbf{X}}) + \varepsilon_i, \quad i = 1, \dots, n \end{aligned} \quad (2.15)$$

$$\Sigma = \ell \Omega \ell^T + \ell_0 \Omega_0 \ell_0^T. \quad (2.16)$$

Se sigue de (2.9) y $\beta = \Sigma^{-1} \sigma = \Omega^{-1} \sigma$ que

$$\text{var}(Y) = \sigma_Y^2 = \beta^T \Sigma \beta + \tau^2 = \frac{\sigma^T \sigma}{\Omega} + \tau^2. \quad (2.17)$$

2.3. Resultados para p fijo

Como ya fue mencionado, en [Cook et al., 2013] se obtienen los primeros resultados sobre la distribución asintótica del estimador PLS. Básicamente la conexión se refleja en el hecho de que un algoritmo como NIPALS o SIMPLS puede usarse para estimar una base de $\mathcal{E}_{\Sigma}(\mathcal{B})$. Dicha base juega el rol de matriz reductora tal como se mencionó al principio de la Sección 2.1.

Uno de los resultados principales fue establecido para $r = q = 1$ ya que cualquier caso en donde $r > 1$ o $q > 1$ implican una complejidad superior y que aún son objeto de investigación. En [Cook et al., 2013] se probó, en primer lugar, que si $\widehat{\beta}_{pls}$ es el estimador PLS del coeficiente β dado en (2.5) entonces si $n > p$ con p fijo, se tiene una expansión

$$\sqrt{n}(\widehat{\beta}_{pls} - \beta) = \sqrt{n}\Gamma(\mathbf{X}_i, Y_i, \beta) + O_p(n^{-1/2}), \quad (2.18)$$

donde $\Gamma(\mathbf{X}_i, Y_i, \beta)$ es una expresión que depende de la muestra y del vector β . Dicha expresión ahora no es relevante pero volveremos a ella más adelante cuando unifiquemos conceptos (ver Sección 3.3). La clave fundamental, que discutiremos ampliamente en las demostraciones del siguiente capítulo, es que la expansión anterior permite deducir el siguiente resultado

Proposición 2.7. [Cook et al., 2013, Proposición 10] Para $n \rightarrow \infty$ y p fijo se tiene que $\sqrt{n}(\hat{\beta}_{pls} - \beta)$ es asintóticamente normal con media 0 y varianza asintótica

$$avar(\sqrt{n}\hat{\beta}_{pls}) = \Omega^{-2} (\Sigma\tau^2 + \beta^T \Sigma \beta \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma) \quad (2.19)$$

que depende de los momentos de (X, Y) y del modelo (2.9).

En [Naik and Tsai, 2000] se probó que bajo un modelo un poco más general que (2.9) y bajo ciertas condiciones sobre el espectro de Σ , el estimador $\hat{\beta}_{pls}$ es un estimador consistente de β con una constante de proporcionalidad. Es decir $\hat{\beta}_{pls} \xrightarrow{p} C\beta$ cuando $n \rightarrow \infty$ y p es fijo. Por otro lado, en [Chun and Keles, 2010] se probó que el estimador PLS de β era consistente cuando $p/n \rightarrow k = 0$, pero era inconsistente si $k > 0$ bajo condiciones de *sparsity*. Por esta razón se limitó la discusión al estudio de propiedades del estimador muestral para el caso $n > p$ con p fijo y $n \rightarrow \infty$, mientras que el caso $p > n$ quedaba abierto.

Volveremos a la expresión (2.19) más adelante y veremos que juega un papel importante en los resultados a desarrollar. Una observación de este resultado es que permite comparar performance de PLS versus OLS, que depende de medidas como el cociente $\tau^2/\sigma_Y^2 \leq 1$ y del nivel de colinealidad. Se puede ver que existen configuraciones donde PLS tiene una performance peor que OLS a nivel asintótico, así como también existen niveles de colinealidad para los cuales PLS es mejor que OLS asintóticamente. Por ejemplo, puede verse que si $\text{var}(\mathbf{X}) = \Sigma = \sigma^2 \mathbf{I}$ entonces la varianza asintótica del estimador PLS nunca es menor que el del estimador OLS.

RESULTADOS ASINTÓTICOS

3.1. Predicción PLS

Como ya fue mencionado, en [Cook et al., 2013] se obtuvieron resultados sobre el comportamiento del estimador PLS para el caso en que $n \rightarrow \infty$ y p es fijo. En [Cook and Forzani, 2018] se dio un primer paso al estudiar comportamientos asintóticos en alta dimensión de la predicción PLS para $q = 1$ en regresión univariada, con la esperanza de asentar las bases para una posible extensión a casos más generales ($q > 1, r > 1$). Hablamos de predicción porque ahora nos encontramos en una configuración donde $n, p \rightarrow \infty$. Este es un nuevo caso donde estamos agregando predictoras a nuestra matriz de diseño. Podría suceder que n y p crezcan con distinto orden, permitiendo que $p \gg n$, en cuyo caso estamos añadiendo más información que observaciones. Esto implica que $\boldsymbol{\sigma}$ y $\boldsymbol{\Sigma}$ ya no son de dimensión fija (por cada nueva variable predictora, la matriz $\boldsymbol{\Sigma}$ agrega una columna y una fila), y por supuesto el vector de coeficientes $\boldsymbol{\beta}$ tampoco es de dimensión fija. Obsevamos que esto es un problema a la hora de establecer un resultado de convergencia asintótica para $n, p \rightarrow \infty$ pues deberíamos definir una distribución límite con matriz de covarianza de dimensión no fija. La clave en el trabajo [Cook and Forzani, 2018] para evitar este problema es estudiar el producto escalar entre el estimador de PLS y una nueva observación (\mathbf{X}_N, Y_N) independiente de la muestra, que sigue el modelo

$$Y_N = \mu_Y + \boldsymbol{\beta}^T(\mathbf{X}_N - \boldsymbol{\mu}_X) + \varepsilon_N$$

y compararla con el valor predicho $\hat{Y}_N = \bar{Y} + \hat{\boldsymbol{\beta}}_{pls}^T(\mathbf{X}_N - \bar{\mathbf{X}})$, lo cual nos da una expresión para la diferencia

$$\begin{aligned}\widehat{Y}_N - Y_N &= (\bar{Y} - \mu_Y) + (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\mathbf{X}_N - \boldsymbol{\mu}_X) - (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_X) \\ &\quad - \boldsymbol{\beta}^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_X) + \varepsilon_N.\end{aligned}$$

A continuación los autores prueban que el comportamiento n, p asintótico de la predicción está determinada por el orden de $(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\mathbf{X}_N - \boldsymbol{\mu}_X) \in \mathbb{R}$ y que bajo ciertas condiciones se obtiene la consistencia. Notemos que la ventaja de estudiar este término es que se trata de una cantidad escalar, la cual depende de p (pero no en la dimensión). Tomando esta idea como punto de partida, vamos a deducir algunas propiedades del estimador PLS e ir presentando otros resultados que se deducen de nuestro análisis.

En este capítulo vamos a extender el trabajo de [Cook and Forzani, 2018] al hallar y estudiar la distribución asintótica de $\widehat{\boldsymbol{\beta}}_{pls}^T \mathbf{G}$ en regresión lineal con una componente ($q = 1$), donde $\mathbf{G} \in \mathbb{R}^p$ es un vector no estocástico seleccionado por el usuario. Por ejemplo, seleccionando $\mathbf{G}^T = (1, 0, 0, \dots, 0)$ la distribución asintótica de $\widehat{\boldsymbol{\beta}}_{pls}^T \mathbf{G}$ podría usarse para inferir sobre la primera componente de $\boldsymbol{\beta}$. Tomando \mathbf{G} como un nuevo valor no observado de la matriz de \mathbf{X} , podríamos usar la distribución asintótica de $\widehat{\mu}_Y + \widehat{\boldsymbol{\beta}}_{pls}^T \mathbf{G}$ para estudiar valores promedio. A su vez notaremos que esto también es una extensión de la Proposición 2.7 del trabajo [Cook et al., 2013]. La restricción que hacemos a la hora de tomar una componente es una limitación notable. Sin embargo, existen muchos casos de regresiones que requieren sólo una componente así que puede esperarse que nuestros resultados tengan un valor práctico sustancial [Chiappini et al., 2021], [Cook, 2018, Sección 4.4.4], [Nadler and Coifman, 2005], [Berntsson and Wold, 1986], [Kettaneh-Wold, 1992] y este trabajo será el puntapié inicial para estudiar a futuro el caso $q > 1$.

3.2. Cantidades que gobiernan las distribuciones

En esta sección vamos a mencionar e interpretar algunas de las cantidades que controlan los resultados asintóticos de $\widehat{\boldsymbol{\beta}}_{pls}^T \mathbf{G}$ y de \widehat{Y}_N que presentaremos luego. Primero, recordemos el estimador del vector de coeficientes $\boldsymbol{\beta}$ y su correspondiente contraparte poblacional

$$\widehat{\boldsymbol{\beta}}_{pls} = (\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}})^{-1} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}} \boldsymbol{\sigma}, \quad \boldsymbol{\beta} = (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} \boldsymbol{\sigma}^T \boldsymbol{\sigma} \boldsymbol{\sigma}.$$

Además recordamos que el estimador de μ_Y es $\widehat{\mu}_Y = \bar{Y} - \widehat{\boldsymbol{\beta}}_{pls}^T \bar{\mathbf{X}}$ así que con predictores centrados, los valores ajustados pueden expresarse como $\widehat{Y}_i = \bar{Y} + \widehat{\boldsymbol{\beta}}_{pls}^T (\mathbf{X}_i - \bar{\mathbf{X}})$. El valor predicho de Y para un nuevo valor no observado \mathbf{X}_N de \mathbf{X} resulta $\widehat{Y}_N = \widehat{E}(Y|\mathbf{X}_N) = \bar{Y} + \widehat{\boldsymbol{\beta}}_{pls}^T (\mathbf{X}_N - \bar{\mathbf{X}})$.

Ahora continuamos con algunas de las definiciones más importantes para los resultados a presentar a continuación. Sean

$$C(p, k) = \text{tr} \left((\mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma)^k \right) = \text{tr}(\Omega_0^k) \quad (3.1)$$

$$K = \left(\frac{\Omega}{n\sigma^T \sigma} \right)^{1/2} + \frac{C(p, 1)}{n\sigma^T \sigma} \quad (3.2)$$

$$J = \left(\frac{\Omega}{n\sigma^T \sigma} \right)^{1/2} + \frac{C(p, 1)}{n\sigma^T \sigma} + \frac{C(p, 2)}{n(\sigma^T \sigma)^2}, \quad (3.3)$$

con Ω y Ω_0 definidos en (2.13) y (2.14). En lo que resta, vamos a suponer que $\Omega/(n\sigma^T \sigma)$ y $C(p, k)/n(\sigma^T \sigma)^k$ para $k = 1, 2, 3, 4$ están acotados cuando $n, p \rightarrow \infty$. Notemos que esto implica que K y J están acotados también. Nuestros resultados asintóticos requieren que $J \rightarrow 0$ cuando $n, p \rightarrow \infty$, lo cual vale sólo si los tres sumandos distintos de J convergen a 0. De manera similar, vamos a pedir que $C(p, 4)/n(\sigma^T \sigma)^4$ converja a 0 con n, p . Las expresiones de K, J fueron introducidas en [Cook and Forzani, 2018] aunque levemente distintas, sin embargo ambas formas son comparables (ver Apéndice A).

Consideremos el cociente $\Omega/\sigma^T \sigma$ que aparece como primer sumando de J . Por (2.17), $(\Omega/\sigma^T \sigma)^{-1} \leq \sigma_Y^2$ y, como σ_Y^2 se mantiene constante cuando $p \rightarrow \infty$ (ya que la respuesta es unidimensional y no depende de p), vemos que $\Omega/\sigma^T \sigma$ está acotado inferiormente por $1/\sigma_Y^2$. Si $(\Omega/\sigma^T \sigma)$ crece cuando $p \rightarrow \infty$ entonces correspondientemente τ^2 debe decrecer y es técnicamente posible que ocurra $\tau^2 \rightarrow 0$. Encontramos que esto es implausible en las aplicaciones y por eso vamos a suponer que τ^2 está acotado lejos de 0, lo cual implica que $(\Omega/\sigma^T \sigma)^{-1} < \sigma_Y^2$ y $\Omega/\sigma^T \sigma > 1/\sigma_Y^2$.

Si, cuando $n, p \rightarrow \infty$, $\sqrt{\Omega} = \text{var}^{1/2}(\ell^T \mathbf{X})$ crece mucho más rápido que $\sqrt{\sigma^T \sigma} = \text{cov}(Y, \ell^T X)$ entonces es posible que $\Omega/\sigma^T \sigma \rightarrow \infty$, en cuyo caso la correlación $\rho(Y, \ell^T X) \rightarrow 0$ cuando $p \rightarrow \infty$. En este caso para mantener una señal, el tamaño de muestra debe crecer más rápido que $\Omega/\sigma^T \sigma$ para obligar al sumando $(\Omega/\sigma^T \sigma n)^{1/2}$ de J a converger a 0 cuando $n, p \rightarrow \infty$. Esperamos que en casi todos los problemas de regresión será apropiado acotar $\Omega/\sigma^T \sigma$ inferiormente cuando $p \rightarrow \infty$, por lo que siempre habrá algo de señal presente en la regresión, y entonces podemos considerar $\Omega/\sigma^T \sigma = 1$ en J .

Los sumandos restantes de J pueden ser representados como una comparación del ruido en \mathbf{X} representado por Ω_0 y la señal representada por $\sigma^T \sigma$: $C(p, k)/n(\sigma^T \sigma)^k = n^{-1} \text{tr}\{(\Omega_0/\sigma^T \sigma)^k\}$. Si los autovalores del ruido $\Omega_0 = \text{var}(\ell_0^T \mathbf{X})$ están acotados cuando $p \rightarrow \infty$, lo cual podría ser una condición razonable en muchas aplicaciones, entonces cuando $n, p \rightarrow \infty$ se tiene $C(p, k)/n(\sigma^T \sigma)^k \asymp p/n(\sigma^T \sigma)^k$, el cual debe estar acotado cuando $n, p \rightarrow \infty$ para $k = 1, 2, 3, 4$.

En regresión sparse, $\sigma^T \sigma \asymp 1$ y entonces p/n debe estar acotado. En regresión abundante, $\sigma^T \sigma \asymp p$ y entonces $C(p, k)/n(\sigma^T \sigma)^k \asymp 1/np^{k-1} \rightarrow 0$ para $k = 1, 2, 3, 4$.

3.3. Distribución del estimador

En esta sección describimos la distribución asintótica de una versión escalada de $(\hat{\beta}_{pls} - \beta)^T \mathbf{G}$ cuando $n, p \rightarrow \infty$. Aunque estamos considerando el caso de alta dimensionalidad al hacer tender ambos parámetros a infinito, nuestros resultados pueden adaptarse directamente al caso $n \rightarrow \infty$ con p fijo. En este sentido nuestro enfoque también cubre los resultados de [Cook et al., 2013] para $q = 1$. Lo desarrollado a continuación requiere varias cantidades que vamos a definir, antes de presentar los resultados de nuestro estudio. Por conveniencia notacional, sea $m = n - 1$. Ahora, definamos

$$\begin{aligned} \mathbf{R} &= \Omega^{-1} \left(\hat{\sigma} - \sigma - \Omega^{-1} \mathbf{P}_\sigma (\hat{\Sigma} - \Sigma) \sigma \right) \\ A &= - \left(\frac{2}{m} + \frac{C(p, 1)}{m^2 \Omega} + \sigma_Y^2 \frac{C(p, 1)}{m^2 \sigma^T \sigma} + \sigma_Y^2 \frac{C(p, 2)}{m^2 \Omega \sigma^T \sigma} \right) \\ \Delta &= (\sigma^T \hat{\Sigma} \sigma)^{-1} (\hat{\sigma} - \sigma)^T (\hat{\sigma} + \sigma) (\hat{\sigma} - \sigma) + \frac{(\hat{\sigma} - \sigma)^T \sigma}{\sigma^T \Sigma \sigma} \left(\frac{\hat{\sigma}^T \sigma}{\sigma^T \sigma} - \frac{\hat{\sigma}^T \hat{\Sigma} \sigma}{\sigma^T \Sigma \sigma} \right) \sigma + A \beta \\ U &= (\hat{\sigma} - \sigma)^T \left[\frac{\mathbf{Q}_\sigma (\hat{\sigma} - \hat{\Sigma} \hat{\sigma} \Omega^{-1}) + \sigma - \hat{\Sigma} \sigma \Omega^{-1}}{\sigma^T \sigma} \right] + A. \end{aligned}$$

El término A lo definimos para facilitar las definiciones de Δ y U . Estas cantidades se derivan de una expansión de $(\hat{\beta}_{pls} - \beta)^T \mathbf{G}$ como se muestra en el siguiente resultado, cuya prueba se encuentra en el Apéndice D.

Proposición 3.1. *Supongamos que vale el modelo de regresión PLS para una componente dado por (2.15) y (2.16). Entonces*

$$(\hat{\beta}_{pls} - \beta)^T \mathbf{G} = \left(\frac{\hat{\sigma}^T \hat{\Sigma} \hat{\sigma}}{\sigma^T \Sigma \sigma} \right)^{-1} (\mathbf{R} + \Delta + \beta U)^T \mathbf{G}. \quad (3.4)$$

El comportamiento asintótico de estas cantidades en esta expansión de $(\hat{\beta}_{pls} - \beta)^T \mathbf{G}$ se describen en el siguiente resultado. En preparación, definamos

$$V(\mathbf{G}) = \frac{\Omega^{-2}}{n} \mathbf{G}^T (\Sigma \tau^2 + \beta^T \Sigma \beta \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma) \mathbf{G} \quad (3.5)$$

$$= \frac{\Omega^{-2}}{n} [(\mathbf{G}^T \ell)^2 \Omega \tau^2 + \sigma_Y^2 \mathbf{G}^T \ell_0 \Omega_0 \ell_0^T \mathbf{G}] \quad (3.6)$$

$$= \frac{\Omega^{-2}}{n} \mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}$$

$$b = - \left(\left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} - \tau^2 \right) \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \sigma_Y^2 \frac{C^2(p, 1)}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \sigma_Y^2 \frac{C(p, 2)}{m \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right), \quad (3.7)$$

donde $V(\mathbf{G})$ es usada como la varianza asintótica de $(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \mathbf{G}$ y b juega el papel de un sesgo. Cada una de las expresiones para $V(\mathbf{G})$ son útiles y su uso dependerá del contexto. A menos que sea necesario aclararlo, suprimiremos el argumento de la varianza y escribiremos simplemente V para hacer referencia a $V(\mathbf{G})$. En efecto, esta expresión de V coincide con el caso no escalar de p fijo dado en [Cook et al., 2013] tal como fue descrito en la Proposición 2.7.

Proposición 3.2. *Supongamos que valen las hipótesis de la Proposición 3.1, y consideremos las siguientes hipótesis adicionales:*

- (a) $\mathbf{X} \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma})$
- (b) $E(\varepsilon^4)$ del modelo (2.15) está acotada
- (c) $J \rightarrow 0$ cuando $n, p \rightarrow \infty$
- (d) $C(p, k)/n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^k$ está acotado cuando $n, p \rightarrow \infty$ para $k = 3, 4$.

Entonces

1. $\frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} = 1 + O_p(J)$
2. $b = O(J)$
3. bajo la hipótesis (a), $E(U) = b$
4. bajo las hipótesis (a) y (b), $V^{-1/2} \mathbf{R}^T \mathbf{G} \rightsquigarrow N(0, 1)$
5. bajo las hipótesis (a), (b) y (c), $V^{-1/2} \boldsymbol{\Delta}^T \mathbf{G} = O_p(J)$
6. bajo las hipótesis (a), (b), (c) y (d),

$$V^{-1/2} (U - E(U)) \boldsymbol{\beta}^T \mathbf{G} = O_p \left(J^{1/2} + \left(\frac{C(p, 4)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \right)^{1/2} \right).$$

Por (3.4) tenemos que

$$\begin{aligned} V^{-1/2} \left(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta}(1 + b) \right)^T \mathbf{G} &= \left(\frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right)^{-1} \left[V^{-1/2} \mathbf{R}^T \mathbf{G} + V^{-1/2} \boldsymbol{\Delta}^T \mathbf{G} \right. \\ &\quad \left. + V^{-1/2} (U - E(U)) \boldsymbol{\beta}^T \mathbf{G} + V^{-1/2} E(U) \boldsymbol{\beta}^T \mathbf{G} \right]. \end{aligned}$$

Para obtener nuestros resultados asintóticos vamos a requerir que

$$\frac{C(p, 4)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} \rightarrow 0, \quad \left(\frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right)^{-1} \left(\frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} - 1 \right) V^{-1/2} \boldsymbol{\beta}^T \mathbf{G} b \rightarrow 0.$$

Usando la conclusión 1 de la Proposición 3.2 para controlar el cociente $\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}} / \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}$, tenemos nuestro resultado principal

Teorema 3.3. *Supongamos que valen todas las hipótesis de la Proposición 3.2.*

(I) *Si $JV^{-1/2}\beta^T \mathbf{G}b \rightarrow 0$ y $C(p, 4)/n(\sigma^T \sigma)^4 \rightarrow 0$ cuando $n, p \rightarrow \infty$, entonces*

$$V^{-1/2} \left(\widehat{\beta}_{pls} - \beta(1+b) \right)^T \mathbf{G} \rightsquigarrow N(0, 1). \quad (3.8)$$

(II) *Si*

$$\frac{\Omega}{\sqrt{n}\sigma^T \sigma} = O(1), \quad \frac{C(p, 1)}{n^{3/4}\sigma^T \sigma} \rightarrow 0, \quad \frac{C(p, 2)}{n^{3/4}(\sigma^T \sigma)^2} \rightarrow 0 \quad y \quad \frac{C(p, 4)}{n(\sigma^T \sigma)^4} \rightarrow 0 \quad (3.9)$$

cuando $n, p \rightarrow \infty$, entonces se satisface la condición (I).

Aunque la sucesión no estocástica representada por b converge a 0 cuando $J \rightarrow 0$, representa un potencial sesgo al utilizar el Teorema 3.3 como punto de partida para inferencia en aplicaciones. Escribiendo

$$V^{-1/2} \left(\widehat{\beta}_{pls} - \beta(1+b) \right)^T \mathbf{G} = V^{-1/2} \left(\widehat{\beta}_{pls} - \beta \right)^T \mathbf{G} - V^{-1/2} \beta^T \mathbf{G}b \quad (3.10)$$

vemos que el término que es afectado por el sesgo es despreciable si $V^{-1/2}\beta^T \mathbf{G}b \rightarrow 0$ cuando $n, p \rightarrow \infty$. Como estamos pidiendo que $J \rightarrow 0$, la condición $V^{-1/2}\beta^T \mathbf{G}b \rightarrow 0$ implica la condición suficiente $JV^{-1/2}\beta^T \mathbf{G}b \rightarrow 0$ del Teorema 3.3, pero no viceversa. Si $\mathbf{G} \in \text{span}^\perp(\beta)$ entonces $\beta^T \mathbf{G} = 0$ y la contribución del sesgo es nula. Si $\mathbf{G} \notin \text{span}^\perp(\beta)$ entonces el sesgo puede jugar un rol notable. Para computar la importancia del sesgo, usamos cotas inferiores de V para encontrar condiciones en las que este término converge a 0. Notemos que

$$\begin{aligned} V^{-1}(\beta^T \mathbf{G})^2 &= n \frac{(\sigma^T \sigma / \Omega) \|\mathbf{G}^T \sigma\|^2}{\tau^2 \|\mathbf{G}^T \sigma\|^2 + \sigma_Y^2 (\sigma^T \sigma / \Omega) \mathbf{G}^T \ell_0 \Omega_0 \ell_0^T \mathbf{G}} \\ &\leq n \frac{\sigma^T \sigma}{\Omega \tau^2}, \end{aligned} \quad (3.11)$$

donde la cota superior se alcanza cuando $\mathbf{G} \in \text{span}(\beta)$. Como la elección de \mathbf{G} no afecta a b (ver (3.7)), esto indica que el efecto del sesgo será más prominente cuando $\mathbf{G} \in \text{span}(\beta)$. Se sigue de (3.11) que $V^{-1/2} |\beta^T \mathbf{G}b| \leq \sqrt{n}|b| (\sigma^T \sigma / \Omega \tau^2)^{1/2}$. En consecuencia, $\sqrt{n}|b| \rightarrow 0$ es una condición suficiente para evitar el efecto del sesgo cuando $\mathbf{G} \notin \text{span}^\perp(\beta)$. Observemos que

$$\sqrt{n}|b| \leq \sqrt{n} \left(\left| \frac{\sigma^T \sigma}{\Omega} - \tau^2 \right| \frac{C(p, 1)}{m\sigma^T \sigma} + \sigma_Y^2 \frac{C^2(p, 1)}{m^2 \Omega \sigma^T \sigma} + \sigma_Y^2 \frac{C(p, 2)}{m\Omega \sigma^T \sigma} \right).$$

Esta desigualdad nos da condiciones suficientes para $\sqrt{n}|b| \rightarrow 0$. Como $|\sigma^T \sigma / \Omega - \tau^2|$ y σ_Y^2 son acotados, tenemos

$$\sqrt{n} \frac{C(p, 1)}{m\sigma^T \sigma} \asymp \frac{C(p, 1)}{\sqrt{n}\sigma^T \sigma} \rightarrow 0,$$

la cual implica que $\sqrt{n}C^2(p, 1)/m^2\Omega\boldsymbol{\sigma}^T\boldsymbol{\sigma} \rightarrow 0$. Esto nos conduce a la última condición

$$\sqrt{n}\frac{C(p, 2)}{m\Omega\boldsymbol{\sigma}^T\boldsymbol{\sigma}} \asymp \frac{C(p, 2)}{\sqrt{n}(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2} \rightarrow 0.$$

Notemos que si $C(p, 2)/\sqrt{n}(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2$ se va a 0, entonces $C(p, 4)/n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4$ se va a 0. Resumimos todo esto en el siguiente resultado

Corolario 3.4. *Bajo todas las hipótesis de la Proposición 3.2*

(I) Si $\sqrt{n}|b| \rightarrow 0$ y $C(p, 4)/n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4 \rightarrow 0$ cuando $n, p \rightarrow \infty$, entonces

$$V^{-1/2}(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \mathbf{G} \rightsquigarrow N(0, 1). \quad (3.12)$$

(II) Si

$$\frac{C(p, 1)}{n^{1/2}\boldsymbol{\sigma}^T\boldsymbol{\sigma}} \rightarrow 0 \quad \text{y} \quad \frac{C(p, 2)}{n^{1/2}(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2} \rightarrow 0$$

cuando $n, p \rightarrow \infty$, entonces vale la hipótesis (I).

Las hipótesis de la parte (II) del Corolario 3.4 son importantes porque nos dan condiciones suficientes bajo las cuales podemos despreciar el potencial efecto del sesgo y por lo tanto construir intervalos de confianza asintóticos para $\boldsymbol{\beta}^T \mathbf{G}$ en la manera usual, como se describirá en la siguiente sección. Supongamos que los autovalores del ruido $\boldsymbol{\Omega}_0 = \text{var}(\boldsymbol{\ell}_0^T X)$ están acotados cuando $p \rightarrow \infty$ y que $\boldsymbol{\sigma}^T\boldsymbol{\sigma} \asymp p^t$ para $0 \leq t \leq 1$. Entonces si $n, p \rightarrow \infty$, $C(p, k)/\sqrt{n}(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^k \asymp p/\sqrt{n}(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^k \asymp n^{-1/2}p^{1-kt}$. Como $n^{-1/2}p^{1-t} \geq n^{-1/2}p^{1-2t}$, la conclusión del Corolario 3.4 valdrá si $n^{-1/2}p^{1-t} \rightarrow 0$.

