



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**REGRESION POR MÍNIMOS CUADRADOS  
PARCIALES**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

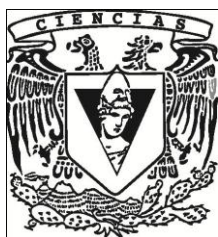
**ACTUARIO**

**P R E S E N T A:**

**YISEL GUADALUPE OLIVO IBARRA**

**TUTORA  
MATEMÁTICA**

**MARGARITA ELVIRA CHÁVEZ CANO  
2010**





Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno  
Apellido paterno  
Apellido materno  
Nombre (s)  
Teléfono  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Carrera  
Número de cuenta

2. Datos del tutor  
Grado  
Nombre (s)  
Apellido paterno  
Apellido materno

3. Datos del Sinodal 1  
Grado  
Nombre (s)  
Apellido paterno  
Apellido materno

4. Datos del sinodal 2  
Grado  
Nombre (s)  
Apellido paterno  
Apellido materno

5. Datos del sinodal 3  
Grado  
Nombre (s)  
Apellido paterno  
Apellido materno

6. Datos del sinodal 4  
Grado  
Nombre (s)  
Apellido paterno  
Apellido materno

7. Datos del trabajo escrito  
Título  
Subtítulo  
Número de páginas  
Año

1. Datos del alumno  
Olivo  
Ibarra  
Yisel Guadalupe  
56 44 92 04  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
303262585

2. Datos del tutor  
Mat  
Margarita Elvira  
Chavéz  
Cano

3. Datos del Sinodal 1  
Act  
Jaime  
Vázquez  
Alamilla

4. Datos del sinodal 2  
Dra  
Ruth Selene  
Fuentes  
García

5. Datos del sinodal 3  
M en A P  
María del Pilar  
Alonso  
Reyes

6. Datos del sinodal 4  
M en C  
José Antonio  
Flores  
Díaz

7. Datos del trabajo escrito.  
Regresión por mínimos cuadrados  
parciales  
120 p  
2011

# DEDICATORIAS

**A MI PADRE**, por su gran esfuerzo para hacer posible este sueño, por sus consejos, por enseñarme que la inspiración es la mejor compañía para un estudiante, que la pasión debe acompañar cualquier meta y sobre todo, por enseñarme la pasión por las matemáticas.

¡Gracias papi, este logro es de los dos!

**A MI MADRE**, por sus grandes consejos y gran apoyo, por su dedicación y comprensión en todo momento, por su ayuda incondicional en todos los aspectos de mi vida.

¡Gracias por ser mi madre!

**A LA UNAM**, por ser la casa máxima de estudios, al concentrar a los mejores profesionistas de todas las áreas.

**A MARGARITA CHÁVEZ**, por pertenecer a la UNAM, ser mi mejor aliada en el cumplimiento de este sueño y por su paciencia, dedicación y pasión en todas sus enseñanzas; por convertirse en este periodo, en mi amiga y confidente para el resto de mi vida.

**A TODOS MIS PROFESORES**, que en cada clase me enseñaron que la perseverancia es la mejor arma para salir adelante, y que nunca es demasiado, pues falta mucho por aprender.

# ÍNDICE

Introducción.....	3
Objetivo.....	5
<b>1. REGRESIÓN LINEAL MÚLTIPLE (RLM)</b>	<b>6</b>
1.1 Introducción.....	6
1.2 Fundamento matemático.....	6
1.3 Propiedades de $\hat{\beta}$ , el estimador por mínimos cuadrados.....	13
1.4 El vector $\underline{y}$ y el vector de residuales $\underline{e}$ .....	14
1.5 Estimación del parámetro $\sigma^2$ .....	16
1.6 Modelo lineal centrado.....	17
<b>2. EL PROBLEMA DE MULTICOLINEALIDAD</b>	<b>19</b>
2.1 Multicolinealidad.....	19
2.1.1 Detección de la multicolinealidad.....	21
2.2 Pre-procesamiento de los datos.....	24
<b>3. MÉTODOS DE REGRESIÓN POR VARIABLES LATENTES (RVL)</b>	<b>26</b>
3.1 Introducción.....	26
3.2 Análisis de componentes principales (ACP).....	29
3.2.1 Introducción.....	29
3.2.2 Invarianza de escala.....	30
3.2.3 Fundamento matemático.....	31
3.2.4 Selección del número de componentes principales.....	35

3.2.5 Algoritmo NIPALS .....	37
3.2.6 Regresión por componentes principales (RCP).....	39
<b>4. MÍNIMOS CUADRADOS PARCIALES (MCP)</b>	<b>40</b>
4.1 Introducción.....	40
4.2 Mínimos cuadrados parciales univariado (MCP1).....	42
4.2.1 Invarianza de escala.....	43
4.2.2 Planteamiento matemático.....	43
4.2.3 Regresión por mínimos cuadrados parciales univariado (RMCP1).....	51
4.2.4 Predicción.....	53
4.2.5 Propiedades básicas de MCP1.....	53
4.3 Mínimos cuadrados parciales multivariado (MCP2).....	64
4.3.1 Planteamiento matemático.....	67
4.3.2 Propiedades básicas de MCP2.....	73
4.3.3 Regresión por mínimos cuadrados parciales multivariado (RMCP2).....	82
4.3.4 Predicción.....	83
4.3.5 Intervalos de confianza para $\underline{\beta}_{MCP}$ y $\underline{y}_{MCP}$ .....	84
4.4 Selección del modelo.....	84
4.4.1 Por capacidad predictiva.....	84
4.4.2 Por variabilidad explicada por las primeras $j$ – variables latentes en el modelo de RMCP.....	86
4.4.3 Interpretación de los resultados obtenidos por MCP.....	87
4.5 Otras variantes de MCP.....	88
4.6 Consideraciones finales sobre MCP.....	90
4.7 Un breve ejemplo de MCP1 en R.....	93
4.8 Software para MCP.....	104
<b>COMENTARIOS FINALES</b> .....	<b>105</b>
<b>ANEXO A</b>	<b>106</b>
<b>ANEXO B</b>	<b>109</b>
<b>REFERENCIAS BIBLIOGRÁFICAS</b>	<b>113</b>
<b>REFERENCIAS ELECTRÓNICAS</b>	<b>121</b>

# Introducción

En cualquier rama de la ciencia, la necesidad básica de un investigador es entender de forma clara el problema a resolver, identificando los factores que lo generen y que, al ser manipulados se encuentren los criterios para dar prevención, solución o predicción, según sea el objetivo. La metodología estadística es aplicada exitosamente para modelar muchos fenómenos y por ende, se le considera la tecnología del método científico. Apoyándose de la teoría de la probabilidad, la estadística desarrolla respuestas ante situaciones de incertidumbre, y lo hace basándose en información histórica del fenómeno para después entrever su futuro.

La técnica estadística más conocida y utilizada para analizar y predecir el comportamiento de algún fenómeno es la de Regresión Lineal Múltiple (RLM), que rinde los mejores resultados de predicción, sin embargo es una técnica restrictiva pues su uso se limita a datos con ciertas características: no debe existir multicolinealidad entre las variables regresoras, de lo contrario se dice que la información es redundante, y debe de contar con un mayor número de observaciones que variables regresoras. Por un lado, desde finales del siglo XX y hasta el día de hoy, los avances tecnológicos han desarrollado poderosas herramientas computacionales que con bajo costo, espacio y tiempo, generan una gran cantidad de datos listos para su análisis; la desventaja es que muchas veces exceso de información no siempre es sinónimo de información útil sino todo lo contrario, información redundante; por otro lado, existen fenómenos que debido a su naturaleza presentan datos históricos limitados y aunado a esto, se explican por varias o muchas variables regresoras. La RLM ofrece bajo ajuste y predicciones no confiables con este tipo de datos; entonces se puede decir que el proceso de evolución de las herramientas tecnológicas trae consigo la necesidad de desarrollar nuevas técnicas estadísticas de análisis y manipulación de datos.

Herman Wold en la década de los sesenta desarrolló la técnica de Mínimos Cuadrados Parciales<sup>1</sup> (MCP), como una técnica económica, con la finalidad de modelar datos multicolineales y con alta dimensionalidad de variables regresoras. La técnica de MCP nació siguiendo ideas heurísticas de parte de su creador, quien al principio implementó

---

<sup>1</sup> Partial Least Squares (PLS).

la técnica de Análisis de Componentes Principales (ACP) en el algoritmo Non-linear Partial Least Squares (NIPALS), después modificó algunos pasos e insertó otros, para desarrollar MCP univariado mejor conocido como NIPALS PLS1 y la versión multivariada de MCP, NIPALS PLS2, esta última considera varias variables de respuesta, por esta razón contempla un algoritmo más complejo y con mayor número de pasos.

Debido a la carencia de rigurosidad matemática en el planteamiento de MCP, es un método desconocido para muchos estadísticos por catalogarlo como un simple algoritmo más que un modelo estadístico riguroso<sup>2</sup>, de hecho, la mayor parte de la literatura que lo expone lo hace como tal, sin mayor especificación de los pasos, sin embargo varios autores<sup>3</sup> se han dado a la tarea de investigar sus propiedades estadísticas.

De manera general, MCP combina algunos aspectos de ACP y RLM; la idea básica es llevar la dependencia lineal entre las variables regresoras y las de respuesta a través de otras, comúnmente llamadas variables latentes, las cuales son elegidas por explicar lo más exacto posible la estructura común entre las variables originales, para posteriormente implementar la RLM utilizando las variables latentes como nuevas variables regresoras, dando origen al modelo de regresión por mínimos cuadrados parciales (RMCP). A esta técnica, se le ha clasificado como de “modelado suave<sup>4</sup>” por sus bondades para ajustar datos con estas características. MCP pertenece a los métodos de regresión por variables latentes (RVL), los cuales son conceptualmente muy diferentes al modelo de regresión tradicional.