TABLA 3.1. Implicaciones prácticas de la condición (II) del Corolario 3.4 cuando los autovalores de $\boldsymbol{\Omega}_0$ están acotados y $\boldsymbol{\sigma}^T\boldsymbol{\sigma} \asymp p^t$ para $0 \leq t \leq 1$.

t	0	1/2	3/4	1
$n \gg$	p^2	p	\sqrt{p}	1

Algunas implicaciones de esta condición se dan en la Tabla 3.1. En regresión sparse $t = 0$ debemos tener n relativamente grande con respecto a p^2 , lo cual es generalmente una condición muy compleja en regresión de alta dimensionalidad. Sin embargo, si la señal crece como $p^{3/4}$ entonces sólo necesitamos n del orden \sqrt{p} , lo cual podría ser alcanzable en muchas regresiones. Si la señal es abundante $t = 1$ entonces p no juega un rol y el comportamiento asintótico es equivalente a una configuración con p fijo. A su vez nuestros resultados son aplicables en el caso

p fijo siendo válidas las condiciones del Corolario 3.4 y sus conclusiones, coincidiendo con la Proposición 2.7.

Además de todo lo mencionado anteriormente, una conclusión interesante de nuestros resultados es que nos permiten concluir el resultado central de [Cook and Forzani, 2018, Teorema 1] cuya demostración se encuentran en el Apéndice F, a saber

Teorema 3.5. *Bajo las condiciones de la Proposición 3.2,*

$$(\widehat{\beta}_{pls} - \beta)^T \Sigma (\widehat{\beta}_{pls} - \beta) = O_p \left(n^{-1} + \left(\frac{C(p, 1)}{n \sigma^T \sigma} \right)^2 + \frac{C(p, 2)}{n (\sigma^T \sigma)^2} \right)$$

y por lo tanto para $\mathbf{X}_N - \boldsymbol{\mu}_X \sim N(\mathbf{0}, \Sigma)$ con Σ definida en (2.16) se tiene

$$(\widehat{\beta}_{pls} - \beta)^T (\mathbf{X}_N - \boldsymbol{\mu}_X) = O_p \left(n^{-1/2} + \frac{C(p, 1)}{n \sigma^T \sigma} + \left(\frac{C(p, 2)}{n (\sigma^T \sigma)^2} \right)^{1/2} \right).$$

3.4. Intervalos de confianza

Una vez hallada la distribución asintótica de estos estimadores, podemos construir intervalos de confianza asintóticos para $(1 + b)\beta^T \mathbf{G}$. El Teorema 3.3 nos permite concluir directamente que

$$CI_\alpha((1 + b)\beta^T \mathbf{G}) = \left[\widehat{\beta}_{pls}^T \mathbf{G} \pm z_{\alpha/2} V^{1/2} \right] \quad (3.13)$$

donde $z_{\alpha/2}$ denota un percentil seleccionado de la distribución normal estándar. Bajo el Corolario 3.4 el mismo intervalo se convierte en un intervalo de confianza para $\beta^T \mathbf{G}$, en cuyo caso nos referiremos a él como $CI_\alpha(\beta^T \mathbf{G})$.

Versiones muestrales pueden construirse para el intervalo dado en (3.13) usando un estimador plug-in para V . Para construir un estimador de V simplemente hacemos plug-in en sus términos constituyentes dados previamente, es decir

$$\widehat{V} = \frac{\widehat{\Omega}_{pls}^{-2}}{n} \mathbf{G}^T (\widehat{\Sigma} \widehat{\sigma}_Y^2 - \widehat{\sigma} \widehat{\sigma}^T) \mathbf{G}, \quad (3.14)$$

donde $\widehat{\Omega}_{pls} = \widehat{\sigma}^T \widehat{\Sigma} \widehat{\sigma} / \widehat{\sigma}^T \widehat{\sigma}$. El siguiente resultado, cuya demostración se encuentra en el Apéndice G muestra que asintóticamente es razonable reemplazar V por \widehat{V} en la construcción de intervalos de confianza

Teorema 3.6. *Bajo todas las hipótesis de la Proposición 3.2 se tiene que*

$$\frac{\widehat{V}}{V} \xrightarrow{p} 1, \quad n, p \rightarrow \infty. \quad (3.15)$$

Adicionalmente, tenemos una condición suficiente para la cual podemos reemplazar b por \widehat{b} , el cual nos permite obtener un intervalo de confianza para $\boldsymbol{\beta}^T \mathbf{G}$

Teorema 3.7. *Bajo las condiciones de la Proposición 3.2 y la condición (II) del Teorema 3.3, $\sqrt{n}(\widehat{b} - b) \xrightarrow{p} 0$ cuando $n, p \rightarrow \infty$, donde*

$$\widehat{b} = - \left[(\widehat{\sigma}_Y^2 - 2\widehat{\tau}^2) \frac{\widehat{C}(p, 1)}{m\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} + \widehat{\sigma}_Y^2 \frac{\widehat{C}^2(p, 1)}{m^2 \widehat{\Omega} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} + \frac{\widehat{C}(p, 2)}{m \widehat{\Omega} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} \right]$$

y $\widehat{C}(p, k) = \text{tr}(\widehat{\boldsymbol{\Sigma}}^k) - \widehat{\Omega}_{pls}^k$. En consecuencia, un intervalo de confianza asintótico para $\boldsymbol{\beta}^T \mathbf{G}$ es

$$CI_\alpha(\boldsymbol{\beta}^T \mathbf{G}) = \frac{1}{1 + \widehat{b}} \left[\widehat{\boldsymbol{\beta}}_{pls}^T \mathbf{G} \pm z_{\alpha/2} \widehat{V}^{1/2} \right].$$

En el siguiente capítulo veremos algunas simulaciones que nos ayudarán a comprobar la veracidad de estos intervalos.

3.5. Intervalos para predicción

Los intervalos de confianza discutidos en la sección anterior para $\boldsymbol{\beta}^T \mathbf{G}$ y $(1+b)\boldsymbol{\beta}^T \mathbf{G}$ requieren que \mathbf{G} sea un vector no estocástico de \mathbb{R}^p . La estimación para la media condicional en una nueva observación \mathbf{X}_N de \mathbf{X} es $\widehat{E}(Y|\mathbf{X}_N) = \bar{Y} + \widehat{\boldsymbol{\beta}}_{pls}^T (\mathbf{X}_N - \bar{\mathbf{X}})$. Si bien \mathbf{X}_N es fijo, tenemos que $\bar{\mathbf{X}}$ es estocástico y por lo tanto el correspondiente vector $\mathbf{G} = \mathbf{X}_N - \bar{\mathbf{X}}$ resulta ser estocástico. En consecuencia, los intervalos presentados en la sección anterior pueden no ser apropiados para una media condicional $E(Y|\mathbf{X}_N)$ o para predicción. En esta sección presentamos los intervalos para esas cantidades.

Sean

$$M(Y|\mathbf{X}_N) = \mu_Y + (1+b)\boldsymbol{\beta}^T (\mathbf{X}_N - \boldsymbol{\mu}_X)$$

$$D_N = \widehat{E}(Y|\mathbf{X}_N) - M(Y|\mathbf{X}_N)$$

$$\omega(\mathbf{X}_N) = V(\mathbf{X}_N - \boldsymbol{\mu}_X) + \frac{\tau^2}{n}$$

donde $V(\cdot)$ está definido en (3.6). La distribución asintótica de D_N está dada por el siguiente resultado

Teorema 3.8. *Supongamos que los datos (\mathbf{X}_i, Y_i) , $i = 1, \dots, n$ son observaciones independientes de un vector aleatorio normal multivariado de dimensión $p+1$ (\mathbf{X}, Y) que sigue el modelo dado por (2.15) y que además satisface las hipótesis de la Proposición 3.2. Entonces*

$$\omega^{-1/2}(\mathbf{X}_N) D_N \rightsquigarrow N(0, 1). \quad (3.16)$$

La demostración de este teorema se construye de la siguiente forma. Primero, haciendo un poco de álgebra se tiene que

$$\begin{aligned}\omega^{-1/2}(\mathbf{X}_N)D_N &= \omega^{-1/2}(\mathbf{X}_N) (\bar{Y} - \mu_Y - \beta^T(\bar{\mathbf{X}} - \boldsymbol{\mu}_X)) \\ &\quad + \omega^{-1/2}(\mathbf{X}_N) \left((\hat{\boldsymbol{\beta}}_{pls} - (1+b)\boldsymbol{\beta})^T (\mathbf{X}_N - \boldsymbol{\mu}_X) \right) \\ &\quad - \omega^{-1/2}(\mathbf{X}_N) (\hat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_X).\end{aligned}\tag{3.17}$$

El último término converge a 0 en probabilidad así que por el teorema de Slutsky no juega ningún rol en la distribución asintótica del término izquierdo de (3.16).

Lema 3.9. *Bajo las condiciones del Teorema 3.8,*

$$\omega^{-1/2}(\mathbf{X}_N) (\hat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_X) \xrightarrow{p} 0$$

cuando $n, p \rightarrow \infty$.

El siguiente lema muestra características de los términos restantes

Lema 3.10. *Bajo las condiciones del Teorema 3.8,*

$$W_{n,p} = V(\mathbf{X}_N - \boldsymbol{\mu}_X)^{-1/2} \left(\hat{\boldsymbol{\beta}}_{pls} - (1+b)\boldsymbol{\beta} \right)^T (\mathbf{X}_N - \boldsymbol{\mu}_X) \rightsquigarrow N(0, 1)\tag{3.18}$$

$$Z_{n,p} = \sqrt{n}\tau^{-1} (\bar{Y} - \mu_Y - \beta^T(\bar{\mathbf{X}} - \boldsymbol{\mu}_X)) \sim N(0, 1)\tag{3.19}$$

$$\bar{Y} - \mu_Y - \beta^T(\bar{\mathbf{X}} - \boldsymbol{\mu}_X) \perp\!\!\!\perp \left(\hat{\boldsymbol{\beta}}_{pls} - (1+b)\boldsymbol{\beta} \right)^T (\mathbf{X}_N - \boldsymbol{\mu}_X).\tag{3.20}$$

La conclusión (3.18) se sigue inmediatamente del Teorema 3.3; (3.19) se sigue de la distribución normal multivariada y porque vale el modelo de regresión lineal; (3.20) se sigue ya que bajo normalidad $(\bar{\mathbf{X}}, \bar{Y}) \perp\!\!\!\perp \hat{\boldsymbol{\beta}}_{pls}$.

Sean

$$a_{n,p} = \frac{\tau/\sqrt{n}}{\omega^{1/2}(\mathbf{X}_N)} \quad y \quad b_{n,p} = \frac{V^{1/2}(\mathbf{X}_N - \boldsymbol{\mu}_X)}{\omega^{1/2}(\mathbf{X}_N)}.$$

Despreciando el término que ya dijimos era irrelevante para la distribución asintótica, sólo nos interesa trabajar con

$$\omega^{-1/2}(\mathbf{X}_N)D_N = a_{n,p}Z_{n,p} + b_{n,p}W_{n,p}.\tag{3.21}$$

De aquí vemos que asintóticamente $\omega^{-1/2}(\mathbf{X}_N)D_N$ es una combinación lineal no estocástica de dos variables aleatorias independientes, donde cada una converge en distribución a una normal estándar.

A partir de (3.18) y (3.19) y observando que $a_{n,p}^2 + b_{n,p}^2 = 1$, el siguiente resultado general concluye la prueba del Teorema 3.8.

Lema 3.11. *Sea $(a_{n,p}, b_{n,p})$ una sucesión en \mathbb{R}^2 tal que $a_{n,p}^2 + b_{n,p}^2 = 1$ para todo n, p , y sea $(Z_{n,p}, W_{n,p})$ un vector aleatorio en \mathbb{R}^2 tal que $\begin{pmatrix} Z_{n,p} \\ W_{n,p} \end{pmatrix} \rightsquigarrow N_2(\mathbf{0}, \mathbf{I})$, entonces*

$$a_{n,p}Z_{n,p} + b_{n,p}W_{n,p} \rightsquigarrow N(0, 1).$$

El Teorema 3.8 nos permite construir intervalos de confianza basándonos en $\omega^{-1/2}(\mathbf{X}_N)D_N$. Sin embargo, nuestro principal interés es deducir resultados para $E(Y|\mathbf{X})$ a partir de $\widehat{E}(Y|\mathbf{X}) - E(Y|\mathbf{X}) = D_N + b\boldsymbol{\beta}^T(\mathbf{X}_N - \boldsymbol{\mu}_X)$. Escalando esta cantidad con la varianza tenemos

$$\omega^{-1/2}(\mathbf{X}_N) \left(\widehat{E}(Y|\mathbf{X}) - E(Y|\mathbf{X}) \right) = \omega^{-1/2}(\mathbf{X}_N)D_N + \omega^{-1/2}(\mathbf{X}_N)b\boldsymbol{\beta}^T(\mathbf{X}_N - \boldsymbol{\mu}_X).$$

Para el cuadrado del último sumando

$$\omega^{-1}(\mathbf{X}_N)b^2 (\boldsymbol{\beta}^T(\mathbf{X}_N - \boldsymbol{\mu}_X))^2 \leq V^{-1}(\mathbf{X}_N - \boldsymbol{\mu}_X)b^2 (\boldsymbol{\beta}^T(\mathbf{X}_N - \boldsymbol{\mu}_X))^2 \leq nb^2 \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega \tau^2}.$$

En consecuencia, terminamos concluyendo un argumento similar a aquel que nos condujo al Corolario 3.4 y obtenemos el siguiente resultado

Corolario 3.12. *Supongamos que valen las condiciones del Teorema 3.8 y que $\sqrt{n}|b| \rightarrow 0$ cuando $n, p \rightarrow \infty$. Entonces*

$$\omega^{-1/2}(\mathbf{X}_N) \left(\widehat{E}(Y|\mathbf{X}) - E(Y|\mathbf{X}) \right) \rightsquigarrow N(0, 1).$$

Con este corolario podemos construir intervalos aproximados para $E(Y|\mathbf{X}_N)$ y para $Y_N = E(Y|\mathbf{X}_N) + \varepsilon_N$ usando un método plug-in para obtener un estimador $\widehat{\omega}$ de ω , es decir

$$CI_\alpha(E(Y|\mathbf{X}_N)) = \left[\widehat{E}(Y|\mathbf{X}_N) \pm z_{\alpha/2} \widehat{\omega}^{1/2}(\mathbf{X}_N) \right] \quad (3.22)$$

$$CI_\alpha(Y_N) = \left[\widehat{E}(Y|\mathbf{X}_N) \pm z_{\alpha/2} (\widehat{\omega}(\mathbf{X}_N) + \widehat{\tau}^2)^{1/2} \right]. \quad (3.23)$$

De manera similar a las interpretaciones que dimos en la sección anterior, si $\sqrt{n}|b|$ no converge a 0, entonces (3.22) y (3.23) pueden ser interpretadas, bajo el Teorema 3.8, como intervalos de confianza para $M(Y|\mathbf{X}_N)$ y $M(Y|\mathbf{X}_N) + \varepsilon_N$. En el caso en que $\sqrt{n}|b|$ no converge a 0 pero las condiciones del Teorema 3.7 se satisfacen, aún podemos usar \widehat{b} para construir intervalos de confianza.

CAPÍTULO 4

SIMULACIONES Y APLICACIONES

4.1. Distribución asintótica

Ilustramos ahora una simulación que muestra el potencial impacto que puede tener el sesgo sobre la distribución asintótica de $\widehat{\beta}_{pls}^T \mathbf{G}$. En este caso, la proporción que elegimos es $n = p/2$. En el marco del modelo para una componente dado por las ecuaciones (2.15) y (2.16), vamos a tomar $\mu_Y = 0$, $\tau = 1/2$, $\Omega = \boldsymbol{\sigma}^T \boldsymbol{\sigma}$ y $\boldsymbol{\Omega}_0 = \mathbf{I}_{p-1}$. El vector $\boldsymbol{\sigma}$ fue generado con $\lfloor p^{1/2} \rfloor$ elementos con distribución normal estándar y los restantes igualados a 0. De esta forma se tiene que $\boldsymbol{\sigma}^T \boldsymbol{\sigma} \asymp \sqrt{p}$. Notemos que bajo esta configuración $C(p, k)/n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^k \asymp p^{-k/2}$ y por lo tanto $J \rightarrow 0$ y $C(p, 4)/n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4 \rightarrow 0$. Nuevamente para enfatizar el efecto del sesgo tomamos $\mathbf{G} = \boldsymbol{\sigma}$ como en la discusión de la ecuación (3.11). Con esta configuración, \mathbf{X} fue generado con distribución $N(\mathbf{0}, \boldsymbol{\Sigma})$ y luego generamos Y acorde a (2.15) con errores que siguen una distribución $N(0, \tau^2)$. Para cada par (n, p) la simulación fue repetida 500 veces y se construyeron dos histogramas. Uno representa la expresión $D_1 = V^{-1/2} \left(\widehat{\beta}_{pls} - \beta(1+b) \right)^T \mathbf{G}$ y el otro $D_2 = V^{-1/2} \left(\widehat{\beta}_{pls} - \beta \right)^T \mathbf{G}$. El resultado se muestra en la Figura 4.1.

Como $\sqrt{n}|b|J \asymp J \rightarrow 0$, tal como lo predice el Teorema 3.3 el histograma de D_1 se desplaza para coincidir con la densidad de una normal estándar, mientras que el histograma para D_2 no converge en distribución a una normal estándar pues no se satisfacen las condiciones del Corolario 3.4 con las configuraciones de esta simulación, por ejemplo $\sqrt{n}|b| \asymp 1$.

4.2. Coverage

Hemos construido una serie de simulaciones para ilustrar el comportamiento de la versión muestral de los intervalos de confianza dados en (3.13) cuando $n = \sqrt{p}$ crece con p . Para cada

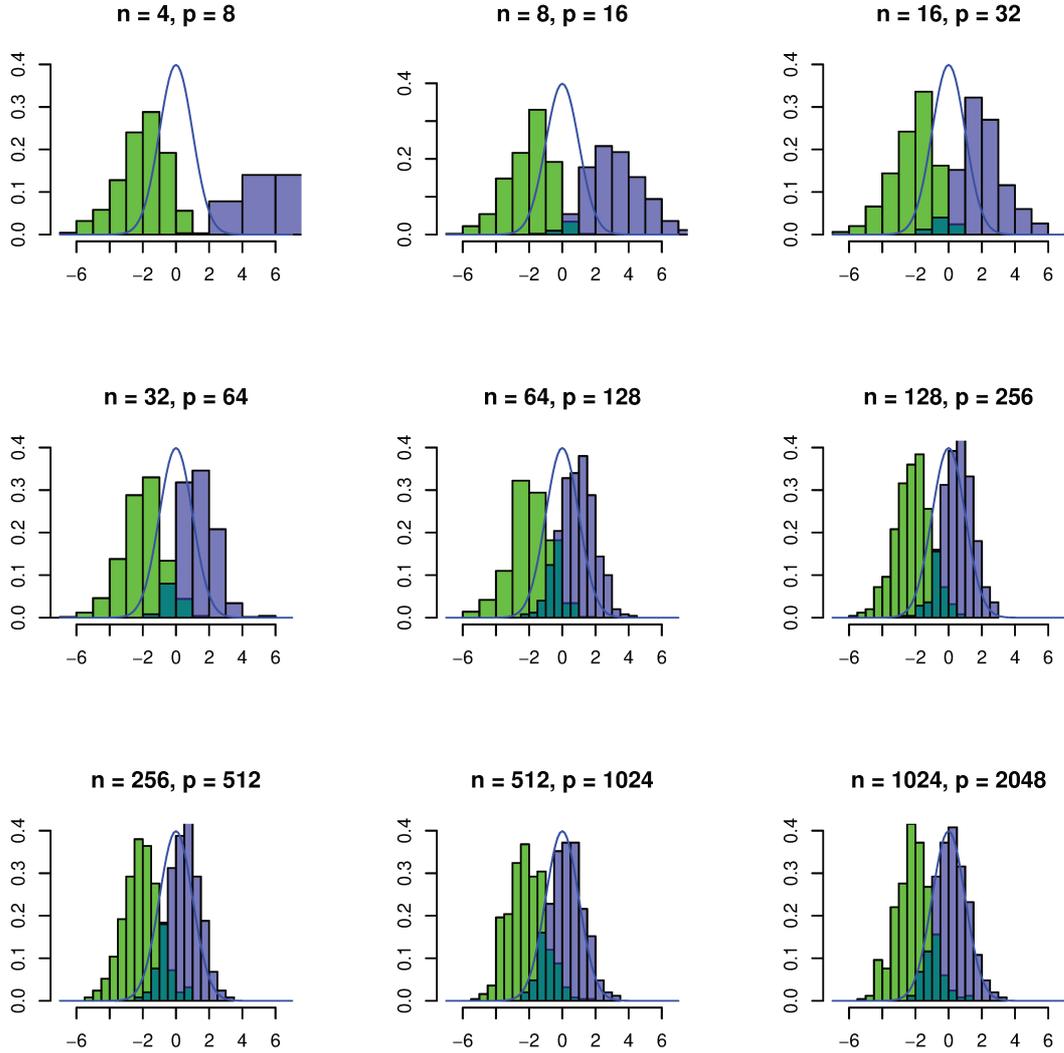


FIGURA 4.1. Resultado de la simulación para las distribuciones de D_1 y D_2 : El histograma derecho (azul) representa el término D_1 mientras que el histograma izquierdo (verde) es para D_2 . La curva azul representa la densidad de una normal estándar.

par (n, p) construimos σ de forma que tenga $\lfloor p^{3/4} \rfloor$ elementos generados independientemente de una distribución normal estándar con los restantes $p - \lfloor p^{3/4} \rfloor$ elementos seteados a 0. Con $p = 1024$ como el valor más grande considerado en nuestra simulación, σ tiene 181 elementos distintos de 0. Luego establecemos $\Omega = \sigma^T \sigma$, $\Omega_0 = \mathbf{I}_{p-1}$ y (σ, σ_0) que sea una matriz ortogonal de $p \times p$. A continuación creamos Σ de acuerdo a (2.16), $\beta = \Omega^{-1} \sigma$ y con esto ya podemos crear nuestros datos $(\mathbf{X}_i, Y_i), i = 1, \dots, n$ independientes tal que $\mathbf{X}_i \sim N(\mathbf{0}, \Sigma)$, $Y|\mathbf{X} \sim N(\beta^T \mathbf{X}_i, 1)$. Por conveniencia hemos elegido tomar \mathbf{G} como un vector con distribución multinormal estándar

y luego se construye el intervalo de confianza aproximado de nivel 95% de acuerdo a (3.13) usando el estimador plugin de V . Esta simulación fue repetida 1000 veces y se contaron la cantidad de veces que el intervalo contenía a $(1+b)\beta^T \mathbf{G}$ y $\beta^T \mathbf{G}$. Los resultados se presentan en la Tabla 4.1.

Este ejemplo satisface las condiciones de la Proposición 3.2 y por lo tanto vale el Teorema 3.3. Por ejemplo

$$\frac{C(p, k)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^k} \asymp \frac{1}{p^{3k/4-1/2}}.$$

Más aún vale la condición (I) del Corolario 3.4 pues $\sqrt{n}|b| \asymp n^{-1/2}$. Sin embargo, no vale la condición (II) de Corolario 3.4 pues $\frac{C(p, 1)}{n^{1/2}\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \asymp 1$.

TABLA 4.1. Resultados del coverage para los intervalos de confianza (3.13) para el valor indicado en la primera fila. La tercera y cuarta columna son para la configuración donde $\sqrt{n}|b| \rightarrow 0$, la quinta y sexta columna son para la situación donde $\sqrt{n}|b| \not\rightarrow 0$, y la última columna es para el intervalo ajustado dado en el Teorema 3.7.

n	p	$\beta^T \mathbf{G}$	$(1+b)\beta^T \mathbf{G}$	$\beta^T \mathbf{G}$	$(1+b)\beta^T \mathbf{G}$	$\beta^T \mathbf{G}$
16	256	0.924	0.897	0.766	0.887	0.856
22	512	0.949	0.920	0.734	0.927	0.903
32	1024	0.947	0.946	0.742	0.952	0.920

Cabe mencionar algo importante que sucede con esta configuración. Hemos tomado $\Omega = \boldsymbol{\sigma}^T \boldsymbol{\sigma}$ y $\text{var}(Y|\mathbf{X}) = \tau^2 = 1$. Si recordamos que $\beta^T \boldsymbol{\Sigma} \beta = \Omega^{-1} \boldsymbol{\sigma}^T \boldsymbol{\sigma}$ y observamos la ecuación (3.7) vemos que el primer término se anula. Como los términos restantes del sesgo son de menor orden, esta configuración hace menos prominente el efecto del sesgo, permitiendo acelerar la convergencia.

Para enfatizar el efecto que puede tener el sesgo en los intervalos de confianza, vamos a repetir la simulación pero esta vez tomando $\tau = 1/2$, $\mathbf{G} = \boldsymbol{\sigma}$ y usamos el verdadero valor de V . Como fue remarcado anteriormente, esta elección de \mathbf{G} donde $\mathbf{G} \in \text{span}(\boldsymbol{\beta})$ maximizará el efecto del sesgo, y ahora el primer sumando de (3.7) va a contribuir al sesgo. En la Tabla 4.1 vemos que en esta situación los resultados para $CI_\alpha(\beta^T \mathbf{G})$ sufren en comparación con $CI_\alpha((1+b)\beta^T \mathbf{G})$ ya que $\sqrt{n}|b|$ no converge a 0, sino que $\sqrt{n}|b| \asymp 1$. Sin embargo, $JV^{-1/2}|\beta^T \mathbf{G} b| \leq$

$\sqrt{n}J|b|(\boldsymbol{\sigma}^T \boldsymbol{\sigma} / \Omega \tau^2)^{1/2} = \sqrt{2n}J|b| \rightarrow 0$ y vale el Teorema 3.3. Más aún, valen las condiciones (II) del Teorema 3.3 y por lo tanto el intervalo ajustado del Teorema 3.7.

4.3. Datos: mejillones

En [Cook, 2018, Sección 4.4.4] se utilizó PLS para analizar una base de datos provenientes de un estudio ecológico de mejillones, donde la muestra fue tomada en Marlborough Sounds, una extensa red de valles de la isla sur de Nueva Zelanda. La variable respuesta Y es el logaritmo de la masa muscular del mejillón, y 4 variables predictoras fueron consideradas, siendo éstas los logaritmos de la altura, ancho y largo del caparazón (medidas en milímetros), y de la masa del caparazón (medido en gramos). El tamaño de muestra es $n = 79$ y $p = 4$, es decir estamos en el caso clásico de p fijo y $n > p$. En la Figura 4.2 vemos el resultado de hacer un método de cross-validation para determinar el número de componentes al ajustar el modelo PLS con estos datos. Observamos que una sola componente es adecuada, por lo tanto estamos bajo el contexto $q = 1$ de nuestro estudio.

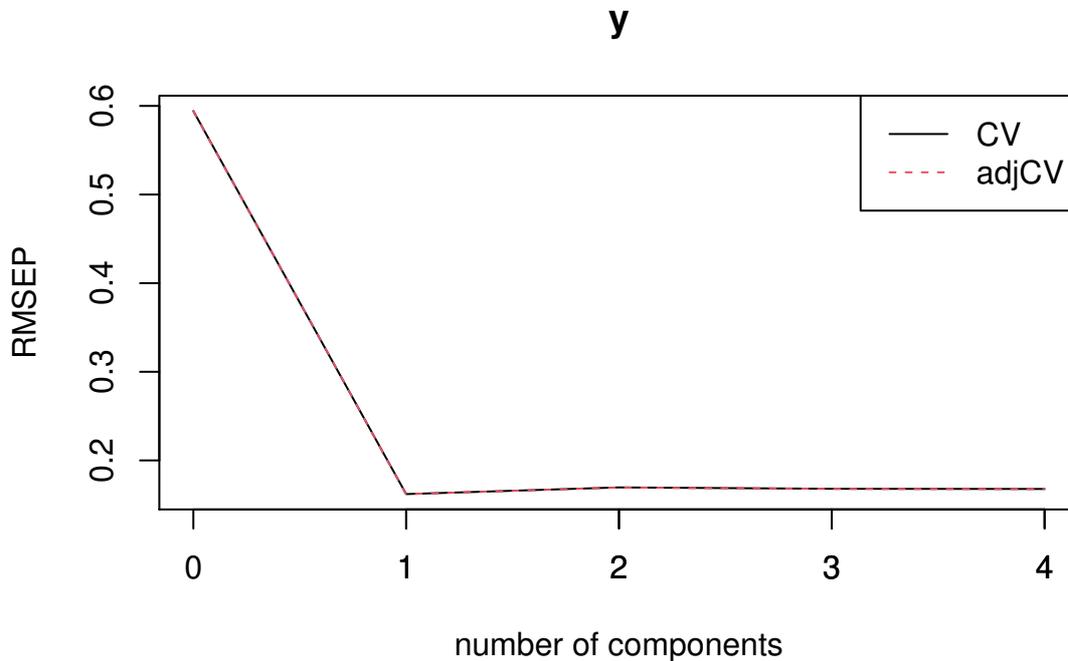


FIGURA 4.2. Determinación del número de componentes por cross-validation para los datos de mejillones.