A pesar de ser relativamente reciente y creada para las ciencias sociales, la popularidad de MCP se ha incrementando en diversas disciplinas tales como psicología, medicina, farmacología, ecología, etc., aunque sus principales proponentes son los quimiométricos, quienes la clasifican como su herramienta estándar.

Es importante señalar que MCP no es un solo algoritmo sino una clase de algoritmos, pues los pasos del algoritmo original pueden ser modificados, dando lugar a otras versiones que responden al nombre de MCP.

---

<sup>2</sup> Helland, I. (1988)

<sup>3</sup> Helland, I. (1988), Krämer, N. (2007), Lingjaerde, O. C. & Christophersen, N. (2000), etc.

<sup>4</sup> Pirouz, D. (2006)



# Objetivo

El objetivo de este trabajo es desarrollar y describir la técnica de MCP a través del algoritmo NIPALS para el caso univariado y para el multivariado, detallando los pasos de ambos mediante el fundamento matemático riguroso que representa la construcción de las variables latentes, así como enunciar las propiedades de los conjuntos originados por los algoritmos.

Ya que MCP combina características de la RLM y el ACP, en el capítulo 1 se dará una breve descripción de la RLM, con el fin de que el lector se familiarice con los conceptos fundamentales de regresión. En el capítulo 2, se describe detalladamente el problema de multicolinealidad así como los factores comunes que permiten detectarla; también se describe el pre-procesamiento a los datos, la cual es una etapa previa que se realiza antes de implementar un análisis de regresión. En el capítulo 3, se describe el fundamento de los métodos RVL. La técnica del ACP es un método RVL, por tanto se dará una descripción breve, pero detallada de ésta; utilizando NIPALS ACP, como algoritmo base para el desarrollo de la teoría de MCP.

El capítulo 4 está dividido fundamentalmente en dos partes, MCP1 y MCP2; se enuncia, desarrolla y describe, la teoría concerniente al método de MCP así como los criterios para la validación del modelo de RMCP. Un breve ejemplo de MCP1 es presentado en este capítulo, y se da una interpretación de los resultados.

# CAPÍTULO 1

## REGRESIÓN LINEAL MÚLTIPLE (RLM)

### 1.1 Introducción

En varios campos de la investigación, ya sea en las ciencias sociales como economía y psicología, en las médicas y de la salud, en la química e ingeniería entre otras, el método de regresión lineal múltiple es ampliamente utilizado; este permite llevar la relación lineal entre una variable dependiente (respuesta) y otras, denominadas variables independientes (regresoras).

### 1.2 Fundamento matemático

En concreto, la RLM pretende evidenciar las relaciones lineales entre variables respuesta y regresoras bajo una relación funcional, que puede ser expresada como:

$$y = f(x_1, x_2, \dots, x_p) \quad (1.1)$$

Sin embargo, de este planteamiento surgen dos problemas:

1. La forma analítica de  $f$  puede ser desconocida o bien, conocida pero muy complicada
2. El número de variables independientes que intervienen en la relación funcional puede ser tan grande que sea imposible manipular correctamente a  $f$ .

Existen alternativas para los problemas anteriores, en el primero se puede aproximar a  $f$  mediante una función auxiliar  $f'$ ; para el segundo, existen técnicas estadísticas que realizan una selección de variables al retener las consideradas relevantes para la

## Capítulo 1. Regresión Lineal Múltiple

descripción de la respuesta y desechar las que no. Un modelo sustituto que aproxima la relación funcional con la mayor exactitud está dado por:

$$y = f'(x_1, x_2, \dots, x_p) + \varepsilon \quad (1.2)$$

Donde se descompone a la variable  $y$ , en dos partes, la parte sistemática o causal  $f'(x_1, x_2, \dots, x_p)$  y la parte aleatoria o casual  $\varepsilon$ , que engloba aquellas variables que no fueron incluidas en la explicación de  $y$  al ser consideradas irrelevantes, en el planteamiento del modelo.

Dadas  $n$  unidades de observación para estas variables, consideremos a  $\mathbf{X}$  y a  $\underline{y}$  como la matriz y el vector, que comprenden dicha información. En regresión se pretende aproximar, de la manera más exacta posible, la dependencia lineal entre  $\mathbf{X}$  y  $\underline{y}$ ; si esta dependencia lineal fuera exacta, el modelo de regresión no tendría ningún sentido al resultar un sistema de ecuaciones determinístico  $\underline{y} = \mathbf{X}\underline{b}$ , por resolver. El modelo de RLM, estipula que la media verdadera de la variable dependiente cambia en razón constante cuando los valores de  $\{x_1, x_2, \dots, x_p\}$  crecen o decrecen, entonces:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (1.3)$$

Para la  $i$ -ésima observación se tiene:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad \text{para } i = 1, \dots, n$$

Matricialmente se escribe:

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon} \quad (1.4)$$
$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Donde:

$\underline{y}$  := Vector columna que recaba la información observada de la variable dependiente (respuesta) por cada unidad observacional  $y_i$ , con  $i=1,2,\dots,n$ .