A modo de ilustración, en la Figura 4.3 vemos el gráfico de la respuesta versus los valores predichos del ajuste PLS con una componente. El gráfico muestra una clara relación lineal sin

dar señales de que haya algún comportamiento anómalo. En la Figura 4.4 vemos el gráfico de los valores ajustados por el modelo PLS con una componente versus los valores ajustados por el método clásico de OLS.

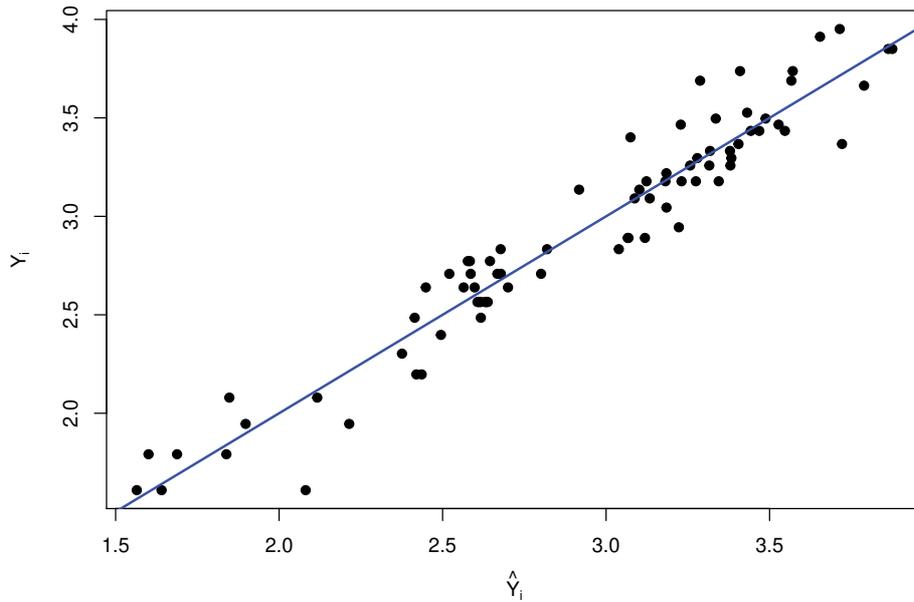


FIGURA 4.3. Datos de mejillones: Plot con las respuestas observadas Y_i versus los valores ajustados \hat{Y}_i que provienen del modelo PLS con una componente, $q = 1$.

Los valores de las estimaciones de los coeficientes dados por los ajustes OLS, PLS y envelope, así como también sus errores estándares asintóticos se muestra en Tabla 4.2. Los errores estándar de OLS son los usuales, mientras que los errores estándar de PLS son $\text{diag}(\hat{V}^{1/2})$, donde \hat{V} está definida en (3.14). Los errores estándar de envelope fueron obtenidos usando estimación plug-in de la matriz de covarianza asintótica de los estimadores de los coeficientes en [Cook et al., 2013]. Ver también Teorema 4.5 de [Cook, 2018]. Vemos que los errores de la estimación dada por OLS son entre 6 y 60 veces más que aquellos correspondientes al estimador de PLS. Este tipo de diferencias no son inusuales en la literatura de envelopes, tal como lo muestran los ejemplos en [Cook, 2018]. Sin embargo, no notamos una diferencia radical en los errores estándar entre las estimaciones de envelope y PLS. Aun los valores para PLS son ligeramente más grandes, los órdenes parecen coincidir.

Las diferencias en la Tabla 4.2 se reflejan visualmente en la Figura 4.5 la cual contiene la representación visual de los intervalos de confianza para los coeficientes de OLS, PLS y

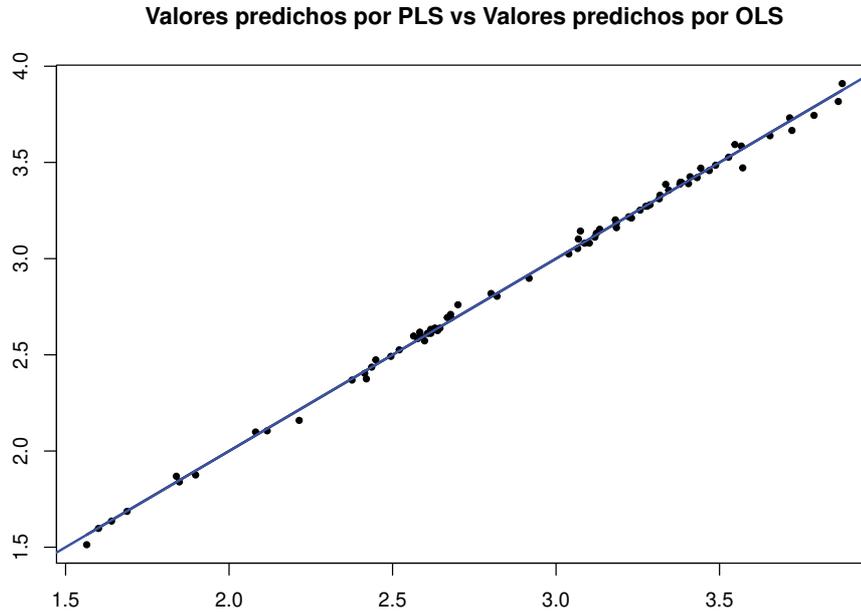


FIGURA 4.4. Datos de mejillones: Plot de los valores ajustados \hat{Y}_i dados por el modelo PLS con una componente $q = 1$, versus los valores ajustados utilizando el modelo OLS. La recta azul representa $y = x$ y se incluye para mejor claridad.

TABLA 4.2. Estimación de los coeficientes y sus correspondientes desvíos estándar (S.S.) correspondientes a tres ajustes para los datos de mejillones: Ordinary least squares (OLS), Partial Least Squares con una componente (PLS), y envelope en una dimensión (ENV).

	OLS		PLS, $q = 1$		ENV, $q = 1$	
	Estimación	S.D.	Estimación	S.D.	Estimación	S.D.
$\hat{\beta}_1$	0.741	0.410	0.142	0.0063	0.141	0.0052
$\hat{\beta}_2$	-0.113	0.399	0.153	0.0067	0.154	0.0056
$\hat{\beta}_3$	0.567	0.118	0.625	0.0199	0.625	0.0194
$\hat{\beta}_4$	0.170	0.304	0.206	0.0086	0.206	0.0073

envelope, donde cada intervalo se computó como estimador $\pm z_{\alpha/2}S.D.$ Como era de esperarse, los intervalos de confianza para OLS dominan el gráfico, mientras que los correspondientes a PLS y envelope se perciben casi como puntos a la derecha del intervalo dado por OLS. Nuevamente, esta diferencia entre los intervalos dados por OLS y los otros dos modelos es bastante típica.

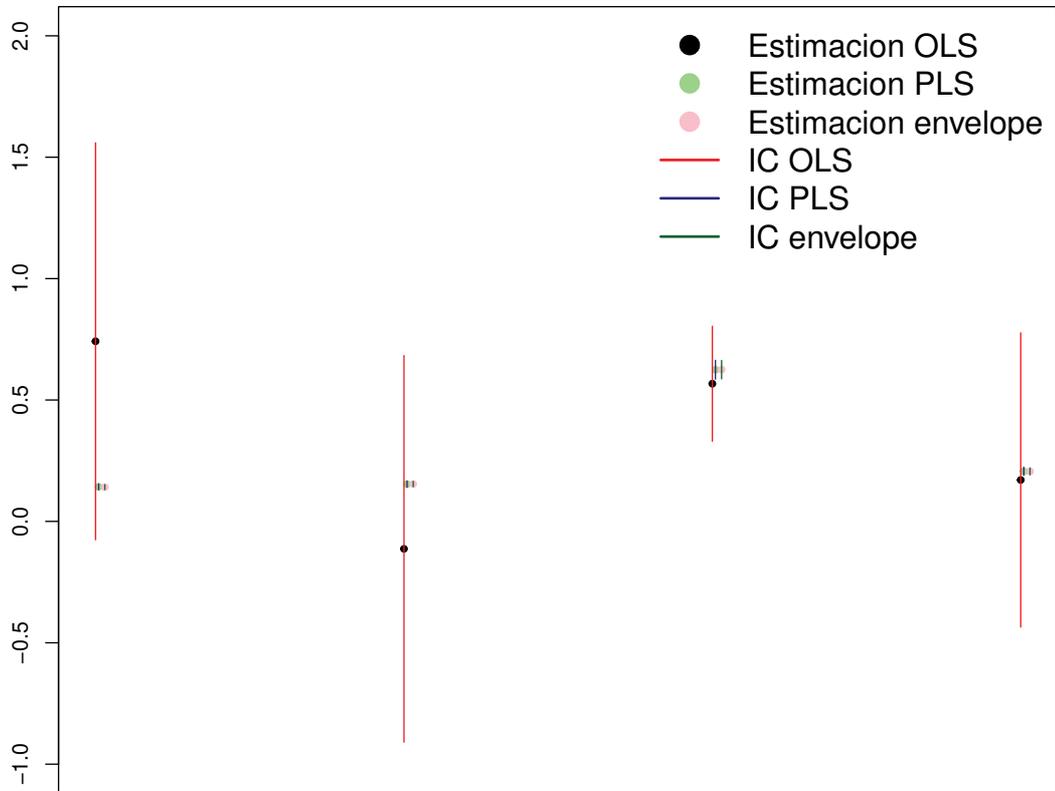


FIGURA 4.5. Datos de mejillones: Plot de los intervalos de confianza para cada uno de los cuatro coeficientes del vector β , utilizando los métodos OLS, PLS y envelope.

4.4. Datos: dopamina

En [Chiappini et al., 2021] presentaron los resultados de una serie de experimentos para predecir la concentración de dopamina pura Y . Para el vector de predictores, ellos midieron $p = 706$ puntos muestrales a lo largo de curvas de espectrogramas para cada una de las $n = 10$ preparaciones que fijaron. Como cada muestra consiste de un analito (sustancia química a ser estudiada) puro de dopamina, aplicaciones de la ley de Beer indican que la regresión PLS sólo posee una componente [Nadler and Coifman, 2005]. Nuestro análisis de componentes para PLS en estos datos (similar a lo que vemos en la Figura 4.2) mostraron como conclusión que podemos tomar $q = 1$. Usamos estos datos como base para una simulación en la que deseamos estudiar la precisión de los intervalos hallados en la Sección 3.5.

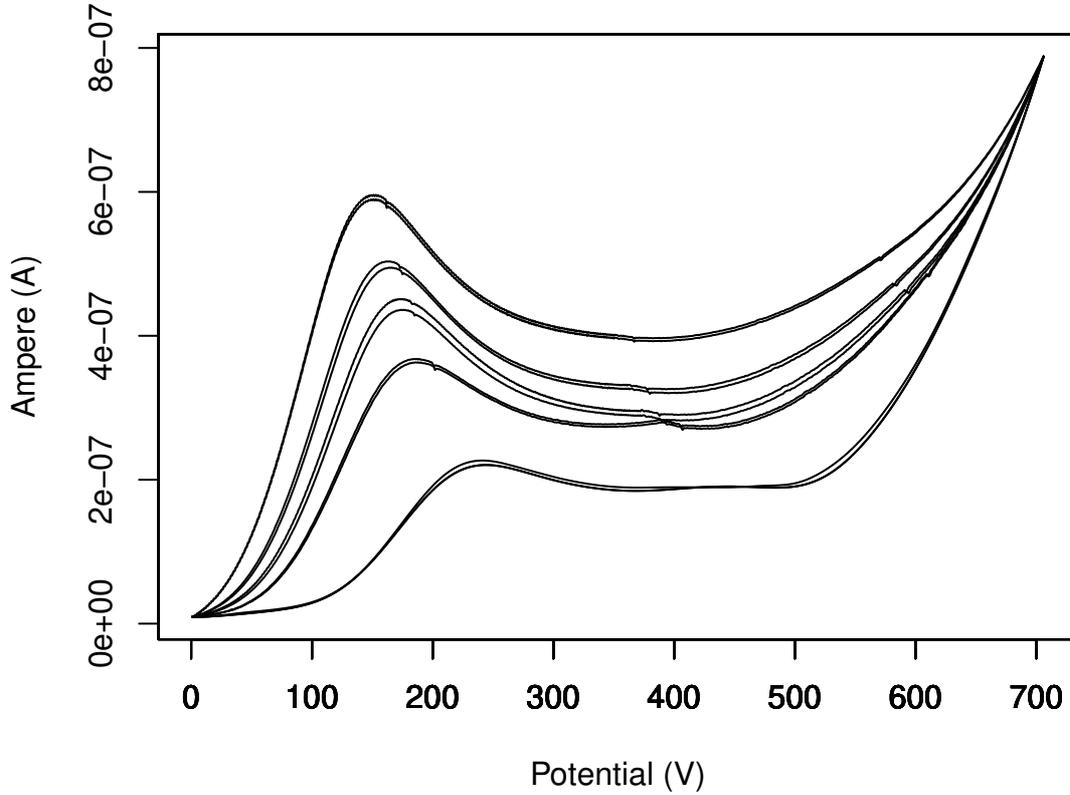


FIGURA 4.6. Voltamperograma: en esta figura se ven lo que los quimiométricos obtienen al hacer sus estudios. Cada curva representa una muestra preparada a una cierta concentración. El vector de predictores se forma tomando una discretización de cada curva.

Luego de ajustar el modelo correspondiente a PLS con una componente, hallamos $\hat{\beta}_{pls}, \hat{\tau}_{pls}$ y los tratamos como parámetros poblacionales para generar 10 nuevas respuestas tal que $y_i = \bar{Y} + \hat{\beta}_{pls}^T (\mathbf{X}_i - \bar{\mathbf{X}}) + \hat{\tau}_{pls} \varepsilon_i$ para $i = 1, \dots, 10$ donde los ε_i son copias independientes de una variable aleatoria con distribución normal estándar. Usando el método *leave-one-out*, removemos la j -ésima observación, usamos PLS con una componente para ajustar las restantes 9 observaciones, construimos los intervalos de confianza para $E(Y|\mathbf{X}_j)$ y $E(Y|\mathbf{X}_j) + \varepsilon_j$, y luego evaluamos si $E(Y|\mathbf{X}_j)$ y Y_j caen o no en los correspondientes intervalos. Repetimos el proceso para cada $j = 1, \dots, 10$ y luego replicamos toda la simulación 1000 veces. Además hemos repetido toda la simulación para $n = 100$.

Los resultados se presentan en la Tabla 4.3. La segunda y tercer columna nos muestra los valores de coverage de los intervalos para $E(Y|\mathbf{X}_N)$ y para Y_N , como fueron presentados en (3.22)

y (3.23). Los índices de coverage para $n = 10$ son bajos, pero se observa un comportamiento muy bueno para $n = 100$. Notamos además que los valores de coverage para $n = 10$ son similares a aquellos que vimos para tamaños de muestra chico en Tabla 4.1. Desde luego, el tamaño de muestra requerido para alcanzar valores cercanos al nominal van a depender de muchas características de la regresión.

TABLA 4.3. Simulación para los datos de dopamina: Coverage observados para intervalos de confianza aproximados de 95 % para $E(Y | X_N)$ y Y_N .

n	$CI_{0.05}(E(Y X_N))$	$CI_{0.05}(Y_N)$
10	0.864	0.886
100	0.943	0.946

Los intervalos de confianza que hemos usado para ilustrar este caso se basan en la suposición de la normalidad multivariada de \mathbf{X} tal como vimos en la Proposición 3.2. Si bien nosotros simulamos errores con distribución normal, no hemos estudiado la distribución del vector de predictores $\mathbf{X} \in \mathbb{R}^{706}$, sino que usamos los valores observados en [Chiappini et al., 2021]. Como los índices mostrados en la Tabla 4.3 se acercan al valor nominal de 0.95, la distribución normal de los predictores puede no ser un requerimiento estricto para resultados representativos siempre que se disponga de tamaños de muestra suficientemente grandes.

BIBLIOGRAFÍA

- [Antunes Percíncula et al., 2020] Antunes Percíncula, C., Forzani, L., and Toledano, R. (2020). Wishart moments calculator.
- [Berntsson and Wold, 1986] Berntsson, P. and Wold, S. (1986). Comparison between x-ray crystallographic data and physicochemical parameters with respect to their information about the calcium channel antagonist activity of 4-phenyl-1,4-dihydropyridines. *Quantitative Structure-Activity Relationships*, 5(2):45–50.
- [Bickel and Levina, 2008] Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577 – 2604.
- [Chiappini et al., 2021] Chiappini, F. A., Gutierrez, F., Goicoechea, H. C., and Olivieri, A. C. (2021). Interference-free calibration with first-order instrumental data and multivariate curve resolution. when and why? *Analytica Chimica Acta*, 1161:338465.
- [Chun and Keles, 2010] Chun, H. and Keles, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society Series B*, 72(1):3–25.
- [Cook, 2018] Cook, R. D. (2018). *An Introduction to Envelopes*. New York: Wiley.
- [Cook and Forzani, 2018] Cook, R. D. and Forzani, L. (2018). Big data and partial least-squares prediction. *Canadian Journal of Statistics*, 46(1):62–78.
- [Cook et al., 2013] Cook, R. D., Helland, I. S., and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society Series B*, 75(5):851–877.
- [Cook et al., 2007] Cook, R. D., Li, B., and Chiaromonte, F. (2007). Dimension reduction in regression without matrix inversion. *Biometrika*, 94(3):569–584.
- [Cook et al., 2010] Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate regression (with discussion). *Statist. Sin.*, 20:927–1010.
- [de Jong, 1993] de Jong, S. (1993). Simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263.
- [Fisher, 1922] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368.
- [Galton, 1886] Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature.
- [Geladi, 1988] Geladi, P. (1988). Notes on the history and nature of partial least squares (pls) modelling. *Journal of Chemometrics*, 2(4):231–246.
- [Giraud, 2014] Giraud, C. (2014). *Introduction to High-Dimensional Statistics*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

- [Gottfries et al., 1995] Gottfries, J., Blennow, K., Wallin, A., and Gottfries, C. G. (1995). Diagnosis of dementias using partial least squares discriminant analysis. *Dementia*, 6(2):83–88.
- [Helland, 1988] Helland, I. S. (1988). On the structure of partial least squares regression. *Communications in Statistics - Simulation and Computation*, 17(2):581–607.
- [Helland, 1990] Helland, I. S. (1990). Partial least squares regression and statistical models. *Scandinavian Journal of Statistics*, 17(2):97–114.
- [Huang et al., 2006] Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93(1):85–98.
- [Jagpal and Brick, 1982] Jagpal, H. S. and Brick, I. E. (1982). The marketing mix decision under uncertainty. *Marketing Science*, 1(1):79–92.
- [Kettaneh-Wold, 1992] Kettaneh-Wold, N. (1992). Analysis of mixture data with partial least squares. *Chemometrics and Intelligent Laboratory Systems*, 14(1):57–69. Proceedings of the 2nd Scandinavian Symposium on Chemometrics.
- [Marcenko and Pastur, 1967] Marcenko, V. and Pastur, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Math USSR Sb*, 1:457–483.
- [Nadler and Coifman, 2005] Nadler, B. and Coifman, R. R. (2005). Partial least squares, beer’s law and the net analyte signal: statistical modeling and analysis. *Journal of Chemometrics*, 19(1):45–54.
- [Naik and Tsai, 2000] Naik, P. and Tsai, C.-L. (2000). Partial least squares estimator for single-index models. *Journal of the Royal Statistical Society Series B*, 62:763–771.
- [Noonan and Wold, 1977] Noonan, R. and Wold, H. (1977). Nipals path modelling with latent variables. *Scandinavian Journal of Educational Research*, 21:33–61.
- [Rothman et al., 2008] Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2(none):494 – 515.
- [S. Wold and et al., 1984] S. Wold, A. R. and et al., H. W. (1984). The collinearity problem in linear regression. the partial least squares (pls) approach to generalized inverses. *SIAM J. Sci. and Stat. Comput.*, 5(3):735–743.
- [Saaksjarvi et al., 1989] Saaksjarvi, E., Khalighi, M., and Minkkinen, P. (1989). Waste water pollution modelling in the southern area of lake saimaa, finland, by the simca pattern recognition method. *Chemometrics and Intelligent Laboratory Systems*, 7(1):171–180. Proceedings of the First Scandinavian Symposium on Chemometrics.
- [Schwartz et al., 2009] Schwartz, W. R., Kembhavi, A., Harwood, D., and Davis, L. S. (2009). Human detection using partial least squares analysis. In *2009 IEEE 12th International Conference on Computer Vision*, pages 24–31.
- [Stigler, 1981] Stigler, S. M. (1981). Gauss and the invention of least squares. *The Annals of Statistics*, 9(3):465–474.
- [Vinzi et al., 2010] Vinzi, V., Chin, W., and Henseler, J. (2010). *Handbook of Partial Least Squares*. Springer handbooks of computational statistics. Springer My Copy UK.
- [Vong et al., 1988] Vong, R., Geladi, P., Wold, S., and Esbensen, K. (1988). Source contributions to ambient aerosol calculated by discriminat partial least squares regression (pls). *Journal of Chemometrics*, 2(4):281–296.
- [Wold, 1938] Wold, H. (1938). A study in analysis of stationary time series. *Almqvist & Wiksell, Stockholm*.

-
- [Wold, 1966a] Wold, H. (1966a). Estimation of principal components and related models by iterative least squares. In *P.R. Krishnaiah (Ed.). Multivariate Analysis*, pages 391–420.
- [Wold, 1966b] Wold, H. (1966b). Nonlinear estimation by iterative least squares procedures. *Research Papers in Statistics; Festschrift for J. Neyman*, pages 411–444.
- [Wold, 2001] Wold, S. (2001). Personal memories of the early pls development. *Chemometrics and Intelligent Laboratory Systems - CHEMOMETR INTELL LAB SYST*, 58:83–84.
- [Wold et al., 1996] Wold, S., Kettaneh, N., and Tjessem, K. (1996). Hierarchical Multiblock PLS and PC Models for Easier Model Interpretation and as an Alternative to Variable Selection. *Journal of Chemometrics*, 10:463–482.
- [Wold et al., 2001] Wold, S., Sjostrom, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130. PLS Methods.
- [Zotto, 2012] Zotto, S. D. (2012). *Ph.D. Thesis. The PLS regression model: algorithms and application to chemometric data*. Universita degli Studi di Udine Dipartimento di Matematica e Informatica Dottorato di Ricerca in Informatica Ciclo xxv, Udine, Italia.

APÉNDICE A

NOTACIÓN

Introducimos nuevamente la notación y el modelo que usaremos de ahora en adelante para facilitar la lectura del apéndice. Suponemos que vale el modelo lineal PLS con $q = 1$; i.e.

$$Y = \mu_Y + \boldsymbol{\beta}^T(\mathbf{X} - \boldsymbol{\mu}_X) + \varepsilon, \quad Y, \varepsilon \in \mathbb{R}, \quad \boldsymbol{\beta}, \mathbf{X} \in \mathbb{R}^{p \times 1}, \quad \varepsilon \perp \mathbf{X} \quad (\text{A.1})$$

con $\varepsilon \sim N(0, \tau^2)$, $\boldsymbol{\beta} = \boldsymbol{\sigma}^T \boldsymbol{\sigma} (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} \boldsymbol{\sigma}$ y

$$\widehat{\boldsymbol{\beta}}_{pls} = (\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}})^{-1} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}. \quad (\text{A.2})$$

Ahora, para un nuevo vector \mathbf{X}_N la predicción es

$$\widehat{Y} = \bar{Y} + \widehat{\boldsymbol{\beta}}_{pls}^T (\mathbf{X}_N - \bar{\mathbf{X}}).$$

Además en lo que sigue usaremos las siguientes notaciones y resultados directos

$$\sigma_Y^2 = \text{var}(Y) = \tau^2 + \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} \quad \tau^2 = \text{var}(\varepsilon) = \text{var}(Y|\mathbf{X}) \quad (\text{A.3})$$

$$\boldsymbol{\ell} = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}}$$

$\boldsymbol{\ell}_0 \in \mathbb{R}^{p \times (p-1)}$ el complemento ortogonal de $\boldsymbol{\ell}$

$$\mathbf{P}_\sigma = \boldsymbol{\ell} \boldsymbol{\ell}^T = \frac{\boldsymbol{\sigma} \boldsymbol{\sigma}^T}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}$$

$$\mathbf{Q}_\sigma = \boldsymbol{\ell}_0 \boldsymbol{\ell}_0^T = \mathbf{I} - \mathbf{P}_\sigma$$

$$\boldsymbol{\Sigma} = \boldsymbol{\ell} \boldsymbol{\Omega} \boldsymbol{\ell}^T + \boldsymbol{\ell}_0 \boldsymbol{\Omega}_0 \boldsymbol{\ell}_0^T$$

$$\boldsymbol{\Omega} = \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}$$

$$\boldsymbol{\Omega}_0 = \boldsymbol{\ell}_0^T \boldsymbol{\Sigma} \boldsymbol{\ell}_0$$

(A.4)

Con esta notación, $\boldsymbol{\beta} = \boldsymbol{\Omega}^{-1} \boldsymbol{\sigma}$. Más aún denotamos

$$C(p, k) = \text{tr}(\mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma)^k, \quad (\text{A.5})$$

$$K = \left(\frac{\Omega}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}} \right)^{1/2} + \frac{C(p,1)}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}}, \quad (\text{A.6})$$

$$J = \left(\frac{\Omega}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}} \right)^{1/2} + \frac{C(p,1)}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}} + \frac{C(p,2)}{n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2}, \quad (\text{A.7})$$

y suponemos que $K = O(1)$, $J = O(1)$.

Vamos a suponer que

$$\frac{C(p,3)}{n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^3} = O(1) \quad \text{y} \quad \frac{C(p,4)}{n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4} = O(1).$$

Observemos que como σ_Y^2 es una constante positiva, τ^2 y $\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} = \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega}$ están acotados superiormente:

$$\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} = \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \leq \sigma_Y^2, \quad \tau^2 \leq \sigma_Y^2 \quad (\text{A.8})$$

Adicionalmente, requerimos que τ^2 esté acotado lejos de 0, así $\tau^2 \asymp 1$. Si además tenemos que $\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}$ está acotado lejos de 0, entonces obtenemos que $\boldsymbol{\sigma}^T \boldsymbol{\sigma} \asymp \Omega$. En este caso K y J serán reemplazados por

$$K = \frac{1}{\sqrt{n}} + \frac{C(p,1)}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}}, \quad (\text{A.9})$$

$$J = \frac{1}{\sqrt{n}} + \frac{C(p,1)}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}} + \frac{C(p,2)}{n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2}. \quad (\text{A.10})$$

Notemos que estos son equivalentes a los definidos en [Cook and Forzani, 2018], ya que

$$\left(\frac{C(p,3)}{n^2(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^3} \right)^{1/2} \leq \left(\frac{C(p,1)}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}} \frac{C(p,2)}{n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2} \right)^{1/2} \leq \frac{1}{\sqrt{2}} \left(\frac{C(p,1)}{n\boldsymbol{\sigma}^T\boldsymbol{\sigma}} + \frac{C(p,2)}{n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2} \right).$$

Adaptando la demostración de [Cook and Forzani, 2018, Proposición 1] a (A.6) y (A.7) tenemos el siguiente resultado

Teorema A.1. *Si $K, J \rightarrow 0$ cuando $n, p \rightarrow \infty$, entonces*

$$\begin{aligned} \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} &= 1 + O_p(K) \\ \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} &= 1 + O_p(K) \\ \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} &= 1 + O_p(J) \\ \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} &= 1 + O_p(J) \\ \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} &= \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} (1 + O_p(J)) \\ \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} &= \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} (1 + O_p(J)). \end{aligned} \quad (\text{A.11})$$

APÉNDICE B

ALGUNOS RESULTADOS CONOCIDOS

Lema B.1. (Momentos de la distribución χ^2) Si $X \sim \chi_m^2$, entonces

$$E(X^k) = m(m+2)(m+4)\dots(m+2(k-1)).$$

Los siguientes momentos pueden encontrarse en [Antunes Percíncula et al., 2020].

Lema B.2. (Momentos de la distribución Wishart) Si $\mathbf{W} \sim \text{Wish}(m, \Sigma)$ para $\Sigma \in \mathbb{R}^{p \times p}$ simétrica definida positiva, entonces

$$E(\mathbf{W}) = m\Sigma$$

$$E(\text{tr}\mathbf{W}) = m\text{tr}\Sigma$$

$$\text{var}(\text{tr}\mathbf{W}) = 2m\text{tr}(\Sigma^2)$$

$$E(\text{tr}\mathbf{W}^2) = m\text{tr}^2(\Sigma) + (m^2 + m)\text{tr}(\Sigma^2)$$

$$\text{var}(\text{tr}\mathbf{W}^2) = (8m^3 + 20m^2 + 20m)\text{tr}(\Sigma^4) + 4(m^2 + m)\text{tr}^2(\Sigma^2)$$

$$+ 8m\text{tr}(\Sigma^2)\text{tr}^2(\Sigma) + 16(m^2 + m)\text{tr}(\Sigma^3)\text{tr}(\Sigma)$$

$$E(\text{tr}\mathbf{W}^4) = m\text{tr}^4(\Sigma) + (m^4 + 6m^3 + 21m^2 + 20m)\text{tr}(\Sigma^4) + (2m^3 + 5m^2 + 5m)\text{tr}^2(\Sigma^2)$$

$$+ 6(m^2 + m)\text{tr}(\Sigma^2)\text{tr}^2(\Sigma) + 4(m^3 + 3m^2 + 4m)\text{tr}(\Sigma^3)\text{tr}(\Sigma)$$

$$E(\mathbf{W}^3) = m\Sigma\text{tr}^2(\Sigma) + (m^2 + m)\left(\Sigma\text{tr}(\Sigma^2) + 2\Sigma^2\text{tr}(\Sigma)\right) + (m^3 + 3m^2 + 4m)\Sigma^3.$$

Lema B.3. Si $\mathbf{W} \sim \text{Wish}(m, \Sigma)$ para $\Sigma \in \mathbb{R}^{p \times p}$ simétrica definida positiva y $\mathbf{a} \in \mathbb{R}^p$, entonces

$$\frac{\mathbf{a}^T \mathbf{W} \mathbf{a}}{\mathbf{a}^T \Sigma \mathbf{a}} \sim \chi_m^2.$$

Lema B.4. Sea $\tilde{\varepsilon} \in \mathbb{R}^n$ que cumple $E(\tilde{\varepsilon}) = 0$, $\text{var}(\tilde{\varepsilon}) = \tau^2 \mathbf{I}_n$ y $E(\tilde{\varepsilon}_i^4) = O(\tau^4)$. Entonces para matrices \mathbf{Q}_n de $n \times n$ y \mathbf{A} , con \mathbf{Q}_n semidefinida positiva y tal que $\mathbf{A}\mathbf{Q}_n = \mathbf{A}$ tenemos que para $\varepsilon = \mathbf{Q}_n \tilde{\varepsilon}$,

$$E(\varepsilon^T \mathbf{A} \varepsilon) = \tau^2 \text{tr}(\mathbf{A}) \quad \text{var}(\varepsilon^T \mathbf{A} \varepsilon) = O(\tau^4 (\text{tr}(\mathbf{A}^2 + \mathbf{A}^T \mathbf{A}))). \quad (\text{B.1})$$

Demostración. Sólo necesitamos calcular la varianza para el caso de \mathbf{A} simétrica, ya que el caso general se deduce considerando $\mathbf{A} + \mathbf{A}^T$. Ahora,

$$\begin{aligned} \text{var}(\varepsilon^T \mathbf{A} \varepsilon) &= \text{var}(\tilde{\varepsilon}^T \mathbf{Q}_n \mathbf{A} \mathbf{Q}_n \tilde{\varepsilon}) \\ &= \text{var}(\text{tr}(\mathbf{Q}_n \mathbf{A} \mathbf{Q}_n \tilde{\varepsilon} \tilde{\varepsilon}^T)) \\ &= \text{vec}^T(\mathbf{Q}_n \mathbf{A} \mathbf{Q}_n) \text{var}(\text{vec} \tilde{\varepsilon} \tilde{\varepsilon}^T) \text{vec}(\mathbf{Q}_n \mathbf{A} \mathbf{Q}_n), \end{aligned}$$

pero $\text{var}(\text{vec} \tilde{\varepsilon} \tilde{\varepsilon}^T)$ es una matriz diagonal cuyos elementos diagonales son de orden τ^4 . De aquí obtenemos el resultado, ya que $\text{tr}(\mathbf{Q}_n \mathbf{A} \mathbf{Q}_n)^2 = \text{tr}(\mathbf{A}^2)$. ■

APÉNDICE C

EL MODELO Y LOS ESTIMADORES

Sin pérdida de generalidad vamos a asumir para las demostraciones que $E(Y) = 0$ y $E(\mathbf{X}) = \mathbf{0}$. De ahora en más consideramos $m = n - 1$. Suponiendo la normalidad de $\mathbf{X}_1, \dots, \mathbf{X}_n$ sea

$$\mathcal{X} = \begin{pmatrix} \mathbf{X}_1^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix} \in \mathbb{R}^{n \times p}, \quad \text{vec}(\mathcal{X}^T) \sim N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma).$$

Se sigue que

$$\begin{aligned} \mathbf{F} &= (\mathbf{X}_1 - \bar{\mathbf{X}}, \dots, \mathbf{X}_n - \bar{\mathbf{X}}) = \mathcal{X}^T \mathbf{Q}_n \in \mathbb{R}^{p \times n}, \\ \mathbf{Q}_n &= \mathbf{I} - \frac{\mathbf{e}\mathbf{e}^T}{n} \in \mathbb{R}^{n \times n} \text{ con } \mathbf{e} = (\mathbf{1}, \dots, \mathbf{1})^T, \\ \text{vec}(\mathbf{F}) &\sim N(\mathbf{0}, \mathbf{Q}_n \otimes \Sigma), \end{aligned}$$

y luego como $\text{rank}(\mathbf{Q}_n) = m$, $\mathbf{W} = \mathbf{F}\mathbf{F}^T \sim \text{Wish}(m, \Sigma)$. Por lo tanto

$$\begin{aligned} \hat{\boldsymbol{\sigma}} &= \frac{1}{m} \sum_{i=1}^n (Y_i - \bar{Y})(\mathbf{X}_i - \bar{\mathbf{X}}) \\ &= \frac{1}{m} \sum_{i=1}^n (\boldsymbol{\beta}^T \mathbf{X}_i + \varepsilon_i - \boldsymbol{\beta}^T \bar{\mathbf{X}} - \bar{\varepsilon})(\mathbf{X}_i - \bar{\mathbf{X}}) \\ &= \frac{1}{m} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\beta} + \frac{1}{m} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})(\mathbf{X}_i - \bar{\mathbf{X}}) \\ &= \frac{1}{m} \mathbf{W}\boldsymbol{\beta} + \frac{1}{m} \mathbf{F}\boldsymbol{\varepsilon} \\ &= \frac{\mathbf{W}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\varepsilon}}{m} \end{aligned} \tag{C.1}$$

donde $\boldsymbol{\varepsilon}$ es un vector con componentes $\varepsilon_i - \bar{\varepsilon}$. Vamos a suponer que $\boldsymbol{\varepsilon}$ satisface $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{var}(\boldsymbol{\varepsilon}) = \tau^2 \mathbf{Q}_n$ y $E(\varepsilon_i^4) = O(\tau^4)$.

Observemos que $\beta^T \Sigma^k \beta = \Omega^{k-2} \sigma^T \sigma$, y

$$\hat{\Sigma} = \frac{1}{m} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T = \frac{1}{m} \mathbf{F} \mathbf{F}^T = \frac{1}{m} \mathbf{W}. \quad (\text{C.2})$$

Lema C.1. Con las definiciones de arriba,

$$\mathbf{F}^T \beta \sim N(\mathbf{0}, \beta^T \Sigma \beta \mathbf{Q}_n), \quad \text{vec}(\mathbf{F}^T \mathbf{Q}_\sigma) \sim N(\mathbf{0}, \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \otimes \mathbf{Q}_n)$$

y $\mathbf{F}^T \beta$ es independiente de $\mathbf{F}^T \mathbf{Q}_\sigma$.

Demostración. Como $\mathbf{F}^T \beta \in \mathbb{R}^n$, $\mathbf{F}^T \beta = \text{vec}(\mathbf{F}^T \beta) = (\beta^T \otimes \mathbf{Q}_n) \text{vec}(\mathcal{X})$. De manera similar, tenemos para $\mathbf{F}^T \mathbf{Q}_\sigma \in \mathbb{R}^{n \times p}$, $\text{vec}(\mathbf{F}^T \mathbf{Q}_\sigma) = (\mathbf{Q}_\sigma \otimes \mathbf{Q}_n) \text{vec}(\mathcal{X})$. Ahora recordemos que $\text{vec}(\mathcal{X}^T) \sim N(\mathbf{0}, \Sigma \otimes \mathbf{I}_n)$ y notemos que

$$\begin{aligned} \begin{pmatrix} \mathbf{F}^T \beta \\ \text{vec}(\mathbf{F}^T \mathbf{Q}_\sigma) \end{pmatrix} &= \begin{pmatrix} (\beta^T \otimes \mathbf{Q}_n) \text{vec}(\mathcal{X}) \\ (\mathbf{Q}_\sigma \otimes \mathbf{Q}_n) \text{vec}(\mathcal{X}) \end{pmatrix} = \begin{pmatrix} \beta^T \otimes \mathbf{Q}_n \\ \mathbf{Q}_\sigma \otimes \mathbf{Q}_n \end{pmatrix} \text{vec}(\mathcal{X}) \\ &\sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \beta^T \Sigma \beta \mathbf{Q}_n & (\beta^T \Sigma \otimes \mathbf{Q}_n)(\mathbf{Q}_\sigma \otimes \mathbf{Q}_n) \\ (\mathbf{Q}_\sigma \otimes \mathbf{Q}_n)(\Sigma \beta \otimes \mathbf{Q}_n) & \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \otimes \mathbf{Q}_n \end{pmatrix} \right) \end{aligned}$$

pero

$$(\beta^T \Sigma \otimes \mathbf{Q}_n)(\mathbf{Q}_\sigma \otimes \mathbf{Q}_n) = \sigma^T \mathbf{Q}_\sigma \otimes \mathbf{Q}_n = \mathbf{0}. \quad \blacksquare$$

Por el resultado anterior,

Corolario C.2.

$$\mathbf{M} = \frac{\mathbf{F}^T \beta \beta^T \mathbf{F}}{\beta^T \Sigma \beta} \sim \text{Wish}(1, \mathbf{Q}_n) \quad (\text{C.3})$$

$$N = \text{tr}(\mathbf{M}) = \frac{\beta^T \mathbf{W} \beta}{\beta^T \Sigma \beta} \sim \chi_m^2 \quad (\text{C.4})$$

$$\mathbf{H} = \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \sim \text{Wish}(m, \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma) \quad \text{y} \quad \mathbf{H} \perp \beta^T \mathbf{F}, \mathbf{M}, N \quad (\text{C.5})$$

$$\mathbf{L} = \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F}, \text{ entonces } \text{tr}(\mathbf{L}^k) = \text{tr}(\mathbf{H}^k) \quad \text{para } k = 1, 2, \dots, \text{ y } \mathbf{L} \perp \beta^T \mathbf{F}, \mathbf{M}, N \quad (\text{C.6})$$

$$\mathbf{T} = \frac{\mathbf{Q}_\sigma \mathbf{W} \beta \beta^T \mathbf{W} \mathbf{Q}_\sigma}{\beta^T \mathbf{W} \beta} \sim \text{Wish}(1, \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma), \quad \text{tr}^2(\mathbf{T}) = \text{tr}(\mathbf{T}^2) \text{ y } \mathbf{T} \perp \beta^T \mathbf{F}, \mathbf{M}, N \quad (\text{C.7})$$

$$\text{tr}((\mathbf{L}\mathbf{M})^k) = \text{tr}^k(\mathbf{T}) N^k, \quad k \in \mathbb{N}. \quad (\text{C.8})$$

Demostración. Sólo demostraremos la distribución de \mathbf{T} , ya que el resto es directo. Por el Lema C.1, $\mathbf{F}^T \mathbf{Q}_\sigma$ es independiente de $\mathbf{F}^T \boldsymbol{\beta}$ y tenemos que

$$\mathbf{Q}_\sigma \mathbf{F} \mathbf{F}^T \boldsymbol{\beta} | \mathbf{F}^T \boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma \boldsymbol{\beta}^T \mathbf{F} \mathbf{F}^T \boldsymbol{\beta}),$$

o

$$\frac{\mathbf{Q}_\sigma \mathbf{F} \mathbf{F}^T \boldsymbol{\beta}}{(\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta})^{1/2}} | \mathbf{F}^T \boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma).$$

Entonces $\mathbf{T} | \mathbf{F}^T \boldsymbol{\beta} \sim \text{Wish}(1, \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma)$, y ya que la distribución de $\mathbf{T} | \mathbf{F}^T \boldsymbol{\beta}$ no depende del valor del argumento condicionante, la distribución sin condicionar es la misma $\mathbf{T} \sim \text{Wish}(1, \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma)$. ■

Con los resultados anteriores podemos mostraron algunas propiedades de $\hat{\boldsymbol{\sigma}}$. Es claro que $E(\hat{\boldsymbol{\sigma}}) = \boldsymbol{\sigma}$, más aún

Lema C.3. $E(\hat{\boldsymbol{\sigma}} \hat{\boldsymbol{\sigma}}^T) = \boldsymbol{\sigma} \boldsymbol{\sigma}^T + \frac{1}{m} (\boldsymbol{\sigma} \boldsymbol{\sigma}^T + \sigma_Y^2 \boldsymbol{\Sigma})$.

Demostración. Recordemos que $\hat{\boldsymbol{\sigma}} = \frac{\mathbf{W} \boldsymbol{\beta} + \mathbf{F} \boldsymbol{\varepsilon}}{m}$ y que $\boldsymbol{\varepsilon}$ es independiente del resto,

$$\begin{aligned} E(\hat{\boldsymbol{\sigma}} \hat{\boldsymbol{\sigma}}^T) &= \frac{1}{m^2} (E(\mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W}) + E(\mathbf{F} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{F})) \\ &= \frac{1}{m^2} (E(\mathbf{P}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{P}_\sigma) + E(\mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma) + E(\mathbf{F} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{F})), \end{aligned}$$

donde la última igualdad se sigue del hecho de que $\mathbf{F}^T \boldsymbol{\beta} \perp \mathbf{F}^T \mathbf{Q}_\sigma$ por Lema C.1. Ahora por Corolario C.2 y el hecho de que $\text{var}(\boldsymbol{\varepsilon}) = \tau^2 \mathbf{Q}_n$,

$$\begin{aligned} E(\hat{\boldsymbol{\sigma}} \hat{\boldsymbol{\sigma}}^T) &= \frac{1}{m^2} (\boldsymbol{\sigma} \boldsymbol{\sigma}^T E(N^2) + \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} E(N) E(\mathbf{T}) + \tau^2 E(\mathbf{W})) \\ &= \frac{1}{m^2} (\boldsymbol{\sigma} \boldsymbol{\sigma}^T m(m+2) + \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} m \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma + \tau^2 m \boldsymbol{\Sigma}) \\ &= \boldsymbol{\sigma} \boldsymbol{\sigma}^T + \frac{1}{m} (\boldsymbol{\sigma} \boldsymbol{\sigma}^T + \sigma_Y^2 \boldsymbol{\Sigma}), \end{aligned}$$

donde usamos Lema B.1 y Lema B.2, junto con $\sigma_Y^2 = \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} + \tau^2$. ■

Está claro que

$$\text{var}(\hat{\boldsymbol{\sigma}}) = \frac{1}{m} (\boldsymbol{\Sigma} \sigma_Y^2 + \boldsymbol{\sigma} \boldsymbol{\sigma}^T) \tag{C.9}$$

$$\begin{aligned} \text{var} \left(\frac{(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) &= \frac{\sigma_Y^2 \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma} + (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2}{m (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \\ &= \frac{\sigma_Y^2 \Omega}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{1}{m} \\ &= O \left(\frac{\Omega}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right), \end{aligned} \tag{C.10}$$

donde el orden se sigue del hecho de que $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$ por (A.8). Adicionalmente, si $\Omega \asymp \boldsymbol{\sigma}^T \boldsymbol{\sigma}$, tenemos que $\text{var} \left(\frac{(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) = O(1/m)$.

Ahora, recordemos que $\boldsymbol{\varepsilon} \perp \mathbf{F}$, y por lo tanto por el Lema B.4

Lema C.4. *Tenemos que $\boldsymbol{\varepsilon}$ satisface $E(\boldsymbol{\varepsilon}) = 0$, $\text{var}(\boldsymbol{\varepsilon}) = \tau^2 \mathbf{Q}_n$, $E(\boldsymbol{\varepsilon}_i^4) = O(\tau^4)$ y $\boldsymbol{\varepsilon} \perp \mathbf{M}, \mathbf{N}, \mathbf{H}, \mathbf{L}, \mathbf{T}$. Además, si $\mathbf{A} \in \mathbb{R}^{n \times n}$ es cualquier función de $\mathbf{M}, \mathbf{N}, \mathbf{H}, \mathbf{L}$ o \mathbf{T} tal que $\mathbf{A} \mathbf{Q}_n = \mathbf{A}$, entonces tenemos que*

$$E(\boldsymbol{\varepsilon}^T \mathbf{A} \boldsymbol{\varepsilon}) = \tau^2 E(\text{tr}(\mathbf{A})), \quad \text{var}(\boldsymbol{\varepsilon}^T \mathbf{A} \boldsymbol{\varepsilon}) = O(\tau^4 E(\text{tr}(\mathbf{A}^2 + \mathbf{A}^T \mathbf{A}))). \quad (\text{C.11})$$

Como consecuencia directa del Corolario C.2 y del Lema B.2, tenemos que

Lema C.5. *Recordando que $\mathbf{H} \sim \text{Wish}(m, \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma)$ y $\mathbf{T} \sim \text{Wish}(1, \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma)$, y también que $\text{tr}((\mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma)^k) = C(p, k)$, tenemos*

$$E(\text{tr} \mathbf{H}) = mC(p, 1)$$

$$E(\text{tr}^2 \mathbf{H}) = m^2 C^2(p, 1) + 2mC(p, 2)$$

$$E(\text{tr} \mathbf{H}^2) = mC^2(p, 1) + (m^2 + m)C(p, 2)$$

$$\text{var}(\text{tr} \mathbf{H}) = 2mC(p, 2)$$

$$E(\text{tr} \mathbf{T}) = C(p, 1)$$

$$E(\text{tr}^2 \mathbf{T}) = E(\text{tr} \mathbf{T}^2) = C^2(p, 1) + 2C(p, 2)$$

$$\text{var}(\text{tr} \mathbf{T}) = 2C(p, 2).$$

Más aún por Lema B.2 y Lema C.5

Corolario C.6.

$$\frac{E(\text{tr} \mathbf{T}^2)}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} = O(J)$$

$$\frac{\text{var}(\text{tr} \mathbf{T})}{m (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} = O(J)$$

$$\frac{E(\text{tr} \mathbf{H}^2)}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} = O(J)$$

$$\frac{\text{var}(\text{tr} \mathbf{H})}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} = O(J)$$

$$\frac{\text{var}(\text{tr} \mathbf{H}^2)}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} = O \left(J + \frac{C(p, 4)}{m (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} \right)$$

$$\frac{E(\text{tr} \mathbf{H}^4)}{m^5 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} = O \left(J + \frac{C(p, 4)}{m (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} \right).$$

Demostración. Las primeras cuatro igualdades son una consecuencia directa de Lema C.5 y la definición de J . Para la quinta, usando (C.5) y Lema B.2 tenemos

$$\frac{\text{var}(\text{tr}\mathbf{H}^2)}{m^4(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4} = O\left(\frac{m^3C(p,4) + m^2C^2(p,2) + mC(p,2)C^2(p,1) + m^2C(p,3)C(p,1)}{m^4(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4}\right),$$

y como $\frac{C(p,k)}{m(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^k}$ están acotados para $k = 1, 2, 3, 4$, tenemos el orden.

Para la última igualdad, nuevamente usando (C.5) y Lema B.2 se sigue que

$$\begin{aligned} \frac{E(\text{tr}\mathbf{H}^4)}{m^5(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4} &= O\left(\frac{mC^4(p,1) + m^4C(p,4) + m^3C^2(p,2)}{m^5(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4}\right. \\ &\quad \left.+ \frac{m^2C(p,2)C^2(p,1) + m^3C(p,3)C(p,1)}{m^5(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4}\right), \end{aligned}$$

y como $\frac{C(p,k)}{m(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^k}$ están acotados para $k = 1, 2, 3, 4$, tenemos el orden. ■

APÉNDICE D

DEMOSTRACIÓN DE LA PROPOSICIÓN 3.1

Recordamos las definiciones

$$\mathbf{R} = \Omega^{-1} \left((\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) - \Omega^{-1} \mathbf{P}_{\boldsymbol{\sigma}} (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \boldsymbol{\sigma} \right), \quad (\text{D.1})$$

$$\begin{aligned} \boldsymbol{\Delta} &= (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T (\hat{\boldsymbol{\sigma}} + \boldsymbol{\sigma}) (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) + \frac{(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \left(\frac{\hat{\boldsymbol{\sigma}}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} - \frac{\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right) \boldsymbol{\sigma} \\ &\quad - \left(\frac{2}{m} + \frac{C(p,1)}{m^2 \Omega} + \sigma_Y^2 \frac{C(p,1)}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{\sigma_Y^2 C(p,2)}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \boldsymbol{\beta}, \end{aligned} \quad (\text{D.2})$$

$$\begin{aligned} U &= (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \left[\frac{\mathbf{Q}_{\boldsymbol{\sigma}} (\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\sigma}} \Omega^{-1}) + \boldsymbol{\sigma} - \hat{\boldsymbol{\Sigma}} \boldsymbol{\sigma} \Omega^{-1}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right] \\ &\quad + \left(\frac{2}{m} + \frac{C(p,1)}{m^2 \Omega} + \sigma_Y^2 \frac{C(p,1)}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{\sigma_Y^2 C(p,2)}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right). \end{aligned} \quad (\text{D.3})$$

Entonces tenemos

$$\begin{aligned} \frac{\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \left(\hat{\boldsymbol{\beta}}_{pls}^T \mathbf{G} - \boldsymbol{\beta}^T \mathbf{G} \right) &= \left[(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} (\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\sigma}}) \hat{\boldsymbol{\sigma}}^T \mathbf{G} - \frac{\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\sigma}}}{(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^2} (\boldsymbol{\sigma}^T \boldsymbol{\sigma}) \boldsymbol{\sigma}^T \mathbf{G} \right] \\ &= [(\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\sigma}}) \hat{\boldsymbol{\sigma}}^T \mathbf{G} - (\boldsymbol{\sigma}^T \boldsymbol{\sigma}) \boldsymbol{\sigma}^T \mathbf{G}] (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} \\ &\quad + \left[(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} - \frac{\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\sigma}}}{(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^2} \right] (\boldsymbol{\sigma}^T \boldsymbol{\sigma}) \boldsymbol{\sigma}^T \mathbf{G} \\ &= T_1 + T_2. \end{aligned}$$

Primero para T_1 , mediante manipulaciones algebraicas

$$\begin{aligned} T_1 &= [(\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\sigma}}) \hat{\boldsymbol{\sigma}}^T \mathbf{G} - (\boldsymbol{\sigma}^T \boldsymbol{\sigma}) \boldsymbol{\sigma}^T \mathbf{G}] (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} \\ &= [(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \mathbf{G} + (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) \boldsymbol{\sigma}^T \mathbf{G} + (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\sigma} (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \mathbf{G} \\ &\quad + (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\sigma} \boldsymbol{\sigma}^T \mathbf{G} + \boldsymbol{\sigma}^T (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \mathbf{G} \end{aligned}$$

$$\begin{aligned}
& + \sigma^T (\hat{\sigma} - \sigma) \sigma^T \mathbf{G} + \sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \Big] (\sigma^T \Sigma \sigma)^{-1} \\
& = \left[(\hat{\sigma} - \sigma)^T (\hat{\sigma} - \sigma) (\hat{\sigma} - \sigma)^T \mathbf{G} + (\hat{\sigma} - \sigma)^T (\hat{\sigma} - \sigma) \sigma^T \mathbf{G} + 2(\hat{\sigma} - \sigma)^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \right. \\
& \quad \left. + 2(\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} + \sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \right] (\sigma^T \Sigma \sigma)^{-1} \\
& = \left[(\hat{\sigma} - \sigma)^T (\hat{\sigma} + \sigma) (\hat{\sigma} - \sigma)^T \mathbf{G} + (\hat{\sigma} - \sigma)^T (\hat{\sigma} - \sigma) \sigma^T \mathbf{G} \right. \\
& \quad \left. + 2(\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} + \sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \right] (\sigma^T \Sigma \sigma)^{-1} \\
& = \frac{1}{\sigma^T \Sigma \sigma} \left[2(\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} + \sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \right] \\
& \quad + \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T (\hat{\sigma} - \sigma) \sigma^T \mathbf{G} + \frac{1}{\sigma^T \Sigma \sigma} \left[(\hat{\sigma} - \sigma)^T (\hat{\sigma} + \sigma) (\hat{\sigma} - \sigma)^T \mathbf{G} \right].
\end{aligned}$$

Ahora para T_2 , nuevamente mediante manipulaciones algebraicas

$$\begin{aligned}
T_2 & = \left[(\sigma^T \Sigma \sigma)^{-1} - \frac{\hat{\sigma}^T \hat{\Sigma} \hat{\sigma}}{(\sigma^T \Sigma \sigma)^2} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& = \frac{-1}{\sigma^T \Sigma \sigma} \left[\frac{\hat{\sigma}^T \hat{\Sigma} \hat{\sigma}}{\sigma^T \Sigma \sigma} - 1 \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& = \frac{-1}{\sigma^T \Sigma \sigma} \left[\frac{(\hat{\sigma} - \sigma)^T (\hat{\Sigma} - \Sigma) (\hat{\sigma} - \sigma)}{\sigma^T \Sigma \sigma} + \frac{(\hat{\sigma} - \sigma)^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} + \frac{(\hat{\sigma} - \sigma)^T \Sigma (\hat{\sigma} - \sigma)}{\sigma^T \Sigma \sigma} \right. \\
& \quad \left. + \frac{(\hat{\sigma} - \sigma)^T \Sigma \sigma}{\sigma^T \Sigma \sigma} + \frac{\sigma^T (\hat{\Sigma} - \Sigma) (\hat{\sigma} - \sigma)}{\sigma^T \Sigma \sigma} + \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} + \frac{\sigma^T \Sigma (\hat{\sigma} - \sigma)}{\sigma^T \Sigma \sigma} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& = \frac{-1}{\sigma^T \Sigma \sigma} \left[2 \frac{(\hat{\sigma} - \sigma)^T \Sigma \sigma}{\sigma^T \Sigma \sigma} + \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& \quad + \frac{-1}{\sigma^T \Sigma \sigma} \left[\frac{(\hat{\sigma} - \sigma)^T (\hat{\Sigma} - \Sigma) (\hat{\sigma} - \sigma)}{\sigma^T \Sigma \sigma} \right. \\
& \quad \quad \left. + 2 \frac{(\hat{\sigma} - \sigma)^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} + \frac{(\hat{\sigma} - \sigma)^T \Sigma (\hat{\sigma} - \sigma)}{\sigma^T \Sigma \sigma} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& = \frac{-1}{\sigma^T \Sigma \sigma} \left[2 \frac{(\hat{\sigma} - \sigma)^T \Sigma \sigma}{\sigma^T \Sigma \sigma} + \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& \quad + \frac{-1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[\frac{\hat{\Sigma} (\hat{\sigma} + \sigma) - 2 \Sigma \sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \right] \sigma^T \mathbf{G}.
\end{aligned}$$