$\mathbf{X}$  := Se le denomina matriz de datos; el primer vector columna es el vector de unos de orden  $n$ , más  $p$  vectores columna referentes a la información observada de las variables que describen a la variable de respuesta (variables regresoras).

$\underline{\beta}$  := Vector columna conformada por los parámetros que serán estimados a partir de los datos, mejor conocidos como coeficientes de regresión o sensibilidades.

$\underline{\varepsilon}$  := Vector que contiene los errores aleatorios para cada unidad observacional.

En el caso de una sola variable independiente, este modelo es conocido como el modelo de regresión lineal simple (RLS), expresado como  $y = \beta_0 + \beta_1 x + \varepsilon$ .

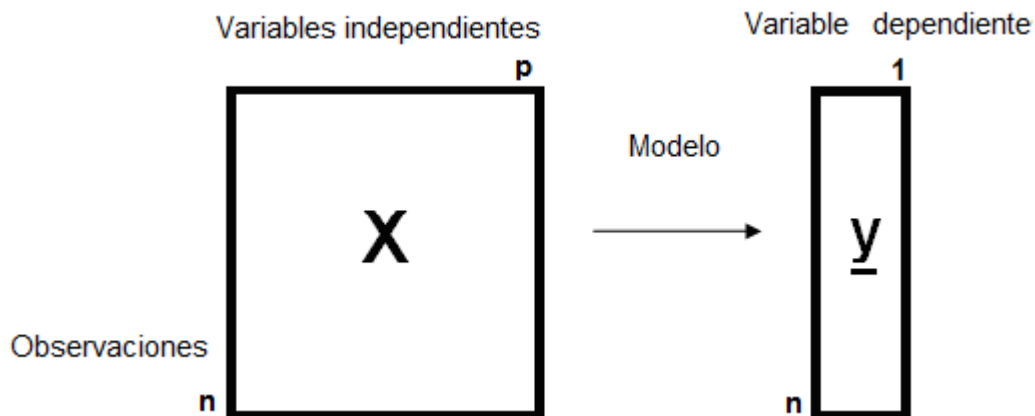


Figura 1<sup>1</sup>: Variables empleadas en la construcción del modelo de regresión.

Los supuestos adicionales que acompañan las ecuaciones del modelo son:

1. Mayor número de observaciones que variables regresoras ( $n \geq p$ ).
2. Las variables regresoras son no redundantes, entonces  $\mathbf{X}$  es una matriz de rango completo por columnas.
3. Las variables regresoras están relacionadas de manera lineal con la variable de respuesta  $y$ .

<sup>1</sup> Fuente: Elaboración propia.

## Capítulo 1. Regresión Lineal Múltiple

4. El vector aleatorio  $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$  tiene la distribución normal multivariada con los siguientes parámetros:

$$\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 \mathbf{I}_n) \quad (1.5)$$

Con matriz de varianzas y covarianzas  $\Sigma_{\underline{\varepsilon}\underline{\varepsilon}}$ , dada por:

$$\Sigma_{\underline{\varepsilon}\underline{\varepsilon}} = \begin{pmatrix} V(\varepsilon_1) & 0 & \dots & 0 \\ 0 & V(\varepsilon_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}_n \quad (1.6)$$

Con  $\mathbf{I}_n$  la matriz identidad de orden  $n$ . La propiedad 4 implica que  $\varepsilon_i \sim N(0, \sigma^2)$  con  $\sigma^2$  constante y desconocida y  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$  para  $i \neq j$ . Los supuestos sobre la parte aleatoria del ajuste a  $y$  en RLM, se interpretan de la siguiente manera: en promedio se espera no haya errores, por ser las variables regresoras estipuladas en (1.4) las suficientes y necesarias para la descripción de la variable de respuesta, y el supuesto de varianza constante, común y desconocida, significa que se espera que los errores no se distribuyan de manera irregular alrededor de la línea media. Las consecuencias directas de estos supuestos sobre el modelo de RLM son:

$$\underline{y} \sim N_n(\mathbf{X}\underline{\beta}, \sigma^2 \mathbf{I}_n) \quad (1.7)$$

Ya que:

$$E(\underline{y}) = E(\mathbf{X}\underline{\beta} + \underline{\varepsilon}) = E(\mathbf{X}\underline{\beta}) + E(\underline{\varepsilon}) = \mathbf{X}\underline{\beta} \quad (1.8)$$

Y para el cálculo de la varianza de  $\underline{y}$ , se tiene:

$$V(\underline{y}) = V(\mathbf{X}\underline{\beta} + \underline{\varepsilon}) = V(\underline{\varepsilon}) = \sigma^2 \mathbf{I}_n \quad (1.9)$$

Puesto que,  $\mathbf{X}$  es una matriz conocida y  $\underline{\beta}$  es un vector columna de parámetros a estimar. En la construcción del modelo, el vector de errores aleatorios justifica la

## 1.2 Fundamento matemático

omisión de algunas variables que no influyen significativamente en  $\underline{y}$ ; sin embargo, si la omisión es sobre variables importantes, entonces  $\underline{X}\underline{\beta}$  no es el valor esperado de  $\underline{y}$ .

La hipótesis de normalidad sobre  $\underline{\varepsilon}$ , permite realizar la teoría de intervalos de confianza, así como pruebas de significancia para la variable de respuesta, siguiendo la distribución dada en (1.7).

Del modelo dado en (1.4), se tienen que estimar a partir de los datos  $p+2$  parámetros, los cuales son: las  $p+1$  entradas del vector de coeficientes de regresión  $\underline{\beta}$  y el parámetro  $\sigma^2$ . El método de estimación de parámetros que se considerará en esta tesis es el de mínimos cuadrados (LS, por sus siglas en inglés Least Squares), el cual sigue el Principio de Mínimos Cuadrados, que estipula lo siguiente: “La solución  $\hat{\underline{\beta}}$ , debe dar la suma de cuadrados de las desviaciones verticales  $y_i$  observadas y los valores estimados  $\hat{y}_i$ , más pequeña posible”. Se define a los residuales  $e_i$  como las desviaciones verticales de las  $y_i$  con respecto a las  $\hat{y}_i$ , es decir:

$$e_i = y_i - \hat{y}_i \quad (1.10)$$

Para  $i = 1, 2, \dots, n$ . El vector de residuales  $\underline{e}$  por consiguiente es:

$$\underline{e} = \underline{y} - \hat{\underline{y}} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{pmatrix} \quad (1.11)$$

Con los valores estimados  $\hat{y}_i$  dados por:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip} \quad (1.12)$$

Tal que  $\underline{e}'\underline{e}$  sea mínimo, lo que equivale a maximizar el ajuste para  $\underline{y}$ . A cada  $\hat{y}_i$  en (1.12), se le llama el valor estimado o ajustado para  $y_i$ . Por tanto, el principio de

## Capítulo 1. Regresión Lineal Múltiple

mínimos cuadrados elige el vector columna de orden  $p + 1$ , tal que se minimice la suma de cuadrados de residuales o errores (SCE), esto es:

$$\begin{aligned}
 \text{SCE} &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 & (1.13) \\
 &= \sum_{i=1}^n \left( y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}) \right)^2 \\
 &= \sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_p x_{ip} \right)^2 \\
 &= \sum_{i=1}^n e_i^2 = \underline{e}' \underline{e}
 \end{aligned}$$

El estimador para los  $p + 1$  parámetros se obtiene de igualar a cero la derivada parcial de SCE con respecto a cada  $\hat{\beta}_i$ , de  $i=1, \dots, p$ . Dichas  $p$  derivadas parciales, dan origen a las llamadas ecuaciones normales, que forman un sistema de ecuaciones lineales que al resolverse dan como resultado  $\underline{\hat{\beta}}$ .

Por otro lado, con expresión matricial en (1.13) se tiene que:

$$\begin{aligned}
 \text{SCE} &= (\underline{y} - \underline{X}\hat{\underline{\beta}})' (\underline{y} - \underline{X}\hat{\underline{\beta}}) \\
 &= (\underline{y}' - \hat{\underline{\beta}}' \underline{X}') (\underline{y} - \underline{X}\hat{\underline{\beta}}) \\
 &= \underline{y}' \underline{y} - \hat{\underline{\beta}}' \underline{X}' \underline{y} - \underline{y}' \underline{X} \hat{\underline{\beta}} + \hat{\underline{\beta}}' \underline{X}' \underline{X} \hat{\underline{\beta}} & (1.14)
 \end{aligned}$$

La derivada parcial de SCE, con respecto al estimador del parámetro  $\underline{\hat{\beta}}$  es:

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial \underline{\hat{\beta}}} = -2\underline{X}' \underline{y} + 2\underline{X}' \underline{X} \underline{\hat{\beta}} \quad (1.15)$$

Igualando a cero se obtiene:

$$\underline{X}' \underline{y} = \underline{X}' \underline{X} \underline{\hat{\beta}} \quad (1.16)$$

Para encontrar la solución de (1.16), utilizando la inversa regular de  $\mathbf{X}'\mathbf{X}$ , este producto debe ser no singular es decir,  $\det(\mathbf{X}'\mathbf{X})^{-1} \neq 0$  (supuesto 2 de la RLM), si esto se cumple se tiene:

$$\hat{\underline{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underline{y} \quad (1.17)$$

El estimador  $\hat{\underline{\beta}}$  puede ser descrito mediante la descomposición en valores singulares<sup>2</sup> de  $\mathbf{X}$ , como sigue:

$$\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}'$$

Donde:

$\mathbf{U}$  y  $\mathbf{V}$  son matrices ortogonales, que contienen como columnas los vectores singulares izquierdos y los vectores singulares derechos, respectivamente.