Usando las expansiones de T_1 y T_2 , y arreglando algunos términos, tomando $(\hat{\sigma} - \sigma)^T$ como factor en el tercer y sexto término, tenemos

$$\begin{aligned}
\frac{\hat{\sigma}^T \hat{\Sigma} \hat{\sigma}}{\sigma^T \Sigma \sigma} \left(\hat{\beta}_{pls}^T \mathbf{G} - \beta^T \mathbf{G} \right) & = \frac{1}{\sigma^T \Sigma \sigma} \left[2(\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} + \sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \right] \\
& \quad + \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T (\hat{\sigma} - \sigma) \sigma^T \mathbf{G} \\
& \quad + \frac{1}{\sigma^T \Sigma \sigma} \left[(\hat{\sigma} - \sigma)^T (\hat{\sigma} + \sigma) (\hat{\sigma} - \sigma)^T \mathbf{G} \right]
\end{aligned}$$

$$\begin{aligned}
& + \frac{-1}{\sigma^T \Sigma \sigma} \left[2 \frac{(\hat{\sigma} - \sigma)^T \Sigma \sigma}{\sigma^T \Sigma \sigma} + \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& + \frac{-1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[\frac{\hat{\Sigma} (\hat{\sigma} + \sigma) - 2 \Sigma \sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \right] \sigma^T \mathbf{G} \\
& = \frac{1}{\sigma^T \Sigma \sigma} \left[2 (\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} + \sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \right] \\
& + \frac{-1}{\sigma^T \Sigma \sigma} \left[2 \frac{(\hat{\sigma} - \sigma)^T \Sigma \sigma}{\sigma^T \Sigma \sigma} + \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& + \frac{1}{\sigma^T \Sigma \sigma} \left[(\hat{\sigma} - \sigma)^T (\hat{\sigma} + \sigma) (\hat{\sigma} - \sigma)^T \mathbf{G} \right] \\
& + \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[(\hat{\sigma} - \sigma) - \frac{\hat{\Sigma} (\hat{\sigma} + \sigma) - 2 \Sigma \sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \right] \sigma^T \mathbf{G}.
\end{aligned} \tag{D.4}$$

Como $\Sigma \sigma = \Omega \sigma$, $\mathbf{P}_\sigma = \frac{\sigma \sigma^T}{\sigma^T \sigma}$ y $\sigma^T \Sigma \sigma = \Omega \sigma^T \sigma$, y por la definición de \mathbf{R} (D.1), los dos primeros sumandos de (D.4) son

$$\begin{aligned}
& \frac{1}{\sigma^T \Sigma \sigma} \left[2 (\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} + \sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} \right] \\
& + \frac{-1}{\sigma^T \Sigma \sigma} \left[2 \frac{(\hat{\sigma} - \sigma)^T \Sigma \sigma}{\sigma^T \Sigma \sigma} + \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} \right] (\sigma^T \sigma) \sigma^T \mathbf{G} \\
& = \frac{1}{\sigma^T \Sigma \sigma} \left[2 (\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} - 2 \frac{(\hat{\sigma} - \sigma)^T \Sigma \sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \sigma^T \mathbf{G} \right] \\
& + \frac{1}{\sigma^T \Sigma \sigma} \left[\sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} - \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \sigma^T \mathbf{G} \right] \\
& = \frac{1}{\Omega \sigma^T \sigma} \left[2 (\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} - 2 \frac{(\hat{\sigma} - \sigma)^T \Omega \sigma}{\Omega \sigma^T \sigma} (\sigma^T \sigma) \sigma^T \mathbf{G} \right] \\
& + \frac{1}{\Omega \sigma^T \sigma} \left[\sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} - \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\Omega \sigma^T \sigma} (\sigma^T \sigma) \sigma^T \mathbf{G} \right] \\
& = \frac{1}{\Omega \sigma^T \sigma} \left[2 (\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} - 2 (\hat{\sigma} - \sigma)^T \sigma \sigma^T \mathbf{G} \right] \\
& + \frac{1}{\Omega \sigma^T \sigma} \left[\sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} - \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\Omega \sigma^T \sigma} (\sigma^T \sigma) \sigma^T \mathbf{G} \right] \\
& = \frac{1}{\Omega \sigma^T \sigma} \left[\sigma^T \sigma (\hat{\sigma} - \sigma)^T \mathbf{G} - \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\Omega \sigma^T \sigma} (\sigma^T \sigma) \sigma^T \mathbf{G} \right] \\
& = \Omega^{-1} \left[(\hat{\sigma} - \sigma) - \frac{\sigma^T (\hat{\Sigma} - \Sigma) \sigma}{\Omega \sigma^T \sigma} \sigma \right]^T \mathbf{G} \\
& = \Omega^{-1} \left[(\hat{\sigma} - \sigma) - \Omega^{-1} \mathbf{P}_\sigma (\hat{\Sigma} - \Sigma) \sigma \right]^T \mathbf{G} \\
& = \mathbf{R}^T \mathbf{G}.
\end{aligned}$$

Insertando en (D.4), sumando y restando $\frac{(\hat{\sigma} - \sigma)^T \sigma}{\sigma^T \Sigma \sigma} \left(\frac{\hat{\sigma}^T \sigma}{\sigma^T \sigma} - \frac{\hat{\sigma}^T \hat{\Sigma} \sigma}{\sigma^T \Sigma \sigma} \right) \sigma^T \mathbf{G}$, y por la definición de Δ (D.2)

$$\begin{aligned}
\frac{\hat{\sigma}^T \hat{\Sigma} \hat{\sigma}}{\sigma^T \Sigma \sigma} \left(\hat{\beta}_{pls}^T \mathbf{G} - \beta^T \mathbf{G} \right) &= \mathbf{R}^T \mathbf{G} + \frac{1}{\sigma^T \Sigma \sigma} \left[(\hat{\sigma} - \sigma)^T (\hat{\sigma} + \sigma) (\hat{\sigma} - \sigma)^T \mathbf{G} \right] \\
&\quad + \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[(\hat{\sigma} - \sigma) - \frac{\hat{\Sigma}(\hat{\sigma} + \sigma) - 2\Sigma\sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \right] \sigma^T \mathbf{G} \\
&= \mathbf{R}^T \mathbf{G} + \frac{1}{\sigma^T \Sigma \sigma} \left[(\hat{\sigma} - \sigma)^T (\hat{\sigma} + \sigma) (\hat{\sigma} - \sigma)^T \mathbf{G} \right] \\
&\quad + \frac{(\hat{\sigma} - \sigma)^T \sigma}{\sigma^T \Sigma \sigma} \left(\frac{\hat{\sigma}^T \sigma}{\sigma^T \sigma} - \frac{\hat{\sigma}^T \hat{\Sigma} \sigma}{\sigma^T \Sigma \sigma} \right) \sigma^T \mathbf{G} \\
&\quad + \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[(\hat{\sigma} - \sigma) - \frac{\hat{\Sigma}(\hat{\sigma} + \sigma) - 2\Sigma\sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \right] \sigma^T \mathbf{G} \\
&\quad - \frac{(\hat{\sigma} - \sigma)^T \sigma}{\sigma^T \Sigma \sigma} \left(\frac{\hat{\sigma}^T \sigma}{\sigma^T \sigma} - \frac{\hat{\sigma}^T \hat{\Sigma} \sigma}{\sigma^T \Sigma \sigma} \right) \sigma^T \mathbf{G} \\
&= \mathbf{R}^T \mathbf{G} + \Delta^T \mathbf{G} \\
&\quad + \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[(\hat{\sigma} - \sigma) - \frac{\hat{\Sigma}(\hat{\sigma} + \sigma) - 2\Sigma\sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \right] \sigma^T \mathbf{G} \\
&\quad - \frac{(\hat{\sigma} - \sigma)^T \sigma}{\sigma^T \Sigma \sigma} \left(\frac{\hat{\sigma}^T \sigma}{\sigma^T \sigma} - \frac{\hat{\sigma}^T \hat{\Sigma} \sigma}{\sigma^T \Sigma \sigma} \right) \sigma^T \mathbf{G} \\
&\quad + \left(\frac{2}{n} + \frac{C(p, 1)}{n^2 \Omega} + \frac{\sigma_Y^2 C(p, 2)}{n^2 \Omega \sigma^T \sigma} \right) \Omega^{-1} \sigma^T \mathbf{G}.
\end{aligned}$$

Ahora, nuevamente usando el hecho de que $\Sigma\sigma = \Omega\sigma$ y $\mathbf{P}_\sigma = \frac{\sigma\sigma^T}{\sigma^T\sigma}$, $\Omega = \frac{\sigma^T \Sigma \sigma}{\sigma^T \sigma}$, $\hat{\sigma}^T \sigma = \sigma^T \hat{\sigma}$, $\hat{\sigma}^T \hat{\Sigma} \sigma = \sigma^T \hat{\Sigma} \hat{\sigma}$,

$$\begin{aligned}
&\frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[(\hat{\sigma} - \sigma) - \frac{\hat{\Sigma}(\hat{\sigma} + \sigma) - 2\Sigma\sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) \right] \sigma^T \mathbf{G} \\
&\quad - \frac{(\hat{\sigma} - \sigma)^T \sigma}{\sigma^T \Sigma \sigma} \left(\frac{\hat{\sigma}^T \sigma}{\sigma^T \sigma} - \frac{\hat{\sigma}^T \hat{\Sigma} \sigma}{\sigma^T \Sigma \sigma} \right) \sigma^T \mathbf{G} \\
&= \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[(\hat{\sigma} - \sigma) - \frac{\hat{\Sigma}(\hat{\sigma} + \sigma) - 2\Sigma\sigma}{\sigma^T \Sigma \sigma} (\sigma^T \sigma) - \sigma \frac{\hat{\sigma}^T \sigma}{\sigma^T \sigma} + \sigma \frac{\hat{\sigma}^T \hat{\Sigma} \sigma}{\sigma^T \Sigma \sigma} \right] \sigma^T \mathbf{G} \\
&= \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[(\hat{\sigma} - \sigma) - \hat{\Sigma} \hat{\sigma} \Omega^{-1} - \hat{\Sigma} \sigma \Omega^{-1} + 2\sigma - \mathbf{P}_\sigma \hat{\sigma} + \mathbf{P}_\sigma \hat{\Sigma} \hat{\sigma} \Omega^{-1} \right] \sigma^T \mathbf{G} \\
&= \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[\hat{\sigma} - \hat{\Sigma} \hat{\sigma} \Omega^{-1} + \sigma - \hat{\Sigma} \sigma \Omega^{-1} - \mathbf{P}_\sigma (\hat{\sigma} - \hat{\Sigma} \hat{\sigma} \Omega^{-1}) \right] \sigma^T \mathbf{G} \\
&= \frac{1}{\sigma^T \Sigma \sigma} (\hat{\sigma} - \sigma)^T \left[\mathbf{Q}_\sigma (\hat{\sigma} - \hat{\Sigma} \hat{\sigma} \Omega^{-1}) + \sigma - \hat{\Sigma} \sigma \Omega^{-1} \right] \sigma^T \mathbf{G},
\end{aligned}$$

así que por la definición de U (D.3),

$$\frac{\hat{\sigma}^T \hat{\Sigma} \hat{\sigma}}{\sigma^T \Sigma \sigma} \left(\hat{\beta}_{pls}^T \mathbf{G} - \beta^T \mathbf{G} \right) = \mathbf{R}^T \mathbf{G} + \Delta^T \mathbf{G} + U \beta^T \mathbf{G}.$$

APÉNDICE E

DEMOSTRACIÓN DE LA PROPOSICIÓN 3.2 Y DEL TEOREMA 3.3

Con la descomposición dada en la anterior sección, la demostración del Teorema 3.3 se divide en los siguientes lemas.

Lema E.1. *Bajo las condiciones (a) y (b) $V^{-1/2}\mathbf{R}^T\mathbf{G} \rightsquigarrow N(0, 1)$.*

Lema E.2. *Bajo las condiciones (a), $E(U) = b$.*

Lema E.3. *Bajo las condiciones (a) – (d),*

$$V^{-1/2}(U - E(U))\beta^T\mathbf{G} = O_p\left(J^{1/2} + \left(\frac{C(p, 4)}{n(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4}\right)^{1/2}\right).$$

Lema E.4. *Bajo las condiciones (a) – (c), $V^{-1/2}\boldsymbol{\Delta}^T\mathbf{G} = O_p(J)$.*

E.1. Demostración de la Proposición 3.2

1. es la Ecuación (A.11) (ver también [Cook and Forzani, 2018]).
2. se sigue de la definición de J y b , y el hecho de que $\frac{\boldsymbol{\sigma}^T\boldsymbol{\sigma}}{\Omega} = O(1)$, $\tau^2 = O(1)$, $\sigma_Y^2 = O(1)$,
 $J = O(1)$.
3. es el Lema E.2.
4. es el Lema E.1.
5. es el Lema E.4.
6. es el Lema E.3.

E.1.1. Demostración del Teorema 3.3

Por la Proposición 3.1 tenemos que

$$V^{-1/2} \left(\widehat{\beta}_{pls} - \beta \right)^T \mathbf{G} = \left(\frac{\widehat{\sigma}^T \widehat{\Sigma} \widehat{\sigma}}{\sigma^T \Sigma \sigma} \right)^{-1} \left\{ V^{-1/2} \mathbf{R}^T \mathbf{G} + V^{-1/2} \Delta^T \mathbf{G} \right. \\ \left. + V^{-1/2} (U - E(U)) \beta^T \mathbf{G} + V^{-1/2} E(U) \beta^T \mathbf{G} \right\},$$

y usando que $E(U) = b$,

$$V^{-1/2} \left(\widehat{\beta}_{pls} - (1+b)\beta \right)^T \mathbf{G} = \left(\frac{\widehat{\sigma}^T \widehat{\Sigma} \widehat{\sigma}}{\sigma^T \Sigma \sigma} \right)^{-1} \left\{ V^{-1/2} \mathbf{R}^T \mathbf{G} + V^{-1/2} \Delta^T \mathbf{G} \right. \\ \left. + V^{-1/2} (U - E(U)) \beta^T \mathbf{G} \right\} \\ + V^{-1/2} \left(\left(\frac{\widehat{\sigma}^T \widehat{\Sigma} \widehat{\sigma}}{\sigma^T \Sigma \sigma} \right)^{-1} - 1 \right) b \beta^T \mathbf{G}.$$

Queremos ver que

$$V^{-1/2} \left(\widehat{\beta}_{pls} - (1+b)\beta \right)^T \mathbf{G} = V^{-1/2} \mathbf{R}^T \mathbf{G} + O_p \left(J^{1/2} + \left(\frac{C(p,4)}{n(\sigma^T \sigma)^4} \right)^{1/2} \right).$$

Por Proposición 3.2 para obtener este resultado vamos a necesitar que

$$\left(\left(\frac{\widehat{\sigma}^T \widehat{\Sigma} \widehat{\sigma}}{\sigma^T \Sigma \sigma} \right)^{-1} - 1 \right) V^{-1/2} |\beta^T \mathbf{G} b| \rightarrow 0,$$

y por (A.11) esto es equivalente a

$$JV^{-1/2} |\beta^T \mathbf{G} b| \rightarrow 0.$$

Esto prueba la primera parte del Teorema 3.3. Para la segunda parte, observemos que

$$V^{-1/2} |\beta^T \mathbf{G}| = O(\sqrt{n}) \text{ por (3.11) y}$$

$$JV^{-1/2} |\beta^T \mathbf{G} b| = O(\sqrt{n}) J |b| \\ = O(\sqrt{n}) \left(\left(\frac{\Omega}{n \sigma^T \sigma} \right)^{1/2} + \frac{C(p,1)}{n \sigma^T \sigma} + \frac{C(p,2)}{n (\sigma^T \sigma)^2} \right) \\ \times \left| - \left(\frac{\sigma^T \sigma}{\Omega} - \tau^2 \right) \frac{C(p,1)}{m \sigma^T \sigma} - \sigma_Y^2 \frac{C^2(p,1)}{m^2 \Omega \sigma^T \sigma} - \sigma_Y^2 \frac{C(p,2)}{m \Omega \sigma^T \sigma} \right| \\ = O \left(\left(\frac{\Omega}{n^{1/2} \sigma^T \sigma} \right)^{1/2} \left(\frac{C(p,1)}{n^{3/4} \sigma^T \sigma} + \frac{C(p,2)}{n^{3/4} (\sigma^T \sigma)^2} \right) \right. \\ \left. + \frac{C^2(p,1)}{n^{3/2} (\sigma^T \sigma)^2} + \frac{C^2(p,2)}{n^{3/2} (\sigma^T \sigma)^4} \right),$$

donde hemos usado el hecho de que $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$, $\tau^2 = O(1)$, $\text{var}(Y) = O(1)$, $J = O(1)$. La parte (II) se sigue entonces por hipótesis.

E.1.2. Demostración del Lema E.1

Observemos primero que por (C.1), y el hecho de que $\boldsymbol{\Sigma}\boldsymbol{\sigma} = \Omega\boldsymbol{\sigma}$, $\mathbf{I}_p = \mathbf{P}_\sigma + \mathbf{Q}_\sigma$, $\boldsymbol{\beta} = \Omega^{-1}\boldsymbol{\sigma}$,

$$\begin{aligned} \mathbf{R} &= \Omega^{-1} \left((\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) - \Omega^{-1} \mathbf{P}_\sigma (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \boldsymbol{\sigma} \right) \\ &= \Omega^{-1} \left(\hat{\boldsymbol{\sigma}} - \Omega^{-1} \mathbf{P}_\sigma \hat{\boldsymbol{\Sigma}} \boldsymbol{\sigma} \right) \\ &= \frac{\Omega^{-1}}{m} \sum_{i=1}^n \left((\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\beta} + (\mathbf{X}_i - \bar{\mathbf{X}})(\boldsymbol{\varepsilon}_i - \bar{\boldsymbol{\varepsilon}}) \right. \\ &\quad \left. - \Omega^{-1} \mathbf{P}_\sigma (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\sigma} \right) \\ &= \frac{\Omega^{-1}}{m} \sum_{i=1}^n \left(\mathbf{Q}_\sigma (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \boldsymbol{\beta} + (\mathbf{X}_i - \bar{\mathbf{X}})(\boldsymbol{\varepsilon}_i - \bar{\boldsymbol{\varepsilon}}) \right) \\ &= \left(\frac{\Omega^{-1}}{m} \sum_{i=1}^n \mathbf{Q}_\sigma \mathbf{X}_i \mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{X}_i \boldsymbol{\varepsilon}_i \right) - \frac{n}{m} \Omega^{-1} (\mathbf{Q}_\sigma \bar{\mathbf{X}} \bar{\mathbf{X}}^T \boldsymbol{\beta} + \bar{\mathbf{X}} \bar{\boldsymbol{\varepsilon}}) \\ &= \frac{n}{m} \tilde{\mathbf{R}} - \frac{n}{m} \Omega^{-1} (\mathbf{Q}_\sigma \bar{\mathbf{X}} \bar{\mathbf{X}}^T \boldsymbol{\beta} + \bar{\mathbf{X}} \bar{\boldsymbol{\varepsilon}}), \end{aligned}$$

con $\tilde{\mathbf{R}} = \frac{\Omega^{-1}}{n} \sum_{i=1}^n \mathbf{Q}_\sigma \mathbf{X}_i \mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{X}_i \boldsymbol{\varepsilon}_i$.

Primero demostraremos que

$$V^{-1/2} \tilde{\mathbf{R}}^T \mathbf{G} \rightsquigarrow N(0, 1),$$

donde $V = V(\mathbf{G}) = \Omega^{-2} n^{-1} \mathbf{G}^T [\boldsymbol{\Sigma} \tau^2 + (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma] \mathbf{G}$ como en (3.5). Luego mostraremos que $V^{-1/2} \Omega^{-1} (\mathbf{Q}_\sigma \bar{\mathbf{X}} \bar{\mathbf{X}}^T \boldsymbol{\beta} + \bar{\mathbf{X}} \bar{\boldsymbol{\varepsilon}})^T \mathbf{G} = O_p(n^{-1/2})$ y esto concluye la demostración del Lema E.1.

Para probar la normalidad asintótica de $\tilde{\mathbf{R}}^T \mathbf{G}$, recordemos que asumimos $E(\mathbf{X}) = \mathbf{0}$ y definamos

$$\begin{aligned} Z_i &= \Omega^{-1} (\mathbf{Q}_\sigma \mathbf{X}_i \mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{X}_i \boldsymbol{\varepsilon}_i)^T \mathbf{G} \\ &= \Omega^{-1} (\mathbf{X}_i^T \boldsymbol{\beta} \mathbf{X}_i^T \mathbf{Q}_\sigma \mathbf{G} + \boldsymbol{\varepsilon}_i \mathbf{X}_i^T \mathbf{G}). \end{aligned}$$

Notemos que $\mathbf{X}_i \perp \boldsymbol{\varepsilon}_i$ y $\mathbf{X}_i^T \boldsymbol{\beta} \perp \mathbf{X}_i^T \mathbf{Q}_\sigma$, más aún $\mathbf{X}_i^T \boldsymbol{\beta} \sim N(0, \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})$, $\mathbf{X}_i^T \mathbf{G} \sim N(0, \mathbf{G}^T \boldsymbol{\Sigma} \mathbf{G})$, $\mathbf{X}_i^T \mathbf{Q}_\sigma \mathbf{G} \sim N(0, \mathbf{G}^T \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma \mathbf{G})$, y $\boldsymbol{\varepsilon}$ satisface $E(\boldsymbol{\varepsilon}) = 0$, $\text{var}(\boldsymbol{\varepsilon}_i) = \tau^2$ y $E(\boldsymbol{\varepsilon}_i^4) = O(\tau^4)$. Por lo tanto

$$\begin{aligned}
E(Z_i) &= 0, \\
E(Z_i^2) &= \Omega^{-2}(\text{var}(\mathbf{X}_i^T \boldsymbol{\beta}) \text{var}(\mathbf{X}_i^T \mathbf{Q}_\sigma \mathbf{G}) + \text{var}(\varepsilon_i) \text{var}(\mathbf{X}_i^T \mathbf{G})) \\
&= \Omega^{-2}((\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})(\mathbf{G}^T \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma \mathbf{G}) + \tau^2(\mathbf{G}^T \boldsymbol{\Sigma} \mathbf{G})) \\
&= nV, \\
E(Z_i^4) &\leq 8\Omega^{-4} (E((\mathbf{X}_i^T \boldsymbol{\beta})^4)E((\mathbf{X}_i^T \mathbf{Q}_\sigma \mathbf{G})^4) + E(\varepsilon_i^4)E((\mathbf{X}_i^T \mathbf{G})^4)) \\
&= O(\Omega^{-4} ((\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2(\mathbf{G}^T \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma \mathbf{G})^2 + \tau^4(\mathbf{G}^T \boldsymbol{\Sigma} \mathbf{G})^2)) \\
&= O(n^2V^2).
\end{aligned}$$

Ahora, como los Z_i son i.i.d., y $\tilde{\mathbf{R}}^T \mathbf{G} = \frac{1}{n} \sum_{i=1}^n Z_i$, entonces la siguiente función característica satisface

$$\varphi_{V^{-1/2} \tilde{\mathbf{R}}^T \mathbf{G}}(t) = \prod_{i=1}^n \varphi_{Z_i} \left(\frac{t}{nV^{1/2}} \right) = \left(\varphi_Z \left(\frac{t}{nV^{1/2}} \right) \right)^n.$$

Por el teorema de Taylor y el hecho de que $E(Z) = 0$ y $E(Z^2) = nV$,

$$\begin{aligned}
\varphi_Z \left(\frac{t}{nV^{1/2}} \right) &= \varphi_Z(0) + \frac{t}{nV^{1/2}} \varphi'_Z(0) + \frac{1}{2} \left(\frac{t}{nV^{1/2}} \right)^2 \varphi''_Z(0) + o \left(\left(\frac{t}{nV^{1/2}} \right)^2 \right) \\
&= 1 + \frac{t}{nV^{1/2}} iE(Z) + \frac{1}{2} \frac{t^2}{n^2V} i^2 E(Z^2) + o \left(\frac{t^2}{n^2V} \right) \\
&= 1 - \frac{1}{2} \frac{t^2}{n^2V} nV + o \left(\frac{t^2}{n^2V} \right) \\
&= 1 - \frac{t^2}{2n} + o \left(\frac{t^2}{n^2V} \right),
\end{aligned}$$

donde

$$\left| o \left(\frac{t^2}{n^2V} \right) \right| \leq \frac{1}{3!} M \left(\frac{t}{nV^{1/2}} \right)^3,$$

con $|\varphi'''| \leq M$ en un entorno de 0. Pero como todos los momentos de la variable Z existen, φ'' es continuamente diferenciable, y $|\varphi'''_Z| \leq 2|EZ^3| \leq 2(EZ^4)^{3/4}$ en un entorno de 0. Entonces

$$\left| o \left(\frac{t^2}{n^2V} \right) \right| \leq \frac{1}{3} \left(\frac{t}{nV^{1/2}} \right)^3 (EZ^4)^{3/4} \leq \frac{1}{3} \left(\frac{t}{nV^{1/2}} \right)^3 (48n^2V^2)^{3/4} = c \left(\frac{t}{\sqrt{n}} \right)^3.$$

Así que claramente

$$\varphi_{V^{-1/2} \tilde{\mathbf{R}}^T \mathbf{G}}(t) \longrightarrow e^{-t^2/2} \quad \text{cuando } n, p \rightarrow \infty \text{ y se sigue que } V^{-1/2} \tilde{\mathbf{R}}^T \mathbf{G} \rightsquigarrow N(0, 1).$$

Notar que esto es una demostración “clásica” del teorema central del límite, pero estamos usando $n, p \rightarrow \infty$.

Nos queda ver que

$$\begin{aligned} V^{-1/2}\Omega^{-1}(\mathbf{Q}_\sigma\bar{\mathbf{X}}\bar{\mathbf{X}}^T\boldsymbol{\beta} + \bar{\mathbf{X}}\bar{\boldsymbol{\varepsilon}})^T\mathbf{G} &= V^{-1/2}\Omega^{-1}((\bar{\mathbf{X}}^T\boldsymbol{\beta})(\bar{\mathbf{X}}^T\mathbf{Q}_\sigma\mathbf{G}) + \bar{\boldsymbol{\varepsilon}}(\bar{\mathbf{X}}^T\mathbf{G})) \\ &= O_p(n^{-1/2}). \end{aligned}$$

Nuevamente $\bar{\mathbf{X}} \perp \bar{\boldsymbol{\varepsilon}}$ y $\bar{\mathbf{X}}^T\boldsymbol{\beta} \perp \bar{\mathbf{X}}^T\mathbf{Q}_\sigma$, con $\bar{\mathbf{X}}^T\boldsymbol{\beta} \sim N(0, n^{-1}\boldsymbol{\beta}^T\Sigma\boldsymbol{\beta})$, $\bar{\mathbf{X}}^T\mathbf{G} \sim N(0, n^{-1}\mathbf{G}^T\Sigma\mathbf{G})$, $\bar{\mathbf{X}}^T\mathbf{Q}_\sigma\mathbf{G} \sim N(0, n^{-1}\mathbf{G}^T\mathbf{Q}_\sigma\Sigma\mathbf{Q}_\sigma\mathbf{G})$ y $\bar{\boldsymbol{\varepsilon}}$ satisface $E(\bar{\boldsymbol{\varepsilon}}) = 0$ y $\text{var}(\bar{\boldsymbol{\varepsilon}}) = \frac{1}{n}\tau^2$. Luego

$$\begin{aligned} E((\bar{\mathbf{X}}^T\boldsymbol{\beta})(\bar{\mathbf{X}}^T\mathbf{Q}_\sigma\mathbf{G}) + \bar{\boldsymbol{\varepsilon}}(\bar{\mathbf{X}}^T\mathbf{G})) &= 0, \\ E\left(\left((\bar{\mathbf{X}}^T\boldsymbol{\beta})(\bar{\mathbf{X}}^T\mathbf{Q}_\sigma\mathbf{G}) + \bar{\boldsymbol{\varepsilon}}(\bar{\mathbf{X}}^T\mathbf{G})\right)^2\right) &= \text{var}(\bar{\mathbf{X}}^T\boldsymbol{\beta})\text{var}(\bar{\mathbf{X}}^T\mathbf{Q}_\sigma\mathbf{G}) + \text{var}(\bar{\boldsymbol{\varepsilon}})\text{var}(\bar{\mathbf{X}}^T\mathbf{G}) \\ &= \frac{1}{n^2}\left((\boldsymbol{\beta}^T\Sigma\boldsymbol{\beta})(\mathbf{G}^T\mathbf{Q}_\sigma\Sigma\mathbf{Q}_\sigma\mathbf{G}) + \tau^2(\mathbf{G}^T\Sigma\mathbf{G})\right) \\ &= \frac{\Omega^2}{n}V. \end{aligned}$$

Se sigue entonces que $(\bar{\mathbf{X}}^T\boldsymbol{\beta})(\bar{\mathbf{X}}^T\mathbf{Q}_\sigma\mathbf{G}) + \bar{\boldsymbol{\varepsilon}}(\bar{\mathbf{X}}^T\mathbf{G}) = O_p\left(\frac{\Omega}{n^{1/2}}V^{1/2}\right)$ y el hecho de que

$$V^{-1/2}\Omega^{-1}(\mathbf{Q}_\sigma\bar{\mathbf{X}}\bar{\mathbf{X}}^T\boldsymbol{\beta} + \bar{\mathbf{X}}\bar{\boldsymbol{\varepsilon}})^T\mathbf{G} = O_p(n^{-1/2})$$

queda probado.

E.1.3. Demostración del Lema E.2 y Lema E.3

Usando que $\Omega\boldsymbol{\sigma}^T\boldsymbol{\sigma} = \boldsymbol{\sigma}^T\Sigma\boldsymbol{\sigma}$, tenemos

$$\begin{aligned} U &= \frac{1}{\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\hat{\boldsymbol{\sigma}}^T\mathbf{Q}_\sigma\hat{\boldsymbol{\sigma}} - \frac{1}{\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\hat{\boldsymbol{\sigma}}^T\mathbf{Q}_\sigma\hat{\Sigma}\hat{\boldsymbol{\sigma}}\Omega^{-1} + \frac{\hat{\boldsymbol{\sigma}}^T\boldsymbol{\sigma}}{\boldsymbol{\sigma}^T\boldsymbol{\sigma}} - \frac{\hat{\boldsymbol{\sigma}}^T\hat{\Sigma}\boldsymbol{\sigma}}{\boldsymbol{\sigma}^T\Sigma\boldsymbol{\sigma}} - 1 + \frac{\boldsymbol{\sigma}^T\hat{\Sigma}\boldsymbol{\sigma}}{\boldsymbol{\sigma}^T\Sigma\boldsymbol{\sigma}} \\ &\quad + \left(\frac{2}{m} + \frac{C(p,1)}{m^2\Omega} + \sigma_Y^2\frac{C(p,1)}{m^2\boldsymbol{\sigma}^T\boldsymbol{\sigma}} + \frac{\sigma_Y^2C(p,2)}{m^2\Omega\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\right) \\ &= \left(\frac{1}{\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\hat{\boldsymbol{\sigma}}^T\mathbf{Q}_\sigma\hat{\boldsymbol{\sigma}}\right) - \left(\frac{1}{\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\hat{\boldsymbol{\sigma}}^T\mathbf{Q}_\sigma\hat{\Sigma}\mathbf{Q}_\sigma\hat{\boldsymbol{\sigma}}\Omega^{-1}\right) - \left(\frac{1}{\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\hat{\boldsymbol{\sigma}}^T\mathbf{Q}_\sigma\hat{\Sigma}\mathbf{P}_\sigma\hat{\boldsymbol{\sigma}}\Omega^{-1}\right) \\ &\quad + \left(\frac{\hat{\boldsymbol{\sigma}}^T\boldsymbol{\sigma}}{\boldsymbol{\sigma}^T\boldsymbol{\sigma}} - \frac{\hat{\boldsymbol{\sigma}}^T\hat{\Sigma}\boldsymbol{\sigma}}{\boldsymbol{\sigma}^T\Sigma\boldsymbol{\sigma}} - 1 + \frac{\boldsymbol{\sigma}^T\hat{\Sigma}\boldsymbol{\sigma}}{\boldsymbol{\sigma}^T\Sigma\boldsymbol{\sigma}}\right) \\ &\quad + \left(\frac{2}{m} + \frac{C(p,1)}{m^2\Omega} + \sigma_Y^2\frac{C(p,1)}{m^2\boldsymbol{\sigma}^T\boldsymbol{\sigma}} + \frac{\sigma_Y^2C(p,2)}{m^2\Omega\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\right) \\ &= U_1 + U_2 + U_3 + U_4 + \left(\frac{2}{m} + \frac{C(p,1)}{m^2\Omega} + \sigma_Y^2\frac{C(p,1)}{m^2\boldsymbol{\sigma}^T\boldsymbol{\sigma}} + \frac{\sigma_Y^2C(p,2)}{m^2\Omega\boldsymbol{\sigma}^T\boldsymbol{\sigma}}\right). \end{aligned} \tag{E.1}$$

E.1.3.1. Orden del primer término

Por el Apéndice C,

$$\begin{aligned} U_1 &= \frac{1}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \widehat{\boldsymbol{\sigma}}^T \mathbf{Q}_\sigma \widehat{\boldsymbol{\sigma}} = \frac{1}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} (\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon} + 2\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon}) \\ &= U_{1.1} + U_{1.2} + U_{1.3}. \end{aligned}$$

Para estudiar el orden de U_1 , vamos a estudiar el orden de $U_{1.1}, U_{1.2}, U_{1.3}$, y adaptamos la notación del Corolario C.2

▪ $U_{1.1}$:

$$\begin{aligned} E(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta}) &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) E(N \text{tr}(\mathbf{T})) \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) E(N) \text{tr}(E(\mathbf{T})) \quad \text{por (C.4) y (C.7)} \end{aligned}$$

$$\begin{aligned} E(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta})^2 &= E(\text{tr}((\mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma)^2)) \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 E(N^2) E(\text{tr}^2(\mathbf{T})) \quad \text{por (C.4) y (C.7)} \end{aligned}$$

$$\begin{aligned} E(U_{1.1}) &= \frac{(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} E(N) \text{tr}(E(\mathbf{T})) \\ &= \frac{(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} m \text{tr}(E(\mathbf{T})) \quad \text{por Lema B.1} \\ &= \frac{\Omega^{-1}}{m} \text{tr}(E(\mathbf{T})) \\ \text{var}(U_{1.1}) &= \frac{1}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} (E(N^2) E(\text{tr}^2(\mathbf{T})) - E(N)^2 \text{tr}^2(E(\mathbf{T}))) \\ &= \frac{(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} (m(m+2) E(\text{tr}^2(\mathbf{T})) - m^2 \text{tr}^2(E(\mathbf{T}))) \quad \text{por Lema B.1} \\ &= \frac{(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} (2m E(\text{tr}(\mathbf{T}^2)) + m^2 \text{var}(\text{tr}(\mathbf{T}))) \quad \text{por (C.7)} \\ &= O\left(\frac{J}{m}\right) \end{aligned}$$

por Corolario C.6, [Cook and Forzani, 2018] y $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$.

▪ $U_{1.2}$:

$$E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}^T \mathbf{L} \boldsymbol{\varepsilon}) = \tau^2 \text{tr}(E(\mathbf{H})) \quad \text{por (C.6) y (C.11)}$$

$$\begin{aligned} E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon})^2 &= E(E_\varepsilon(\boldsymbol{\varepsilon}^T \mathbf{L} \boldsymbol{\varepsilon})^2) \quad \text{por (C.11) y (C.6)} \\ &= E(\text{var}_\varepsilon(\boldsymbol{\varepsilon}^T \mathbf{L} \boldsymbol{\varepsilon}) + E_\varepsilon^2(\boldsymbol{\varepsilon}^T \mathbf{L} \boldsymbol{\varepsilon})) \end{aligned}$$

$$= 2\tau^4 \text{tr}(E(\mathbf{L}^2)) + \tau^4 E(\text{tr}^2(\mathbf{L})) \quad \text{por (C.11)}$$

$$= 2\tau^4 \text{tr}(E(\mathbf{H}^2)) + \tau^4 E(\text{tr}^2(\mathbf{H})) \quad \text{por (C.6)}$$

$$E(U_{1.2}) = \frac{\tau^2}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \text{tr}(E(\mathbf{H}))$$

$$\text{var}(U_{1.2}) = \frac{\tau^4}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} (2\text{tr}(E(\mathbf{H}^2)) + E(\text{tr}^2(\mathbf{H})) - \text{tr}^2(E(\mathbf{H})))$$

$$= \frac{\tau^4}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} (2\text{tr}(E(\mathbf{H}^2)) + \text{var}(\text{tr}(\mathbf{H})))$$

$$= O\left(\frac{J}{m}\right) \quad \text{por Corolario C.6 y } \tau^2 = O(1).$$

■ $U_{1.3}$:

$$E(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon}) = 0 \quad \text{por (C.11)}$$

$$E(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon})^2 = \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) E(\text{tr}(\mathbf{L} \mathbf{M} \mathbf{L})) \quad \text{por (C.11), (C.3) y (C.6)}$$

$$= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(E(\mathbf{L}^2) E(\mathbf{M})) \quad \text{por (C.6)}$$

$$= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(E(\mathbf{L}^2 \mathbf{Q}_n)) \quad \text{por Lema B.2}$$

$$= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(E(\mathbf{H}^2))$$

$$E(U_{1.3}) = 0$$

$$\text{var}(U_{1.3}) = \frac{\tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \text{tr}(E(\mathbf{H}^2))$$

$$= O\left(\frac{J}{m}\right) \quad \text{por Corolario C.6, } \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1) \text{ y } \tau^2 = O(1).$$

Por lo tanto

$$E(U_1) = \left(\frac{\Omega^{-1}}{m} \text{tr}(E(\mathbf{T})) + \frac{\tau^2}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \text{tr}(E(\mathbf{H})) \right)$$

$$= \left(\frac{\Omega^{-1}}{m} C(p, 1) + \frac{\tau^2}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} m C(p, 1) \right) \quad \text{por Lema B.2}$$

$$= \left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} + \tau^2 \right) \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}}$$

$$= \sigma_Y^2 \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \tag{E.2}$$

$$\text{var}(U_1) = O(J/m). \tag{E.3}$$

E.1.3.2. Orden del segundo término

Para U_2 , nuevamente por el Apéndice C

$$\begin{aligned} U_2 &= \frac{-1}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \widehat{\boldsymbol{\sigma}}^T \mathbf{Q}_\sigma \widehat{\boldsymbol{\Sigma}} \mathbf{Q}_\sigma \widehat{\boldsymbol{\sigma}} \\ &= \frac{-1}{m^3(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})} \left(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon} + 2\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \right) \\ &= U_{2.1} + U_{2.2} + U_{2.3}. \end{aligned}$$

Para estudiar el orden de U_2 , vamos a analizar el orden de $U_{2.1}, U_{2.2}, U_{2.3}$.

▪ $U_{2.1}$:

$$\begin{aligned} E \left(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \right) &= E \left(\text{tr} \left(\mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{F} (\mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F})^2 \right) \right) \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(E(\mathbf{M} \mathbf{L}^2)) \quad \text{por (C.3) y (C.6)} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(E(\mathbf{M}) E(\mathbf{L}^2)) \quad \text{por (C.6)} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(E(\mathbf{Q}_n \mathbf{L}^2)) \quad \text{por (C.6) y Lema B.2} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(E(\mathbf{H}^2)) \end{aligned}$$

$$\begin{aligned} E \left((\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta})^2 \right) &= \\ &= E \left(\text{tr} \left(\mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F} \mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{F} \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F} \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F} \mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{F} \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{F} \right) \right) \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 E \left(\text{tr}(\mathbf{L} \mathbf{M} \mathbf{L}^2 \mathbf{M} \mathbf{L}) \right) \quad \text{por (C.3) y (C.6)} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 \text{tr}(E(E_{\mathbf{M}}((\mathbf{L} \mathbf{M} \mathbf{L})^2))) \end{aligned}$$

recordando que $\mathbf{M} \sim W(1, \mathbf{Q}_n)$ y que $\mathbf{M} \perp \mathbf{L}$, se tiene $\mathbf{L} \mathbf{M} \mathbf{L} | \mathbf{L} \sim W(1, \mathbf{L}^2)$

$$\begin{aligned} &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 \text{tr}(E(\mathbf{L}^2 \text{tr}(\mathbf{L}^2) + 2\mathbf{L}^4)) \quad \text{por Lema B.2} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 (E(\text{tr}^2(\mathbf{H}^2)) + 2\text{tr}(E(\mathbf{H}^4))) \quad \text{por (C.6),} \end{aligned}$$

$$\begin{aligned} E(U_{2.1}) &= \frac{-(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})}{m^3(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})} \text{tr}(E(\mathbf{H}^2)) \\ &= \frac{-\Omega^{-2}}{m^3} \text{tr}(E(\mathbf{H}^2)) \\ \text{var}(U_{2.1}) &= \frac{(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2}{m^6(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^2} (E(\text{tr}^2(\mathbf{H}^2)) \\ &\quad + 2\text{tr}(E(\mathbf{H}^4)) - \text{tr}^2(E(\mathbf{H}^2))) \\ &= \frac{\Omega^{-4}}{m^6} (\text{var}(\text{tr}(\mathbf{H}^2)) + 2\text{tr}(E(\mathbf{H}^4))) \end{aligned}$$

$$= O\left(\frac{J}{m} + \frac{C(p,4)}{m^2(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4}\right)$$

por Corolario C.6 y $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$.

▪ $U_{2.2}$:

$$E\left(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon}\right) = \tau^2 E(\text{tr}(\mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{F})) \quad \text{por (C.11)}$$

$$= \tau^2 E(\text{tr}(\mathbf{L}^2)) \quad \text{por (C.6)}$$

$$= \tau^2 \text{tr}(E(\mathbf{H}^2)) \quad \text{por (C.6)}$$

$$E\left((\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon})^2\right) = E\left((\boldsymbol{\varepsilon}^T \mathbf{L}^2 \boldsymbol{\varepsilon})^2\right)$$

$$= E\left(\text{var}_\varepsilon(\boldsymbol{\varepsilon}^T \mathbf{L}^2 \boldsymbol{\varepsilon}) + E_\varepsilon^2(\boldsymbol{\varepsilon}^T \mathbf{L}^2 \boldsymbol{\varepsilon})\right) \quad \text{por (C.11)}$$

$$= 2\tau^4 E(\text{tr}(\mathbf{L}^4)) + \tau^4 E(\text{tr}^2(\mathbf{L}^2)) \quad \text{por (C.11)}$$

$$= 2\tau^4 \text{tr}(E(\mathbf{H}^4)) + \tau^4 E(\text{tr}^2(\mathbf{H}^2)) \quad \text{por (C.6)}$$

$$E(U_{2.2}) = \frac{-\tau^2}{m^3(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})} \text{tr}(E(\mathbf{H}^2))$$

$$\text{var}(U_{2.2}) = \frac{\tau^4}{m^6(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^2} \left(2\text{tr}(E(\mathbf{H}^4)) + E(\text{tr}^2(\mathbf{H}^2)) - \text{tr}^2(E(\mathbf{H}^2))\right)$$

$$= \frac{\tau^4}{m^6(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^2} \left(2\text{tr}(E(\mathbf{H}^4)) + \text{var}(\text{tr}(\mathbf{H}^2))\right)$$

$$= O\left(\frac{J}{m} + \frac{C(p,4)}{m^2(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4}\right)$$

por Corolario C.6, $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$ y $\tau^2 = O(1)$.

▪ $U_{2.3}$:

$$E\left(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta}\right) = 0 \quad \text{por (C.11)}$$

$$E\left((\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta})^2\right) = E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \mathbf{Q}_\sigma \mathbf{F} \boldsymbol{\varepsilon})$$

$$= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) E(\boldsymbol{\varepsilon}^T \mathbf{L}^2 \mathbf{M} \mathbf{L}^2 \boldsymbol{\varepsilon}) \quad \text{por (C.6) y (C.3)}$$

$$= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \tau^2 \text{tr}(E(\mathbf{L}^2 \mathbf{M} \mathbf{L}^2)) \quad \text{por (C.11)}$$

recordando que $\mathbf{M} \sim W(1, \mathbf{Q}_n)$ y que $\mathbf{M} \perp \mathbf{L}$, así que $\mathbf{L}^2 \mathbf{M} \mathbf{L}^2 | \mathbf{L} \sim W(1, \mathbf{L}^4)$

$$= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \tau^2 \text{tr}(E(\mathbf{L}^4)) \quad \text{por Lema B.2}$$

$$= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \tau^2 \text{tr}(E(\mathbf{H}^4)) \quad \text{por (C.6)}$$

$$E(U_{2.3}) = 0$$

$$\begin{aligned}\text{var}(U_{2,3}) &= 4 \frac{(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \tau^2}{m^6 (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^2} \text{tr}(E(\mathbf{H}^4)) \\ &= O\left(\frac{J}{m} + \frac{C(p,4)}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4}\right)\end{aligned}$$

por Corolario C.6, $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$ y $\tau^2 = O(1)$.

Por lo tanto

$$\begin{aligned}E(U_2) &= \frac{-\Omega^{-2}}{m^3} \text{tr}(E(\mathbf{H}^2)) + \frac{-\tau^2}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})} \text{tr}(E(\mathbf{H}^2)) \\ &= -\left(\frac{\Omega^{-2}}{m^3} + \frac{\tau^2}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})}\right) (mC^2(p,1) + (m^2 + m)C(p,2)) \quad \text{por Lema B.2} \\ &= -\left(\frac{(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2}{\Omega^2} + \frac{\tau^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega}\right) \left(\frac{C^2(p,1)}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} + \frac{C(p,2)}{m (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} + \frac{C(p,2)}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2}\right) \\ &= -\sigma_Y^2 \left(\frac{C(p,1)^2}{m^2 \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} + \frac{C(p,2)}{m \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} + \frac{C(p,2)}{m^2 \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})}\right) \quad (\text{E.4})\end{aligned}$$

$$\text{var}(U_2) = O\left(\frac{J}{m} + \frac{C(p,4)}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4}\right). \quad (\text{E.5})$$

E.1.3.3. Orden del tercer término

Para U_3 , usamos el hecho de que $\mathbf{P}_\sigma = \frac{\Omega^2}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \boldsymbol{\beta} \boldsymbol{\beta}^T$, luego

$$\begin{aligned}U_3 &= -\frac{\hat{\boldsymbol{\sigma}}^T \mathbf{Q}_\sigma \hat{\boldsymbol{\Sigma}} \mathbf{P}_\sigma \hat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \\ &= -\frac{\Omega}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \left(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{F} \boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta} \right. \\ &\quad \left. + \boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \right) \\ &= U_{3,1} + U_{3,2} + U_{3,3} + U_{3,4}.\end{aligned}$$

Para hallar el orden de U_3 , vamos a encontrar el orden de $U_{3,1}, U_{3,2}, U_{3,3}, U_{3,4}$.

▪ $U_{3,1}$:

$$\begin{aligned}E(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta}) &= E(\text{tr}(\mathbf{T}) N^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2) \quad \text{por (C.7) y (C.4)} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 E(\text{tr}(\mathbf{T})) E(N^2) \quad \text{por (C.7)} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 m(m+2) E(\text{tr}(\mathbf{T})) \quad \text{por Lema B.1}\end{aligned}$$

$$\begin{aligned}E((\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta})^2) &= E((\text{tr}(\mathbf{T}) N^2 (\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta})^2)^2) \quad \text{por (C.7)} \\ &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^4 E(\text{tr}^2(\mathbf{T})) E(N^4) \quad \text{por (C.4) y (C.7)}\end{aligned}$$

$$\begin{aligned}
E(U_{3.1}) &= -\frac{\Omega}{m^3(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2}(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})^2m(m+2)E(\text{tr}(\mathbf{T})) \\
&= -\frac{\Omega^{-1}(m+2)}{m^2}E(\text{tr}(\mathbf{T})) \\
\text{var}(U_{3.1}) &= \frac{\Omega^2(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})^4}{m^6(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^4}(E(\text{tr}^2(\mathbf{T}))E(N^4) - E^2(\text{tr}(\mathbf{T}))E^2(N^2)) \\
&= O\left(\frac{1}{m^6(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2}\right)(m^3E(\text{tr}^2(\mathbf{T})) + m^4\text{var}(\text{tr}(\mathbf{T})))
\end{aligned}$$

usando el hecho de que $E(N^4) = E^2(N^2) + 8m(m+2)(m+3)$ por Lema B.1

$$\text{y que } \frac{\boldsymbol{\sigma}^T\boldsymbol{\sigma}}{\Omega} = O(1)$$

$$= O(J/m) \quad \text{por Corolario C.6.}$$

▪ $U_{3.2}$:

$$\begin{aligned}
E(\boldsymbol{\varepsilon}^T\mathbf{F}^T\mathbf{Q}_\sigma\mathbf{W}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}\boldsymbol{\varepsilon}) &= \tau^2E(\text{tr}(\mathbf{LM}(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta}))) \quad \text{por (C.11), (C.6) y (C.3)} \\
&= \tau^2(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})\text{tr}(E(\mathbf{L})E(\mathbf{M})) \\
&= \tau^2(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})\text{tr}(E(\mathbf{LQ}_n)) \quad \text{por Lema B.2 y (C.6)} \\
&= \tau^2(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})\text{tr}(E(\mathbf{H}))
\end{aligned}$$

$$\begin{aligned}
E((\boldsymbol{\varepsilon}^T\mathbf{F}^T\mathbf{Q}_\sigma\mathbf{W}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}\boldsymbol{\varepsilon})^2) &= E((\boldsymbol{\varepsilon}^T\mathbf{LM}(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})\boldsymbol{\varepsilon})^2) \quad \text{por (C.3) y (C.6)} \\
&= (\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})^2E(\text{var}_\varepsilon(\boldsymbol{\varepsilon}^T\mathbf{LM}\boldsymbol{\varepsilon}) + E_\varepsilon^2(\boldsymbol{\varepsilon}^T\mathbf{LM}\boldsymbol{\varepsilon})) \quad \text{por (C.11)} \\
&= (\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})^2E(\tau^4\text{tr}((\mathbf{LM})^2 + \mathbf{LMML}) + \tau^4\text{tr}^2(\mathbf{LM})) \quad \text{por (C.11)} \\
&= \tau^4(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})^2E(\text{tr}(\mathbf{T}^2)N^2 + \text{tr}(\mathbf{L}^2\mathbf{M}^2) + \text{tr}^2(\mathbf{T})N^2) \quad \text{por (C.8)} \\
&= \tau^4(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})^2(\text{tr}(E(\mathbf{L}^2)E(\mathbf{M}^2)) + 2\text{tr}(E(N^2)E(\mathbf{T}^2))) \\
&\quad \text{por (C.6) y (C.7)} \\
&= \tau^4(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})^2((m+2)\text{tr}(E(\mathbf{H}^2)) + 2m(m+2)\text{tr}(E(\mathbf{T}^2))) \\
&\quad \text{por Lema B.2, (C.6) y Lema B.1}
\end{aligned}$$

$$\begin{aligned}
E(U_{3.2}) &= -\frac{\Omega}{m^3(\boldsymbol{\sigma}^T\boldsymbol{\sigma})^2}\tau^2(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})\text{tr}(E(\mathbf{H})) \\
&= -\frac{\tau^2}{m^3(\boldsymbol{\sigma}^T\boldsymbol{\sigma})}\text{tr}(E(\mathbf{H}))
\end{aligned}$$

$$\begin{aligned}
\text{var}(U_{3.2}) &= \frac{\tau^4}{m^6(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} ((m+2)\text{tr}(E(\mathbf{H}^2))) \\
&\quad + 2m(m+2)\text{tr}(E(\mathbf{T}^2)) - \text{tr}^2(E(\mathbf{H})) \\
&= O\left(\frac{J}{m^2}\right) \quad \text{por Corolario C.6 y } \tau^2 = O(1).
\end{aligned}$$

■ $U_{3.3}$:

$$\begin{aligned}
E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta}) &= 0 \quad \text{por (C.11)} \\
E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta})^2) &= (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 E(\boldsymbol{\varepsilon}^T \mathbf{L} \mathbf{M}^3 \mathbf{L} \boldsymbol{\varepsilon}) \quad \text{por (C.6) y (C.3)} \\
&= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 \text{tr}(E(\mathbf{L}^2) E(\mathbf{M}^3)) \quad \text{por (C.11) y (C.6)} \\
&= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 (m+2)(m+4) \text{tr}(E(\mathbf{L}^2 \mathbf{Q}_n)) \quad \text{por Lema B.2} \\
&= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 (m+2)(m+4) \text{tr}(E(\mathbf{H}^2)) \quad \text{por (C.6)}
\end{aligned}$$

$$E(U_{3.3}) = 0$$

$$\begin{aligned}
\text{var}(U_{3.3}) &= \frac{\Omega^2}{m^6(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 (m+2)(m+4) \text{tr}(E(\mathbf{H}^2)) \\
&= \frac{\tau^2}{m^6 \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} (m+2)(m+4) \text{tr}(E(\mathbf{H}^2)) \\
&= O\left(\frac{1}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2}\right) \text{tr}(E(\mathbf{H}^2)) \\
&\quad \text{por } \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1) \text{ y } \tau^2 = O(1) \\
&= O(J/m) \quad \text{por Corolario C.6.}
\end{aligned}$$

■ $U_{3.4}$:

$$\begin{aligned}
E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta}) &= 0 \quad \text{por (C.11)} \\
E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta})^2) &= E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\beta} \text{tr}(\mathbf{T}) N (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}))^2) \quad \text{por (C.7)} \\
&= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 \text{tr}(E(\text{tr}^2(\mathbf{T}) N^2 \mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{F})) \quad \text{por (C.11)} \\
&= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 \text{tr}(E(\text{tr}^2(\mathbf{T}) N^2 \mathbf{M})) \quad \text{por (C.3)} \\
&= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 \text{tr}(E(\mathbf{T}^2)) E(N^3) \quad \text{por (C.7) y (C.4)} \\
&= \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 m(m+2)(m+4) \text{tr}(E(\mathbf{T}^2)) \quad \text{por Lema B.1}
\end{aligned}$$

$$\begin{aligned}
E(U_{3.4}) &= 0 \\
\text{var}(U_{3.4}) &= \frac{\Omega^2}{m^6(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} \tau^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 m(m+2)(m+4) \text{tr}(E(\mathbf{T}^2)) \\
&= \frac{\tau^2}{m^6 \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} m(m+2)(m+4) \text{tr}(E(\mathbf{T}^2)) \\
&= O\left(\frac{1}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2}\right) \text{tr}(E(\mathbf{T}^2)) \\
&\text{por } \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1) \text{ y } \tau^2 = O(1) \\
&= O(J/m) \quad \text{por Corolario C.6.}
\end{aligned}$$

Entonces

$$\begin{aligned}
E(U_3) &= -\frac{\Omega^{-1}(m+2)}{m^2} E(\text{tr}(\mathbf{T})) - \frac{\tau^2}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \text{tr}(E(\mathbf{H})) \\
&= -\frac{\Omega^{-1}(m+2)}{m^2} C(p, 1) - \frac{\tau^2}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} m C(p, 1) \quad \text{por Lema B.2} \\
&= -\left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \left(1 + \frac{2}{m}\right) + \frac{\tau^2}{m}\right) \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \\
&= -\frac{C(p, 1)}{m \Omega} - \left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} + \sigma_Y^2\right) \frac{C(p, 1)}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \tag{E.6}
\end{aligned}$$

$$\text{var}(U_3) = O(J/m). \tag{E.7}$$

E.1.3.4. Orden del cuarto término

Finalmente, para U_4

$$\begin{aligned}
U_4 &= \frac{\hat{\boldsymbol{\sigma}}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} - \frac{\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} - 1 + \frac{\boldsymbol{\sigma}^T \hat{\boldsymbol{\Sigma}} \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \\
&= \left(\frac{2\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\sigma}}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} - \frac{\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta}}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} - 1\right) + \left(\frac{\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\sigma}}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} - \frac{\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{W} \boldsymbol{\beta}}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}}\right) \\
&= U_{4.1} + U_{4.2}
\end{aligned}$$

y sólo necesitamos estudiar los órdenes de $U_{4.1}$ y $U_{4.2}$.