$\mathbf{D}$  es una matriz diagonal que contiene los valores singulares de  $\mathbf{X}$  ordenados de manera decreciente, esto es,  $\mathbf{D} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_p})$  tal que  $\lambda_i > \lambda_{i+1} \forall i$ .

Por (1.17), se tiene:

$$\hat{\underline{\beta}} = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}'\underline{y} = \sum_{i=1}^p \frac{1}{\sqrt{\lambda_i}} \underline{v}_i \underline{u}_i' y_i \quad (1.18)$$

Así, la solución  $\hat{\underline{\beta}}$  es expresada en términos de los vectores y los valores propios de la matriz  $\mathbf{X}'\mathbf{X}$ .

Para el modelo de RLS, la recta ajustada o estimada resultante es:  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

Con

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{y} \quad \hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1$$

---

<sup>2</sup> Ver anexo A



## Capítulo 1. Regresión Lineal Múltiple

Las entradas  $\hat{\beta}_j$  representan la variación neta en  $y_j$  por cada unidad de variación en  $x_{ij}$ , para  $i=1,2,\dots,n$ ; es decir, toda la variación del espacio columna de  $\mathbf{X}$  que covaría con el vector  $\underline{y}$  es incluida en  $\hat{\underline{\beta}}$ . Si existe algún “ruido” en el espacio columna de  $\mathbf{X}$  covariando con  $\underline{y}$ , también será incluido, causando una alteración llamada “sobre ajuste” que produce graves errores de predicción.

### 1.3 Propiedades de $\hat{\underline{\beta}}$ , el estimador por mínimos cuadrados

De (1.17), es posible expresar a  $\hat{\underline{\beta}}$  como combinación lineal del vector  $\underline{y}$  como:

$$\hat{\underline{\beta}} = \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \underline{y} \quad (1.19)$$

Si el modelo  $\underline{y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$  es correcto para  $\underline{y}$  y  $\mathbf{X}$ , entonces se puede calcular el valor esperado de  $\hat{\underline{\beta}}$  como sigue:

$$\begin{aligned} E(\hat{\underline{\beta}}) &= E\left( \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \underline{y} \right) \\ &= \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) E(\underline{y}) \\ &= \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \mathbf{X}\underline{\beta} \\ &= \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \right) \underline{\beta} \\ &= \mathbf{I}_p \underline{\beta} \\ &= \underline{\beta} \end{aligned} \quad (1.20)$$

Se obtiene que  $\hat{\underline{\beta}}$  es un estimador insesgado de  $\underline{\beta}$ , es decir, en promedio se espera que las entradas del estimador  $\hat{\underline{\beta}}$  sean las mismas que  $\underline{\beta}$ . Se calcula la matriz de varianzas y covarianzas de  $\hat{\underline{\beta}}$ , como:

$$\begin{aligned} \Sigma_{\hat{\underline{\beta}}\hat{\underline{\beta}}} &= v\left( \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \underline{y} \right) \\ &= \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) v(\underline{y}) \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right)' \end{aligned}$$

### 1.3 Propiedades de $\underline{\hat{\beta}}$ , el estimador por mínimos cuadrados

$$\begin{aligned} &= \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \left( \mathbf{I}_n \sigma^2 \right) \left( \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \right)' \\ &= (\mathbf{X}'\mathbf{X})^{-1} \sigma^2 \end{aligned} \quad (1.21)$$

Los elementos sobre la diagonal de (1.21), corresponden a las varianzas de los coeficientes de regresión en el orden en que aparecen en  $\underline{\hat{\beta}}$ , los elementos fuera de ella son las covarianzas entre  $\hat{\beta}_j$  y  $\hat{\beta}_i$  para  $j=0,1,2,\dots, p$  e  $i=0,1,2,\dots, p$ , con  $i \neq j$ .

La distribución del estimador de coeficientes de regresión  $\underline{\hat{\beta}}$  es:

$$\underline{\hat{\beta}} \sim N \left( \underline{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \quad (1.22)$$

Propiedad necesaria para desarrollar la teoría de intervalos de confianza para cada  $\hat{\beta}_j$  con  $j=0,1,2,\dots, p$ . Esta teoría provee de más información al analista que solo la estimación puntual de los parámetros, sin embargo no es finalidad de este trabajo desarrollar esta teoría.

### 1.4 Los vectores $\underline{\hat{y}}$ y $\underline{e}$

Es posible escribir el valor ajustado de  $\underline{\hat{y}}$  como combinación lineal del vector de respuesta observado  $\underline{y}$ , como:

$$\begin{aligned} \underline{\hat{y}} &= \mathbf{X}\underline{\hat{\beta}} = \mathbf{X} \left( \left( (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \underline{y} \right) \\ &= \left( \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \underline{y} \\ &= \mathbf{H} \underline{y} \end{aligned} \quad (1.23)$$

Con  $\mathbf{H} = \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ . A la matriz  $\mathbf{H}$  se le conoce como matriz sombrero (Hat), que también es una matriz de proyección, la cual juega un papel muy importante en el análisis de regresión.