- $U_{4.1}$: Vamos a usar que $\boldsymbol{\beta}^T \boldsymbol{\Sigma}^k \boldsymbol{\beta} = \Omega^{k-2} \boldsymbol{\sigma}^T \boldsymbol{\sigma}$ y $\boldsymbol{\sigma} = \Omega \boldsymbol{\beta}$.

$$\begin{aligned}
E(\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\sigma}) &= m \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\sigma} \quad \text{por Lema B.2} \\
&= m \boldsymbol{\sigma}^T \boldsymbol{\sigma}
\end{aligned}$$

$$E((\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\sigma})^2) = \Omega^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 E(N^2) \quad \text{por (C.4)}$$

$$\begin{aligned}
&= (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 E(N^2) \\
&= (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 m(m+2) \quad \text{por Lema B.1} \\
\text{var} \left(\frac{\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\sigma}}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) &= \frac{1}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \left((\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 m(m+2) - m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 \right) \\
&= \frac{2}{m} \\
E(\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta}) &= \boldsymbol{\beta}^T (m \boldsymbol{\Sigma} \text{tr}(\boldsymbol{\Sigma}) + (m^2 + m) \boldsymbol{\Sigma}^2) \boldsymbol{\beta} \quad \text{por Lema B.2} \\
&= m(\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \text{tr}(\boldsymbol{\Sigma}) + (m^2 + m) \boldsymbol{\sigma}^T \boldsymbol{\sigma} \\
&= m \boldsymbol{\sigma}^T \boldsymbol{\sigma} (\Omega^{-1} \text{tr}(\boldsymbol{\Sigma}) + m + 1) \\
&= m \boldsymbol{\sigma}^T \boldsymbol{\sigma} (\Omega^{-1} C(p, 1) + m + 2) \tag{E.8}
\end{aligned}$$

$$\begin{aligned}
E((\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta})^2) &= E((\boldsymbol{\beta}^T \mathbf{W}(\mathbf{P}_\sigma + \mathbf{Q}_\sigma) \mathbf{W} \boldsymbol{\beta})^2) \\
&= E((\boldsymbol{\beta}^T \mathbf{W} \mathbf{P}_\sigma \mathbf{W} \boldsymbol{\beta})^2) + E((\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta})^2) \\
&\quad + 2E((\boldsymbol{\beta}^T \mathbf{W} \mathbf{P}_\sigma \mathbf{W} \boldsymbol{\beta})(\boldsymbol{\beta}^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta})) \\
&= \frac{\Omega^4}{(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^4 E(N^4) + (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 E(\text{tr}^2(\mathbf{T}) N^2) \\
&\quad + 2 \frac{\Omega^2}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 E(N^3 \text{tr}(\mathbf{T})) \\
&\quad \text{por (C.7) y (C.4)} \\
&= (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 (E(N^4) + \Omega^{-2} E(N^2) E(\text{tr}^2(\mathbf{T}))) \\
&\quad + 2\Omega^{-1} E(N^3) E(\text{tr}(\mathbf{T})) \quad \text{por (C.7)} \\
&= (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 (E(N^4) + \Omega^{-2} E(N^2) (C^2(p, 1) + 2C(p, 2))) \\
&\quad + 2\Omega^{-1} E(N^3) C(p, 1) \quad \text{por (C.5)} \tag{E.9}
\end{aligned}$$

$$\begin{aligned}
\text{var} \left(\frac{\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta}}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) &= \frac{E(N^4) + \Omega^{-2} E(N^2) (C^2(p, 1) + 2C(p, 2))}{m^4} \\
&\quad + \frac{2\Omega^{-1} E(N^3) C(p, 1)}{m^4} \tag{E.10} \\
&\quad - \frac{m^2 (\Omega^{-2} C^2(p, 1) + 2(m+2)\Omega^{-1} C(p, 1) + (m+2)^2)}{m^4} \\
&= \frac{8}{m} + O_p \left(\frac{J}{m} \right) \quad \text{por Corolario C.6 y } \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)
\end{aligned}$$

$$E((\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\sigma})(\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta})) = \Omega E((\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\beta})(\boldsymbol{\beta}^T \mathbf{W}(\mathbf{P}_\sigma + \mathbf{Q}_\sigma) \mathbf{W} \boldsymbol{\beta}))$$

$$\begin{aligned}
&= \frac{\Omega^3}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^3 E(N^3) + \Omega (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta})^2 E(N^2 \text{tr}(\mathbf{T})) \\
&\quad \text{por (C.7) y (C.4)} \\
&= (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 (E(N^3) + \Omega^{-1} E(N^2) E(\text{tr}(\mathbf{T}))) \quad \text{por (C.7)} \\
&= (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 (E(N^3) + \Omega^{-1} E(N^2) C(p, 1)) \quad \text{por (C.5)} \\
E(U_{4.1}) &= - \left(\frac{\Omega^{-1} C(p, 1)}{m} + \frac{2}{m} \right) \\
\text{var}(U_{4.1}) &= 4 \text{var} \left(\frac{\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\sigma}}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) + \text{var} \left(\frac{\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta}}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \\
&\quad - 4 \left(\frac{E((\boldsymbol{\beta}^T \mathbf{W} \boldsymbol{\sigma})(\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta}))}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} - \frac{E(\boldsymbol{\beta}^T \mathbf{W}^2 \boldsymbol{\beta})}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \\
&= \frac{16}{m} + O\left(\frac{J}{m}\right) - 4 \left(\frac{E(N^3) + \Omega^{-1} E(N^2) E(\text{tr}(\mathbf{T}))}{m^3} \right) \\
&\quad + 4 \left(\frac{m^2 \Omega^{-1} C(p, 1) + m^2 (m + 2)}{m^3} \right) \\
&= \frac{16}{m} + O\left(\frac{J}{m}\right) - 4 \left(\frac{4}{m} + O\left(\frac{J}{m}\right) \right) \\
&\quad \text{por Corolario C.6 y } \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1) \\
&= O\left(\frac{J}{m}\right).
\end{aligned}$$

- $U_{4.2}$: Vamos a usar $\boldsymbol{\beta}^T \boldsymbol{\Sigma}^k \boldsymbol{\beta} = \Omega^{k-2} \boldsymbol{\sigma}^T \boldsymbol{\sigma}$ y $\boldsymbol{\sigma} = \Omega \boldsymbol{\beta}$.

$$E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\sigma}) = 0 \quad \text{por (C.11)}$$

$$E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{W} \boldsymbol{\beta}) = 0 \quad \text{por (C.11)}$$

$$E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\sigma})^2) = \tau^2 \text{tr}(E(\mathbf{F}^T \boldsymbol{\sigma} \boldsymbol{\sigma}^T \mathbf{F})) \quad \text{por (C.11)}$$

$$= \tau^2 \Omega^2 (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) E(N) \quad \text{por (C.4)}$$

$$= \tau^2 m \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma} \quad \text{por Lema B.1}$$

$$E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{W} \boldsymbol{\beta})^2) = \tau^2 \text{tr}(E(\mathbf{F}^T \mathbf{W} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{F})) \quad \text{por (C.11)}$$

$$= \tau^2 \boldsymbol{\beta}^T E(\mathbf{W}^3) \boldsymbol{\beta}$$

$$= \tau^2 \boldsymbol{\beta}^T (m \boldsymbol{\Sigma} \text{tr}^2(\boldsymbol{\Sigma}) + (m^2 + m)(\boldsymbol{\Sigma} \text{tr}(\boldsymbol{\Sigma}^2) + 2 \boldsymbol{\Sigma}^2 \text{tr}(\boldsymbol{\Sigma})))$$

$$+ (m^3 + 3m^2 + 6m) \boldsymbol{\Sigma}^3) \boldsymbol{\beta} \quad \text{por Lema B.2}$$

$$= \tau^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma} (m \Omega^{-1} \text{tr}^2(\boldsymbol{\Sigma}) + (m^2 + m)(\Omega^{-1} \text{tr}(\boldsymbol{\Sigma}^2) + 2 \text{tr}(\boldsymbol{\Sigma})))$$

$$+ (m^3 + 3m^2 + 6m) \Omega)$$

(E.11)

$$\begin{aligned}
E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\sigma})(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{W} \boldsymbol{\beta})) &= \Omega E(\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{F} \boldsymbol{\varepsilon}) \\
&= \Omega \tau^2 \text{tr}(E(\mathbf{F}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{W} \mathbf{F})) \quad \text{por (C.11)} \\
&= \Omega \tau^2 \boldsymbol{\beta}^T E(\mathbf{W}^2) \boldsymbol{\beta} \\
&= \Omega \tau^2 \boldsymbol{\beta}^T (m \boldsymbol{\Sigma} \text{tr}(\boldsymbol{\Sigma}) + (m^2 + m) \boldsymbol{\Sigma}^2) \boldsymbol{\beta} \quad \text{por Lema B.2} \\
&= \tau^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma} (m \text{tr}(\boldsymbol{\Sigma}) + (m^2 + m) \Omega)
\end{aligned}$$

$$E(U_{4.2}) = 0$$

$$\begin{aligned}
\text{var}(U_{4.2}) &= \frac{E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\sigma})^2)}{m^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} + \frac{E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{W} \boldsymbol{\beta})^2)}{m^4 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \\
&\quad - 2 \frac{E((\boldsymbol{\varepsilon}^T \mathbf{F}^T \boldsymbol{\sigma})(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{W} \boldsymbol{\beta}))}{m^3 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \\
&= \frac{\tau^2 m^2 \Omega}{m^3 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \\
&\quad + \frac{\tau^2 (\Omega^{-1} \text{tr}^2(\boldsymbol{\Sigma}) + (m+1)(\Omega^{-1} \text{tr}(\boldsymbol{\Sigma}^2) + 2 \text{tr}(\boldsymbol{\Sigma})))}{m^3 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \\
&\quad + \frac{\tau^2 ((m^2 + 3m + 6)\Omega)}{m^3 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} - 2 \frac{\tau^2 (m \text{tr}(\boldsymbol{\Sigma}) + (m^2 + m)\Omega)}{m^3 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \\
&= \frac{\tau^2 (\Omega^{-1} \text{tr}^2(\boldsymbol{\Sigma}) + (m+1)(\Omega^{-1} \text{tr}(\boldsymbol{\Sigma}^2) + 2 \text{tr}(\boldsymbol{\Sigma})))}{m^3 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \\
&\quad + \frac{(m+6)\Omega - 2\tau^2 m \text{tr}(\boldsymbol{\Sigma})}{m^3 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \\
&= \frac{\tau^2 (\Omega^{-1} \text{tr}^2(\boldsymbol{\Sigma}) + (m+1)\Omega^{-1} \text{tr}(\boldsymbol{\Sigma}^2) + 2 \text{tr}(\boldsymbol{\Sigma}) + (m+6)\Omega)}{m^3 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \\
&= O\left(\frac{J}{m}\right) \quad \text{por } \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1) \text{ y } \tau^2 = O(1).
\end{aligned}$$

Con lo cual

$$E(U_4) = -\left(\frac{C(p, 1)}{m\Omega} + \frac{2}{m}\right) \quad (\text{E.12})$$

$$\text{var}(U_4) = O(J/n). \quad (\text{E.13})$$

E.1.3.5. Orden de U

Por (E.1), (E.2), (E.4), (E.6), (E.12), y usando el hecho de que $\sigma_Y^2 = \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} + \tau^2$, tenemos

$$\begin{aligned}
E(U) &= E(U_1) + E(U_2) + E(U_3) + E(U_4) + \left(\frac{2}{m} + \frac{C(p, 1)}{m^2 \Omega} + \sigma_Y^2 \frac{C(p, 1)}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{\sigma_Y^2 C(p, 2)}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}}\right) \\
&= \sigma_Y^2 \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}}
\end{aligned}$$

$$\begin{aligned}
& -\sigma_Y^2 \left(\frac{C(p,1)^2}{m^2\Omega\sigma^T\sigma} + \frac{C(p,2)}{m\Omega\sigma^T\sigma} + \frac{C(p,2)}{m^2\Omega\sigma^T\sigma} \right) \\
& - \frac{C(p,1)}{m\Omega} - \left(\frac{\sigma^T\sigma}{\Omega} + \sigma_Y^2 \right) \frac{C(p,1)}{m^2\sigma^T\sigma} \\
& - \left(\frac{C(p,1)}{m\Omega} + \frac{2}{m} \right) \\
& + \left(\frac{2}{m} + \frac{C(p,1)}{m^2\Omega} + \sigma_Y^2 \frac{C(p,1)}{m^2\sigma^T\sigma} + \frac{\sigma_Y^2 C(p,2)}{m^2\Omega\sigma^T\sigma} \right) \\
& = \left(\tau^2 - \frac{\sigma^T\sigma}{\Omega} \right) \frac{C(p,1)}{m\sigma^T\sigma} - \sigma_Y^2 \frac{C(p,1)^2}{m^2\Omega\sigma^T\sigma} - \sigma_Y^2 \frac{C(p,2)}{m\Omega\sigma^T\sigma} \\
& = b.
\end{aligned}$$

Adicionalmente

$$U - E(U) = O_p \left(\left(\frac{J}{m} \right)^{1/2} + \frac{1}{m^{1/2}} \left(\frac{C(p,4)}{m(\sigma^T\sigma)^4} \right)^{1/2} \right).$$

Luego, recordando que $V^{-1/2}|\beta^T\mathbf{G}| = O(\sqrt{n})$ por (3.11) obtenemos

$$V^{-1/2}(U - E(U))\beta^T\mathbf{G} = O_p \left(J^{1/2} + \left(\frac{C(p,4)}{n(\sigma^T\sigma)^4} \right)^{1/2} \right).$$

E.1.4. Demostración del Lema E.4

Necesitamos probar que $V^{-1/2}\Delta^T\mathbf{G} = O_p(K)$ (y esto concluye el resultado ya que $K \leq J$), donde

$$\begin{aligned}
\Delta & = (\sigma^T\Sigma\sigma)^{-1}(\hat{\sigma} - \sigma)^T(\hat{\sigma} + \sigma)(\hat{\sigma} - \sigma) + \frac{(\hat{\sigma} - \sigma)^T\sigma}{\sigma^T\Sigma\sigma} \left(\frac{\hat{\sigma}^T\sigma}{\sigma^T\sigma} - \frac{\hat{\sigma}^T\hat{\Sigma}\sigma}{\sigma^T\Sigma\sigma} \right) \sigma \\
& - \left(\frac{2}{m} + \frac{C(p,1)}{m^2\Omega} + \sigma_Y^2 \frac{C(p,1)}{m^2\sigma^T\sigma} + \frac{\sigma_Y^2 C(p,2)}{m^2\Omega\sigma^T\sigma} \right) \beta
\end{aligned} \tag{E.14}$$

$$= \Delta_1 + \Delta_2 + \Delta_3 \tag{E.15}$$

definiendo $\Delta_1 = (\sigma^T\Sigma\sigma)^{-1}(\hat{\sigma} - \sigma)^T(\hat{\sigma} + \sigma)(\hat{\sigma} - \sigma)$, $\Delta_2 = \frac{(\hat{\sigma} - \sigma)^T\sigma}{\sigma^T\Sigma\sigma} \left(\frac{\hat{\sigma}^T\sigma}{\sigma^T\sigma} - \frac{\hat{\sigma}^T\hat{\Sigma}\sigma}{\sigma^T\Sigma\sigma} \right) \sigma$ y

finalmente $\Delta_3 = - \left(\frac{2}{m} + \frac{C(p,1)}{m^2\Omega} + \sigma_Y^2 \frac{C(p,1)}{m^2\sigma^T\sigma} + \frac{\sigma_Y^2 C(p,2)}{m^2\Omega\sigma^T\sigma} \right) \beta$.

Usando las definiciones de $\hat{\Sigma}$, $\hat{\sigma}$ y que $\sigma^T\Sigma\sigma = \Omega\sigma^T\sigma$, $\sigma = \Omega\beta$

$$\frac{\hat{\sigma}^T\hat{\Sigma}\sigma}{\sigma^T\Sigma\sigma} - 1 = \frac{\varepsilon^T\mathbf{F}^T\mathbf{W}\sigma}{m^2\sigma^T\Sigma\sigma} + \frac{\beta^T\mathbf{W}^2\sigma}{m^2\sigma^T\Sigma\sigma} - 1 \tag{E.16}$$

$$E \left(\frac{\hat{\sigma}^T\hat{\Sigma}\sigma}{\sigma^T\Sigma\sigma} - 1 \right) = \frac{C(p,1)}{m\Omega} + \frac{2}{m} \text{ usando (E.8)} \tag{E.17}$$

$$\text{var} \left(\frac{\hat{\sigma}^T\hat{\Sigma}\sigma}{\sigma^T\Sigma\sigma} - 1 \right) = E \left(\frac{\varepsilon^T\mathbf{F}^T\mathbf{W}\sigma}{m^2\sigma^T\Sigma\sigma} \right)^2 + \text{var} \left(\frac{\beta^T\mathbf{W}^2\sigma}{m^2\sigma^T\Sigma\sigma} \right)$$

$$= E \left(\frac{\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{W} \boldsymbol{\sigma}}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right)^2 + O_p \left(\frac{1}{m} \right) \text{ por (E.10)} \quad (\text{E.18})$$

$$= O_p \left(\frac{1}{m} \right) \text{ usando (E.11)}. \quad (\text{E.19})$$

E.1.4.1. Primera cota

Por (C.9) tenemos que $E((\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \mathbf{G}) = 0$ y $\text{var}((\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \mathbf{G}) = \mathbf{G}^T \frac{\sigma_Y^2 \boldsymbol{\Sigma} + \boldsymbol{\sigma} \boldsymbol{\sigma}^T}{m} \mathbf{G}$. Como $\tau^2, \sigma_Y^2 \asymp 1$, $\boldsymbol{\Sigma} = \mathbf{P}_\sigma \Omega \mathbf{P}_\sigma + \mathbf{Q}_\sigma \Omega_0 \mathbf{Q}_\sigma$ y $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$, tenemos que existe una constante C tal que

$$\begin{aligned} \mathbf{G}^T (\sigma_Y^2 \boldsymbol{\Sigma} + \boldsymbol{\sigma} \boldsymbol{\sigma}^T) \mathbf{G} &\leq C \mathbf{G}^T (\tau^2 \boldsymbol{\Sigma} + (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}) \mathbf{Q}_\sigma \Omega_0 \mathbf{Q}_\sigma) \mathbf{G} \\ &= C n \Omega^2 V. \end{aligned}$$

Como consecuencia $\Omega^{-1} V^{-1/2} (\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \mathbf{G} = O_p(1)$. Por otro lado como $(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} \Omega = (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^{-1}$ y (A.11),

$$\begin{aligned} (\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^{-1} \Omega (\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T (\widehat{\boldsymbol{\sigma}} + \boldsymbol{\sigma}) &= \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} - 1 \\ &= O_p(K) \end{aligned}$$

y tenemos la cota para $\boldsymbol{\Delta}_1$.

E.1.4.2. Segunda cota

Usando $\boldsymbol{\Sigma} \boldsymbol{\sigma} = \Omega \boldsymbol{\sigma}$,

$$\begin{aligned} \frac{(\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} &= O_p \left(\frac{1}{\sqrt{n \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}}} \right) \text{ por (C.10)} \\ \frac{\widehat{\boldsymbol{\sigma}}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} - \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} &= \left(\frac{\widehat{\boldsymbol{\sigma}}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} - 1 \right) - \left(\frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} - 1 \right) \\ &= O_p(K) \text{ por (C.10), (E.17) y (E.19)} \end{aligned}$$

$$\mathbf{G}^T \boldsymbol{\sigma} \boldsymbol{\sigma}^T \mathbf{G} \leq C n V \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}.$$

De donde obtenemos que $V^{-1/2} \boldsymbol{\Delta}_2^T \mathbf{G} = O_p(K)$.

E.1.4.3. Tercera cota

La cota es una consecuencia de $\boldsymbol{\beta} = \Omega^{-1} \boldsymbol{\sigma}$ y

$$\Omega^{-2} \mathbf{G}^T \boldsymbol{\sigma} \boldsymbol{\sigma}^T \mathbf{G} = O(nV)$$

$$\left(\frac{2}{m} + \frac{C(p, 1)}{m^2 \Omega} + \sigma_Y^2 \frac{C(p, 1)}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{\sigma_Y^2 C(p, 2)}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) = O\left(\frac{1}{n}\right)$$

de donde se sigue que $V^{-1/2} \boldsymbol{\Delta}_3^T \mathbf{G} = O_p(n^{-1/2})$.

APÉNDICE F

DEMOSTRACIÓN DEL TEOREMA 3.5

Consideremos $\mathbf{X}_N \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma})$ una nueva observación de \mathbf{X} y supongamos sin pérdida de generalidad que $\boldsymbol{\mu}_X = \mathbf{0}$. Si probamos que

$$(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \boldsymbol{\Sigma} (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta}) = O_p \left(\frac{1}{n} + \left(\frac{C(p, 1)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \right)^2 + \frac{C(p, 2)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right),$$

entonces por la independencia de \mathbf{X}_N y $\widehat{\boldsymbol{\beta}}_{pls}$ tenemos que $E((\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \mathbf{X}_N) = 0$ y además

$$\text{var}((\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \mathbf{X}_N) = E((\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \boldsymbol{\Sigma} (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})),$$

lo que concluye el resultado.

Para probar la primera parte del teorema, por Proposición 3.1

$$(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \mathbf{X}_N = \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} (\mathbf{R}^T \mathbf{X}_N + \boldsymbol{\Delta}^T \mathbf{X}_N + U \boldsymbol{\beta}^T \mathbf{X}_N). \quad (\text{F.1})$$

Para encontrar el orden de $(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \boldsymbol{\Sigma} (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})$ podemos proceder de la siguiente manera: consideremos $\left(\frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right)^2 (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \mathbf{X}_N \mathbf{X}_N^T (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})$ y tomemos esperanza en \mathbf{X}_N , usando el hecho de que $E_{\mathbf{X}_N}(\mathbf{X}_N \mathbf{X}_N^T) = \boldsymbol{\Sigma}$. Entonces, para analizar el orden de $(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \boldsymbol{\Sigma} (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})$ sólo necesitamos los órdenes de $\mathbf{R}^T \boldsymbol{\Sigma} \mathbf{R}$, $\boldsymbol{\Delta}^T \boldsymbol{\Sigma} \boldsymbol{\Delta}$ y $U^2 \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}$.

F.1. Orden del primer término

Notemos que $\mathbf{R}^T \boldsymbol{\Sigma} \mathbf{R}$ es positivo, así que $\mathbf{R}^T \boldsymbol{\Sigma} \mathbf{R} = O_p(E_{\mathbf{R}}(\mathbf{R}^T \boldsymbol{\Sigma} \mathbf{R}))$, y como $E_{\mathbf{R}}(\mathbf{R}^T \mathbf{X}_N) = 0$ tenemos que, tomando $V = V(\mathbf{X}_N)$,

$$\begin{aligned} E_{\mathbf{R}}(\mathbf{R}^T \boldsymbol{\Sigma} \mathbf{R}) &= \text{tr}(\boldsymbol{\Sigma} E_{\mathbf{R}}(\mathbf{R} \mathbf{R}^T)) \\ &= E_{\mathbf{X}_N}(\text{tr}(\mathbf{X}_N \mathbf{X}_N^T E_{\mathbf{R}}(\mathbf{R} \mathbf{R}^T))) \end{aligned}$$

$$\begin{aligned}
&= E_{\mathbf{X}_N} (\mathbf{X}_N^T E_{\mathbf{R}} (\mathbf{R} \mathbf{R}^T) \mathbf{X}_N) \\
&= E_{\mathbf{X}_N} (\text{var}_{\mathbf{R}} (\mathbf{R}^T \mathbf{X}_N)) \\
&= \frac{n}{m} E_{\mathbf{X}_N} (V),
\end{aligned}$$

donde $\text{var}_{\mathbf{R}} (\mathbf{R}^T \mathbf{X}_N)$ puede ser hallado análogamente a $\text{var}(\tilde{\mathbf{R}}^T \mathbf{X}_N) = V$ del Apéndice E, usando que $\mathbf{R} = \Omega^{-1} \frac{\mathbf{Q}_\sigma \mathbf{W} \boldsymbol{\beta} + \mathbf{F} \boldsymbol{\varepsilon}}{m}$.

Para mostrar el orden de $E_{\mathbf{X}_N} (V)$, recordemos que

$$\begin{aligned}
V &= \frac{\Omega^{-2}}{n} \mathbf{X}_N^T [\boldsymbol{\Sigma} \tau^2 + \Omega^{-1} (\boldsymbol{\sigma}^T \boldsymbol{\sigma}) \mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma] \mathbf{X}_N \\
&= \frac{\Omega^{-2} \tau^2}{n} \text{tr} (\boldsymbol{\Sigma} \mathbf{X}_N \mathbf{X}_N^T) + \frac{\Omega^{-3} \boldsymbol{\sigma}^T \boldsymbol{\sigma}}{n} \text{tr} (\mathbf{Q}_\sigma \boldsymbol{\Sigma} \mathbf{Q}_\sigma \mathbf{X}_N \mathbf{X}_N^T \mathbf{Q}_\sigma).
\end{aligned}$$

Por lo tanto

$$\begin{aligned}
E_{\mathbf{X}_N} (V) &= \frac{\Omega^{-2} \tau^2}{n} \text{tr} (\boldsymbol{\Sigma}^2) + \frac{\Omega^{-3} \boldsymbol{\sigma}^T \boldsymbol{\sigma}}{n} \text{tr} (\mathbf{Q}_\sigma \boldsymbol{\Sigma}^2 \mathbf{Q}_\sigma) \\
&= \frac{\Omega^{-2} \tau^2}{n} (\Omega^2 + C(p, 2)) + \frac{\Omega^{-3} \boldsymbol{\sigma}^T \boldsymbol{\sigma}}{n} C(p, 2) \\
&\quad \text{usando que } \text{tr} \boldsymbol{\Sigma}^2 = \Omega^2 + C(p, 2), \text{tr} (\mathbf{Q}_\sigma \boldsymbol{\Sigma}^2 \mathbf{Q}_\sigma) = C(p, 2), \\
&\quad \tau^2 = O(1) \text{ y } \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1) \\
&= O_p \left(\frac{1}{n} + \frac{C(p, 2)}{n \Omega^2} \right),
\end{aligned}$$

y entonces podemos concluir que

$$\mathbf{R}^T \boldsymbol{\Sigma} \mathbf{R} = O_p \left(\frac{1}{n} + \frac{C(p, 2)}{n \Omega^2} \right). \tag{F.2}$$

F.2. Orden del segundo término

Recordemos que $\boldsymbol{\Delta} = \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2 + \boldsymbol{\Delta}_3$. Vamos a estudiar el orden de $\boldsymbol{\Delta}^T \boldsymbol{\Sigma} \boldsymbol{\Delta}$ sólo mirando los órdenes cuadráticos $\boldsymbol{\Delta}_i^T \boldsymbol{\Sigma} \boldsymbol{\Delta}_i$.

Comenzamos con

$$\boldsymbol{\Delta}_1^T \boldsymbol{\Sigma} \boldsymbol{\Delta}_1 = \frac{1}{(\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma})^2} (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\Sigma} (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) (\hat{\boldsymbol{\sigma}}^T \hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}^T \boldsymbol{\sigma})^2.$$

Notemos que

$$\begin{aligned}
E \left(\frac{1}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\Sigma} (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) \right) &= \frac{1}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \text{tr} (\boldsymbol{\Sigma} \text{var}(\hat{\boldsymbol{\sigma}})) \\
&= \frac{1}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \text{tr} \left(\frac{\boldsymbol{\Sigma}^2 \sigma_Y^2 + \boldsymbol{\Sigma} \boldsymbol{\sigma} \boldsymbol{\sigma}^T}{m} \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \left(\sigma_Y^2 \left(\frac{\Omega^2 + C(p, 2)}{n} \right) + \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{n} \right) \\
&= O \left(\frac{\Omega}{n \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{C(p, 2)}{n \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \right).
\end{aligned}$$

Entonces, como este es un término positivo

$$\frac{(\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\Sigma} (\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} = O_p \left(\frac{\Omega}{n \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{C(p, 2)}{n \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \right).$$

Para el resto de la expresión

$$\frac{1}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} (\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}^T \boldsymbol{\sigma})^2 = \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \left(\frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} - 1 \right)^2 = \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} O_p(K^2)$$

por (A.11). Luego tenemos

$$\begin{aligned}
\boldsymbol{\Delta}_1^T \boldsymbol{\Sigma} \boldsymbol{\Delta}_1 &= O_p \left(\frac{\Omega}{n \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{C(p, 2)}{n \Omega (\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \right) \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} O_p(K^2) \\
&= O_p \left(\frac{1}{n} + \frac{C(p, 2)}{n \Omega^2} \right) O_p(K^2).
\end{aligned}$$

Para $\boldsymbol{\Delta}_2^T \boldsymbol{\Sigma} \boldsymbol{\Delta}_2$, observemos que por el Apéndice E

$$\begin{aligned}
\boldsymbol{\Delta}_2^T \boldsymbol{\Sigma} \boldsymbol{\Delta}_2 &= \left(\frac{(\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 \left(\frac{\widehat{\boldsymbol{\sigma}}^T \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\sigma}} - \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right)^2 \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} \\
&= O_p(K^2/n)
\end{aligned}$$

y para $\boldsymbol{\Delta}_3^T \boldsymbol{\Sigma} \boldsymbol{\Delta}_3$, notemos que por el Apéndice E

$$\begin{aligned}
\boldsymbol{\Delta}_3^T \boldsymbol{\Sigma} \boldsymbol{\Delta}_3 &= \left(\frac{2}{m} + \frac{C(p, 1)}{m^2 \Omega} + \sigma_Y^2 \frac{C(p, 1)}{m^2 \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{\sigma_Y^2 C(p, 2)}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} \\
&= O_p(1/n^2) \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} \\
&= O_p(1/n^2).
\end{aligned}$$

F.3. Orden del tercer término

Sabiendo que $U = E(U) + O_p \left(\left(J/n + \frac{C(p, 4)}{n^2 (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} \right)^{1/2} \right)$, como $J = O(1)$ y $\frac{C(p, 4)}{n (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^4} = O(1)$, tenemos que

$$\begin{aligned}
U^2 \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} &= (E^2(U) + O_p(1/n)) \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} \\
&= O_p \left(\left(\frac{C(p, 1)}{n \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 + \left(\frac{C(p, 2)}{n (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right)^2 \right) + O_p(1/n) \\
&= O_p \left(\frac{1}{n} + \left(\frac{C(p, 1)}{n \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 + \left(\frac{C(p, 2)}{n (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right)^2 \right).
\end{aligned}$$

F.3.1. Orden general

Por (F.1) podemos concluir

$$\begin{aligned}
(\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \boldsymbol{\Sigma} (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta}) &= (1 + O_p(J))^2 \left[O_p \left(\frac{1}{n} + \frac{C(p, 2)}{n\Omega^2} \right) \right. \\
&\quad + O_p \left(\frac{1}{n} + \frac{C(p, 2)}{n\Omega^2} \right) O_p(K^2) \\
&\quad + O_p(K^2/n) + O_p(1/n^2) \\
&\quad \left. + O_p \left(\frac{1}{n} + \left(\frac{C(p, 1)}{n\boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 + \left(\frac{C(p, 2)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right)^2 \right) \right] \\
&= O_p \left(\frac{1}{n} + \left(\frac{C(p, 1)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \right)^2 + \frac{C(p, 2)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right).
\end{aligned}$$

APÉNDICE G

DEMOSTRACIÓN DEL TEOREMA 3.6

Para probar el resultado, observemos que

$$\begin{aligned}
\frac{\widehat{V}}{V} &= \frac{\widehat{\Omega}_{pls}^{-2} \mathbf{G}^T (\widehat{\Sigma} \widehat{\sigma}_Y^2 - \widehat{\sigma} \widehat{\sigma}^T) \mathbf{G}}{\Omega^{-2} \mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}} \\
&= \frac{\mathbf{G}^T (\widehat{\Sigma} \sigma_Y^2 - \widehat{\sigma} \widehat{\sigma}^T) \mathbf{G}}{\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}} \\
&\quad + \left(\frac{\widehat{\Omega}_{pls}^{-2}}{\Omega^{-2}} - 1 \right) \frac{\mathbf{G}^T (\widehat{\Sigma} \sigma_Y^2 - \widehat{\sigma} \widehat{\sigma}^T) \mathbf{G}}{\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}} \\
&\quad + \frac{\widehat{\Omega}_{pls}^{-2}}{\Omega^{-2}} (\widehat{\sigma}_Y^2 - \sigma_Y^2) \frac{\mathbf{G}^T \widehat{\Sigma} \mathbf{G}}{\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}} \\
&= I + II + III.
\end{aligned}$$

Vamos a probar que $I \xrightarrow{p} 1$ y $II, III \xrightarrow{p} 0$ cuando $n, p \rightarrow \infty$.

Para III , observemos primero que por (3.6) tenemos que $\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G} \asymp \mathbf{G}^T \Sigma \mathbf{G}$, y como $m \frac{\mathbf{G}^T \widehat{\Sigma} \mathbf{G}}{\mathbf{G}^T \Sigma \mathbf{G}} \sim \chi_m^2$ por Lema B.3, entonces $E \left(\frac{\mathbf{G}^T \widehat{\Sigma} \mathbf{G}}{\mathbf{G}^T \Sigma \mathbf{G}} \right) = 1$ por Lema B.1 y

$$\frac{\mathbf{G}^T \widehat{\Sigma} \mathbf{G}}{\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}} = O_p(1).$$

Este hecho junto con $\frac{\widehat{\Omega}_{pls}^{-2}}{\Omega^{-2}} = O_p(1)$ por (A.11) y que $\widehat{\sigma}_Y^2 - \sigma_Y^2 = o_p(1)$, nos dicen que

$$III = \frac{\widehat{\Omega}_{pls}^{-2}}{\Omega^{-2}} (\widehat{\sigma}_Y^2 - \sigma_Y^2) \frac{\mathbf{G}^T \widehat{\Sigma} \mathbf{G}}{\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}} = o_p(1).$$

Observemos ahora que si probamos $I \xrightarrow{p} 1$, entonces $II \xrightarrow{p} 0$ se sigue, pues $\frac{\widehat{\Omega}_{pls}^{-2}}{\Omega^{-2}} - 1 = O_p(J)$ por (A.11).

Vamos a probar que $I = 1 + O_p(m^{-1/2})$. Por Lema C.3 y Lema B.2,

$$\begin{aligned} E\left(\mathbf{G}^T(\widehat{\Sigma}\sigma_Y^2 - \widehat{\sigma}\widehat{\sigma}^T)\mathbf{G}\right) &= \mathbf{G}^T \Sigma \mathbf{G} \sigma_Y^2 - \mathbf{G}^T \left(\frac{1}{m} \Sigma \sigma_Y^2 + \left(1 + \frac{1}{m}\right) \sigma \sigma^T \right) \mathbf{G} \\ &= \mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G} + \left(-\frac{1}{m} \mathbf{G}^T (\Sigma \sigma_Y^2 + \sigma \sigma^T) \mathbf{G} \right), \end{aligned}$$

y como $\frac{\mathbf{G}^T (\Sigma \sigma_Y^2 + \sigma \sigma^T) \mathbf{G}}{\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G}} = O(1)$ tenemos

$$E(I) = 1 + O(m^{-1}). \quad (\text{G.1})$$

Ahora vamos a acotar la varianza de I .

$$\text{var}\left(\mathbf{G}^T(\widehat{\Sigma}\sigma_Y^2 - \widehat{\sigma}\widehat{\sigma}^T)\mathbf{G}\right) = O\left(\text{var}(\mathbf{G}^T \widehat{\Sigma} \mathbf{G}) + \text{var}((\widehat{\sigma}^T \mathbf{G})^2)\right).$$

Recordando que $m \frac{\mathbf{G}^T \widehat{\Sigma} \mathbf{G}}{\mathbf{G}^T \Sigma \mathbf{G}} \sim \chi_m^2$, tenemos que $\text{var}(\mathbf{G}^T \widehat{\Sigma} \mathbf{G}) = \frac{2}{m} (\mathbf{G}^T \Sigma \mathbf{G})^2$. Para el otro término, como $\widehat{\sigma} = \frac{\mathbf{W} \beta + \mathbf{F} \varepsilon}{m}$

$$\text{var}((\widehat{\sigma}^T \mathbf{G})^2) = \frac{1}{m^4} O\left(\text{var}(\varepsilon^T \mathbf{F}^T \mathbf{G} \mathbf{G}^T \mathbf{F} \varepsilon) + \text{var}(\mathbf{G}^T \mathbf{W} \beta \beta^T \mathbf{W} \mathbf{G})\right).$$

Usando Lema B.2, Lema B.4 y Corolario C.2

$$\begin{aligned} \text{var}(\varepsilon^T \mathbf{F}^T \mathbf{G} \mathbf{G}^T \mathbf{F} \varepsilon) &= O(\tau^4 E((\mathbf{G}^T \mathbf{W} \mathbf{G})^2)) = O(\tau^4 m(m+2)(\mathbf{G}^T \Sigma \mathbf{G})^2) \\ &= O(m^2 (\mathbf{G}^T \Sigma \mathbf{G})^2). \end{aligned}$$

Para la segunda parte

$$\begin{aligned} \mathbf{G}^T \mathbf{W} \beta \beta^T \mathbf{W} \mathbf{G} &= \mathbf{G}^T (\mathbf{P}_\sigma + \mathbf{Q}_\sigma) \mathbf{W} \beta \beta^T \mathbf{W} (\mathbf{P}_\sigma + \mathbf{Q}_\sigma) \mathbf{G} \\ &= (\sigma^T \mathbf{G})^2 \left(\frac{\beta^T \mathbf{W} \beta}{\beta^T \Sigma \beta} \right)^2 + 2(\sigma^T \mathbf{G}) \left(\frac{\beta^T \mathbf{W} \beta}{\beta^T \Sigma \beta} \right) (\beta^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{G}) \\ &\quad + (\beta^T \Sigma \beta) \left(\frac{\beta^T \mathbf{W} \beta}{\beta^T \Sigma \beta} \right) \mathbf{G}^T \left(\frac{\mathbf{Q}_\sigma \mathbf{W} \beta \beta^T \mathbf{W} \mathbf{Q}_\sigma}{\beta^T \mathbf{W} \beta} \right) \mathbf{G} \\ &= (\sigma^T \mathbf{G})^2 N^2 + 2(\sigma^T \mathbf{G}) N (\beta^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{G}) + (\beta^T \Sigma \beta) N \mathbf{G}^T \mathbf{T} \mathbf{G}, \end{aligned}$$

y por lo tanto

$$\begin{aligned} \text{var}(\mathbf{G}^T \mathbf{W} \beta \beta^T \mathbf{W} \mathbf{G}) &= O\left(\text{var}((\sigma^T \mathbf{G})^2 N^2) + \text{var}((\beta^T \Sigma \beta) N \mathbf{G}^T \mathbf{T} \mathbf{G})\right. \\ &\quad \left. + \text{var}((\sigma^T \mathbf{G}) N \beta^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{G})\right). \end{aligned}$$

Ahora, por el Corolario C.2 y el Lema B.1 tenemos

$$\text{var}((\sigma^T \mathbf{G})^2 N^2) = (\sigma^T \mathbf{G})^4 O(m^3) = O(m^3 (\mathbf{G}^T \Sigma \mathbf{G})^2).$$

Nuevamente por el Corolario C.2, la independencia de \mathbf{T} y N , y el hecho de que $\frac{\mathbf{G}^T \mathbf{T} \mathbf{G}}{\mathbf{G}^T \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \mathbf{G}} \sim \chi_1^2$,

$$\begin{aligned} \text{var}((\beta^T \Sigma \beta) N \mathbf{G}^T \mathbf{T} \mathbf{G}) &= (\beta^T \Sigma \beta)^2 (\mathbf{G}^T \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \mathbf{G})^2 \left(\text{var}(N) \text{var} \left(\frac{\mathbf{G}^T \mathbf{T} \mathbf{G}}{\mathbf{G}^T \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \mathbf{G}} \right) \right. \\ &\quad \left. + \text{var}(N) E^2 \left(\frac{\mathbf{G}^T \mathbf{T} \mathbf{G}}{\mathbf{G}^T \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \mathbf{G}} \right) + E^2(N) \text{var} \left(\frac{\mathbf{G}^T \mathbf{T} \mathbf{G}}{\mathbf{G}^T \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \mathbf{G}} \right) \right) \\ &= O(m^2 (\mathbf{G}^T \Sigma \mathbf{G})^2). \end{aligned}$$

Para los otros, $E(N \beta^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{G}) = 0$, así que

$$\begin{aligned} \text{var}((\sigma^T \mathbf{G}) N \beta^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{G}) &= (\sigma^T \mathbf{G})^2 E(N^2 \mathbf{G}^T \mathbf{Q}_\sigma \mathbf{W} \beta \beta^T \mathbf{W} \mathbf{Q}_\sigma \mathbf{G}) \\ &= (\sigma^T \mathbf{G})^2 (\beta^T \Sigma \beta) (\mathbf{G}^T \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \mathbf{G}) E(N^3) E \left(\frac{\mathbf{G}^T \mathbf{T} \mathbf{G}}{\mathbf{G}^T \mathbf{Q}_\sigma \Sigma \mathbf{Q}_\sigma \mathbf{G}} \right) \\ &= O(m^3 (\mathbf{G}^T \Sigma \mathbf{G})^2). \end{aligned}$$

Como consecuencia

$$\text{var}(I) = \frac{1}{(\mathbf{G}^T (\Sigma \sigma_Y^2 - \sigma \sigma^T) \mathbf{G})^2} O \left(\frac{(\mathbf{G}^T \Sigma \mathbf{G})^2}{m} \right) = O(1/m),$$

y podemos concluir que $I = 1 + O_p(m^{-1/2})$, lo cual termina la prueba.

APÉNDICE H

DEMOSTRACIÓN DEL TEOREMA 3.7

Vamos a mostrar primero que si $\sqrt{n}(\hat{b} - b) \xrightarrow{p} 0$, entonces podemos usar \hat{b} para obtener un intervalo de confianza para $\beta^T \mathbf{G}$. Observemos que por el Teorema 3.3 y el Teorema 3.6 tenemos que

$$\hat{V}^{-1/2}(\hat{\beta}_{pls} - (1 + b)\beta)^T \mathbf{G} \rightsquigarrow N(0, 1),$$

entonces si mostramos que

$$\hat{V}^{-1/2}(\hat{b} - b)\beta^T \mathbf{G} \xrightarrow{p} 0$$

queda claro que

$$\hat{V}^{-1/2}(\hat{\beta} - (1 + \hat{b})\beta)^T \mathbf{G} \rightsquigarrow N(0, 1),$$

y el resultado se sigue. Recordemos que por (3.11) $V^{-1/2}|\beta^T \mathbf{G}| \leq \sqrt{n}$, tenemos

$$\hat{V}^{-1/2}|\hat{b} - b||\beta^T \mathbf{G}| = \left(\frac{\hat{V}}{V}\right)^{-1/2} \frac{V^{-1/2}|\beta^T \mathbf{G}|}{\sqrt{n}} \sqrt{n}|\hat{b} - b| \xrightarrow{p} 0.$$

Necesitamos probar que bajo las condiciones de la Proposición 3.2 y la condición (II) del Teorema 3.3,

$$\sqrt{n}(\hat{b} - b) \xrightarrow{p} 0,$$

donde

$$\hat{b} = -(\hat{\sigma}_Y^2 - 2\hat{\tau}^2) \frac{\hat{C}(p, 1)}{m\hat{\sigma}^T \hat{\sigma}} - \hat{\sigma}_Y^2 \frac{\hat{C}^2(p, 1)}{m^2 \hat{\Omega}_{pls} \hat{\sigma}^T \hat{\sigma}} - \hat{\sigma}_Y^2 \frac{\hat{C}(p, 2)}{m \hat{\Omega}_{pls} \hat{\sigma}^T \hat{\sigma}}$$

con $\hat{C}(p, k) = \text{tr}(\hat{\Sigma}^k) - \hat{\Omega}_{pls}^k$. Vamos a demostrar que cuando $n, p \xrightarrow{p} \infty$,

$$(i) \quad \sqrt{n} \left(\frac{\hat{C}(p, 1)}{m\hat{\sigma}^T \hat{\sigma}} - \frac{C(p, 1)}{m\sigma^T \sigma} \right) \xrightarrow{p} 0,$$

$$(ii) \quad \sqrt{n} \left(\frac{\widehat{C}(p, 1)^2}{m^2 \widehat{\Omega}_{pls} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 1)^2}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \xrightarrow{p} 0,$$

$$(iii) \quad \sqrt{n} \left(\frac{\widehat{C}(p, 2)}{m \widehat{\Omega}_{pls} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 2)}{m \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \xrightarrow{p} 0.$$

De estos resultados, se sigue que $\frac{\widehat{C}(p, 1)}{m \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}$, $\frac{\widehat{C}^2(p, 1)}{m^2 \widehat{\Omega}_{pls} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}$ y $\frac{\widehat{C}(p, 2)}{m \widehat{\Omega}_{pls} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}$ convergen a 0. Luego, como vale $b \rightarrow 0$ y también $\hat{\sigma}_Y^2$ y $\widehat{\tau}$ son \sqrt{n} -consistentes, tenemos el resultado.

H.1. Demostración de (i)

Por definición de $\widehat{C}(p, 1)$ y $C(p, 1) = \text{tr}(\boldsymbol{\Sigma}) + \Omega$,

$$\begin{aligned} \sqrt{n} \left(\frac{\widehat{C}(p, 1)}{m \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) &= \sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} \left(\frac{\text{tr}(\widehat{\boldsymbol{\Sigma}}) - \text{tr}(\boldsymbol{\Sigma})}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \\ &\quad + \sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} \left(\frac{\widehat{\Omega}_{pls} - \Omega}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \\ &\quad + \sqrt{n} \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - 1 \right). \end{aligned} \tag{H.1}$$

Para el primer término, por el Lema B.2 y el hecho de que $m \widehat{\boldsymbol{\Sigma}} \sim \text{Wish}(m, \boldsymbol{\Sigma})$, sabemos que

$$\text{tr}(\widehat{\boldsymbol{\Sigma}}) - \text{tr}(\boldsymbol{\Sigma}) = O_p \left(\left(\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{m} \right)^{1/2} \right) = O_p \left(\frac{\Omega}{\sqrt{m}} + \left(\frac{C(p, 2)}{m} \right)^{1/2} \right).$$

Por Teorema A.1, tenemos $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - 1 = O_p(K)$. Por lo tanto

$$\sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} \left(\frac{\text{tr}(\widehat{\boldsymbol{\Sigma}}) - \text{tr}(\boldsymbol{\Sigma})}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) = O_p \left(\frac{\Omega}{n \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \frac{1}{\sqrt{n}} \left(\frac{C(p, 2)}{n (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right)^{1/2} \right).$$

Para el segundo término, por (A.11) tenemos que $\frac{\widehat{\Omega}_{pls}}{\Omega} - 1 = O_p(J)$, así que

$$\sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} \left(\frac{\widehat{\Omega}_{pls} - \Omega}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) = O_p \left(\frac{\Omega}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) O_p(J).$$

Para el tercer término

$$\begin{aligned} \sqrt{n} \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - 1 \right) &= O \left(\frac{C(p, 1)}{n^{3/4} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) O_p(n^{1/4} K) \\ &= O_p \left(\left(\frac{\Omega}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^{1/2} \frac{C(p, 1)}{n^{3/4} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} + \left(\frac{C(p, 1)}{n^{3/4} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 \right). \end{aligned}$$

Entonces por las condiciones (II) del Teorema 3.3, $\frac{\Omega}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} = O(1)$ y $\frac{C(p, 1)}{n^{3/4} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \xrightarrow{p} 0$, y como $J \rightarrow 0$, todos los términos de (H.1) tienden a 0.

H.2. Demostración de (ii)

Recordando que $\widehat{\Omega}_{pls} = \frac{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}}$ tenemos

$$\begin{aligned} \sqrt{n} \left(\frac{\widehat{C}(p, 1)^2}{m^2 \widehat{\Omega}_{pls} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 1)^2}{m^2 \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) &= \sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \left(\frac{\widehat{C}(p, 1)}{m \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 \left(\frac{(\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}})^2 \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right) \\ &\quad + 2\sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \left(\frac{\widehat{C}(p, 1)}{m \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) \left(\frac{(\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}})^2 \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right) \\ &\quad + \sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \left(\frac{C(p, 1)}{m \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 \left(\frac{(\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}})^2 \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} - 1 \right). \end{aligned}$$

No es difícil ver de (A.11) que $\frac{(\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}})^2 \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} - 1 = O_p(J)$. Esto, junto con el hecho de que $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$ y $\frac{C(p, 1)}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} = O(1)$ por la condición (II) del Teorema 3.3, y por (i), concluimos (ii).

H.3. Demostración de (iii)

Por la definición $\widehat{C}(p, 2)$, junto con $C(p, 2) = \text{tr}(\boldsymbol{\Sigma}^2) + \Omega^2$, y observando que $\widehat{\Omega}_{pls} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}} = \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}$,

$$\begin{aligned} \sqrt{n} \left(\frac{\widehat{C}(p, 2)}{m \widehat{\Omega}_{pls} \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 2)}{m \Omega \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) &= \sqrt{n} \left(\frac{\widehat{C}(p, 2)}{m \widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} - \frac{C(p, 2)}{m \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right) \\ &= \sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} \left(\frac{\text{tr}(\widehat{\boldsymbol{\Sigma}}^2) - \text{tr}(\boldsymbol{\Sigma}^2)}{m \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right) \\ &\quad + \sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} \left(\frac{\widehat{\Omega}_{pls}^2 - \Omega^2}{m \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right) \tag{H.2} \\ &\quad + \sqrt{n} \frac{C(p, 2)}{m (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} - 1 \right). \end{aligned}$$

Para el primer término, por el Lema B.2 y el hecho de que $m \widehat{\boldsymbol{\Sigma}} \sim \text{Wish}(m, \boldsymbol{\Sigma})$, sabemos que

$$\begin{aligned} \text{tr}(\widehat{\boldsymbol{\Sigma}}^2) - \text{tr}(\boldsymbol{\Sigma}^2) &= O_p \left(\frac{\text{tr}^2(\boldsymbol{\Sigma})}{m} + \frac{\text{tr}(\boldsymbol{\Sigma}^2)}{m} + \left(\frac{\text{tr}(\boldsymbol{\Sigma}^3) \text{tr}(\boldsymbol{\Sigma})}{m^2} \right)^{1/2} + \left(\frac{\text{tr}(\boldsymbol{\Sigma}^4)}{m} \right)^{1/2} \right) \\ &= O_p \left(\frac{\Omega^2}{\sqrt{n}} + \frac{C(p, 1)^2}{n} + \frac{C(p, 2)}{n} + \left(\frac{C(p, 3) C(p, 1)}{n^2} \right)^{1/2} \right. \\ &\quad \left. + \left(\frac{\Omega^3 C(p, 1)}{n^2} \right)^{1/2} + \left(\frac{\Omega C(p, 3)}{n^2} \right)^{1/2} + \left(\frac{C(p, 4)}{n} \right)^{1/2} \right). \end{aligned}$$

Por (A.11), tenemos que $\frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} - 1 = O_p(J)$. Por lo tanto usando los resultados $\frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} = O(1)$, $\frac{\Omega}{n \boldsymbol{\sigma}^T \boldsymbol{\sigma}} = O(1)$ y $\frac{C(p, k)}{n (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^k} = O(1)$ para $k = 1, 2, 3, 4$, se sigue que

$$\sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} \left(\frac{\text{tr}(\widehat{\boldsymbol{\Sigma}}^2) - \text{tr}(\boldsymbol{\Sigma}^2)}{m \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right) = O_p \left(\left(\frac{\Omega}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 + \left(\frac{C(p, 1)}{n^{3/4} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^2 + \frac{C(p, 2)}{n^{3/4} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} + \frac{1}{\sqrt{n}} \right).$$

Para el segundo término, por (A.11) tenemos que $\frac{\widehat{\Omega}_{pls}^2}{\Omega^2} - 1 = O_p(J)$, así que

$$\sqrt{n} \frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} \left(\frac{\widehat{\Omega}_{pls}^2 - \Omega^2}{m \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}} \right) = O_p \left(\frac{\Omega}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right) O_p(J).$$

Para el tercer término,

$$\begin{aligned} \sqrt{n} \frac{C(p, 2)}{m (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \frac{\boldsymbol{\sigma}^T \boldsymbol{\sigma}}{\Omega} \left(\frac{\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma}}{\widehat{\boldsymbol{\sigma}}^T \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\sigma}}} - 1 \right) &= O \left(\frac{C(p, 2)}{n^{3/4} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right) O_p(n^{1/4} J) \\ &= O_p \left(\left(\frac{\Omega}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \right)^{1/2} \frac{C(p, 2)}{n^{3/4} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right. \\ &\quad \left. + \frac{C(p, 1)}{n^{3/4} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} \frac{C(p, 2)}{n^{3/4} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} + \left(\frac{C(p, 2)}{n^{3/4} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right)^2 \right). \end{aligned}$$

Entonces por las condiciones (II) del Teorema 3.3, $\frac{\Omega}{\sqrt{n} \boldsymbol{\sigma}^T \boldsymbol{\sigma}} = O(1)$ y $\frac{C(p, 1)}{n^{3/4} \boldsymbol{\sigma}^T \boldsymbol{\sigma}}, \frac{C(p, 2)}{n^{3/4} (\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \xrightarrow{p} 0$, y como $J \rightarrow 0$, todos los términos de (H.2) convergen a 0 y concluimos (iii).

APÉNDICE I

DEMOSTRACIÓN DEL TEOREMA 3.8

I.1. Demostración del Lema 3.9

Por la normalidad conjunta de (\mathbf{X}, Y) se tiene $\widehat{\boldsymbol{\beta}}_{pls} \perp \bar{\mathbf{X}}$, luego tenemos que $E((\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}})) = 0$. Más aún, podemos proceder como en el Teorema 3.5

$$\text{var}((\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}})) = \frac{1}{n} E((\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T \boldsymbol{\Sigma} (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})),$$

y por Teorema 3.5 tenemos

$$(\tau^2/n)^{-1} (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}}) (\bar{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}})^T (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta}) = O_p \left(\frac{1}{n} + \left(\frac{C(p, 1)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})} \right)^2 + \frac{C(p, 2)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right),$$

y como $\omega^{-1}(\mathbf{X}_N) \leq (\tau^2/n)^{-1}$, tenemos que

$$\omega^{-1/2}(\mathbf{X}_N) (\widehat{\boldsymbol{\beta}}_{pls} - \boldsymbol{\beta})^T (\bar{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}}) = O_p \left(\frac{1}{\sqrt{n}} + \frac{C(p, 1)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})} + \left(\frac{C(p, 2)}{n(\boldsymbol{\sigma}^T \boldsymbol{\sigma})^2} \right)^{1/2} \right)$$

que tiende a 0.

I.2. Demostración del Lema 3.11

Observemos que por compacidad de la bola unitaria en \mathbb{R}^2 para cada subsucesión $(a_{n_k, p_k}, b_{n_k, p_k})$ existe (a, b) con $a^2 + b^2 = 1$ y una subsucesión $(a_{n_{k_j}, p_{k_j}}, b_{n_{k_j}, p_{k_j}}) \rightarrow (a, b)$.

Entonces como $\begin{pmatrix} Z_{n,p} \\ W_{n,p} \end{pmatrix} \rightsquigarrow N_2(\mathbf{0}, \mathbf{I})$ tenemos que

$$a_{n_{k_j}, p_{k_j}} Z_{n_{k_j}, p_{k_j}} + b_{n_{k_j}, p_{k_j}} W_{n_{k_j}, p_{k_j}} = (a_{n_{k_j}, p_{k_j}}, b_{n_{k_j}, p_{k_j}}) \begin{pmatrix} Z_{n,p} \\ W_{n,p} \end{pmatrix} \rightsquigarrow N(0, 1),$$

ya que $(a_{n_{k_j}, p_{k_j}}, b_{n_{k_j}, p_{k_j}}) \mathbf{I} \begin{pmatrix} a_{n_{k_j}, p_{k_j}} \\ b_{n_{k_j}, p_{k_j}} \end{pmatrix} = 1$.

Por lo tanto para cada $t \in \mathbb{R}$, consideremos $y_{n,p} = E \left(e^{it(a_{n,p}Z_{n,p} + b_{n,p}W_{n,p})} \right)$. Para cada subsucesión y_{n_k, p_k} hemos probado que existe una subsucesión $y_{n_{k_j}, p_{k_j}} \rightarrow e^{-\frac{t^2}{2}}$. Esto implica que $y_{n,p} \rightarrow e^{-\frac{t^2}{2}}$ y el resultado se sigue.