Propiedades<sup>3</sup> de  $\mathbf{H}$ :

---

<sup>3</sup> Seber, G.A.F. (1977)

1.  $\mathbf{H}$  y  $(\mathbf{I}_n - \mathbf{H})$  son matrices simétricas e idempotentes
2.  $\text{rango}(\mathbf{I}_n - \mathbf{H}) = \text{tr}(\mathbf{I}_n - \mathbf{H}) = n - (p + 1)$
3.  $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = 0$

Utilizando la expresión (1.23), se obtiene  $E(\underline{\hat{y}})$  y  $V(\underline{\hat{y}})$  como sigue:

$$\begin{aligned}
 E(\underline{\hat{y}}) &= E(\mathbf{H}\underline{y}) \\
 &= \mathbf{H}E(\underline{y}) \\
 &= \mathbf{H}\underline{\mathbf{X}}\underline{\beta} \\
 &= \underline{\mathbf{X}}\underline{\beta}
 \end{aligned} \tag{1.24}$$

Ya que  $\mathbf{H}\mathbf{X} = \left( \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right)\mathbf{X} = \mathbf{X}\mathbf{I}_p = \mathbf{X}$ . Es decir,  $\underline{\hat{y}}$  es un estimador insesgado de la media de  $\underline{y}$  dados los valores de  $\mathbf{X}$ . Calculando la matriz de varianzas y covarianzas de  $\underline{\hat{y}}$ , se obtiene:

$$\begin{aligned}
 \Sigma_{\underline{\hat{y}}\underline{\hat{y}}} &= V(\mathbf{H}\underline{y}) \\
 &= \mathbf{H}(V(\underline{y}))\mathbf{H}' \\
 &= \mathbf{H}(\sigma^2\mathbf{I}_n)\mathbf{H}' \text{ Por propiedad (1)} \\
 &= \mathbf{H}\sigma^2
 \end{aligned} \tag{1.25}$$

Por otro lado, el vector de residuales  $\underline{e}$  refleja la inconsistencia entre los valores observados y los estimados, pues fue definido como:

$$\begin{aligned}
 \underline{e} &= \underline{y} - \underline{\hat{y}} \\
 &= \underline{y} - \underline{\mathbf{X}}\underline{\hat{\beta}} \\
 &= \underline{y} - \mathbf{H}\underline{y} \\
 &= (\mathbf{I}_n - \mathbf{H})\underline{y}
 \end{aligned} \tag{1.26}$$

La esperanza del vector de residuales, se obtiene como sigue:

## 1.4 Propiedades de $\hat{\underline{\beta}}$ , el estimador por mínimos cuadrados

$$E(\underline{e}) = E((\mathbf{I}_n - \mathbf{H})\underline{y}) = (\mathbf{I}_n - \mathbf{H})E(\underline{y}) = (\mathbf{I}_n - \mathbf{H})\mathbf{X}\underline{\beta} = \mathbf{X}\underline{\beta} - \mathbf{H}\mathbf{X}\underline{\beta} = (\mathbf{X} - \mathbf{H}\mathbf{X})\underline{\beta} = \mathbf{0} \quad (1.27)$$

De (1.11) se tiene que  $\underline{y} = \underline{e} - \hat{\underline{y}}$ , así la variable de respuesta  $\underline{y}$  está dividida en dos partes, la que aporta el modelo de RLM y la de los residuales  $\underline{e}$ .

La matriz de varianzas y covarianzas del vector de residuales  $\underline{e}$ , es calculada como:

$$\begin{aligned} \Sigma_{\underline{ee}} &= V((\mathbf{I}_n - \mathbf{H})\underline{y}) \\ &= (\mathbf{I}_n - \mathbf{H}) V(\underline{y}) (\mathbf{I}_n - \mathbf{H})' \\ &= (\mathbf{I}_n - \mathbf{H}) \sigma^2 \mathbf{I}_n (\mathbf{I}_n - \mathbf{H})' \\ &= \sigma^2 (\mathbf{I}_n - \mathbf{H}) (\mathbf{I}_n - \mathbf{H})' \\ &= \sigma^2 (\mathbf{I}_n - \mathbf{H}) \end{aligned} \quad (1.28)$$

## 1.5 Estimación del parámetro $\sigma^2$

Otro de los parámetros implicados en el modelo de RLM es  $\sigma^2$ , éste debe ser estimado a partir de la suma de cuadrados de residuales definida en (1.14). Esta tiene  $(n - p + 1)$  grados de libertad asociados, puesto que existen  $p + 1$  parámetros en el modelo. Se define a CME como el cuadrado medio del error, dado por:

$$CME = \frac{SCE}{n - (p + 1)} \quad (1.29)$$

Se probará que CME es un estimador insesgado para  $\sigma^2$ .

Demostración: Utilizando la expresión para  $\underline{e}$  dada en (1.26), se tiene que:

$$\begin{aligned} (n - (p + 1)) CME &= \underline{e}'\underline{e} \\ &= \underline{y}'(\mathbf{I}_n - \mathbf{H})'(\mathbf{I}_n - \mathbf{H})\underline{y} \\ &= \underline{y}'(\mathbf{I}_n - \mathbf{H})\underline{y} \text{ por propiedad (1)} \end{aligned}$$

Para calcular el valor esperado del CME, se requiere hacer uso del siguiente resultado para el valor esperado de una forma cuadrática:

Teorema 1<sup>4</sup>: Sea  $\underline{\delta}$  un vector aleatorio de orden  $n$ ; sea  $\mathbf{A}_{n \times n}$  una matriz simétrica. Si  $E(\underline{\delta}) = \underline{\theta}$  y  $V(\underline{\delta}) = \Sigma$ , entonces:  $E(\underline{\delta}' \mathbf{A} \underline{\delta}) = \text{tr}(\mathbf{A} \Sigma) + \underline{\delta}' \mathbf{A} \underline{\delta}$

Usando el teorema 1, (1.27) y las propiedades 2 y 3 de la matriz sombrero, se tiene que:

$$\begin{aligned} E((n - (p + 1)) \text{CME}) &= E(\underline{y}' (\mathbf{I}_n - \mathbf{H}) \underline{y}) \\ &= \text{tr}((\mathbf{I}_n - \mathbf{H}) \sigma^2 \mathbf{I}_n) \\ &= \sigma^2 (n - (p + 1)) \\ \Rightarrow E(\text{CME}) &= \frac{\sigma^2 (n - (p + 1))}{n - (p + 1)} = \sigma^2 \end{aligned} \quad (1.30)$$

Entonces el CME es un estimador insesgado para  $\sigma^2$ .

En resumen, de (1.20) y (1.30) los estimadores para  $\underline{\beta}$  y  $\sigma^2$ , respectivamente, son insesgados.

## 1.6 Modelo lineal centrado

En algunas ocasiones es conveniente centrar las variables expuestas en un modelo de regresión al sustraer su media a cada variable en  $\mathbf{X} = (\underline{x}_1 \quad \underline{x}_2 \quad \cdots \quad \underline{x}_p)$ , y de igual forma para  $\underline{y}$ . El vector de medias muestrales de las variables independientes y la media de la variable dependiente, están dadas por:

$$\underline{\mu}_x = \frac{1}{n} \mathbf{X}' \underline{1} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} \text{ con } \bar{x}_i = \frac{\sum_{j=1}^n x_{ji}}{n} \text{ y } \underline{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix}; \text{ para } \underline{y}, \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

---

<sup>4</sup> Seber, G.A.F. (1977)

## 1.6 Modelo lineal centrado

Entonces, las variables independientes centradas son:  $\mathbf{X}_c = \mathbf{X} - \underline{\mathbf{1}}\underline{\mu}_x$  y  $\underline{y}_c = \underline{y} - \bar{y}\underline{\mathbf{1}}$  con  $\underline{\mathbf{1}}$  el vector de unos de orden  $n$ .

El modelo de regresión en (1.4) es equivalente a  $\underline{y}_c = \beta_0 + \mathbf{X}_c\beta + \varepsilon$  si el término intercepción  $\beta_0$  es:  $\beta_0 = \bar{y} + \beta_1\bar{x}_1 + \beta_2\bar{x}_2 + \dots + \beta_p\bar{x}_p$ . Note que  $\underline{\beta}$  es un vector de orden  $p$ , i.e.,  $\underline{\beta} = (\beta_1 \ \beta_2 \ \dots \ \beta_p)'$ , pues en este caso, el vector columna de unos en  $\mathbf{X}$  del modelo (1.4), es omitido. En analogía a (1.18), el estimador de coeficientes de regresión por mínimos cuadrados corresponde a  $\underline{\hat{\beta}} = (\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\underline{y}_c$ ; donde el término de intercepción es estimado como:

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \dots + \hat{\beta}_p\bar{x}_p \\ &= \bar{y} + \underline{\hat{\beta}}'\underline{\mu}_x\end{aligned}$$

Los valores ajustados para la variable de respuesta corresponden a:

$$\underline{\hat{y}}_c = \hat{\beta}_0 + \mathbf{X}_c\underline{\hat{\beta}}$$

# CAPÍTULO 2

## EL PROBLEMA DE MULTICOLINEALIDAD

### 2.1 Multicolinealidad

En este capítulo, se analiza uno de los problemas ligados a la obtención masiva de variables explicativas: regresoras correlacionadas entre sí, ésto provoca que el modelo de RLM no sea útil, pues se dice que unas variables regresoras miden el mismo concepto que otras, pero de forma alternativa. A este fenómeno se le conoce como multicolinealidad, la cual es dividida en dos tipos: multicolinealidad “perfecta” y multicolinealidad “casi perfecta”. La primera, refleja una dependencia lineal exacta entre dos o más regresoras, es decir, la  $j$ -ésima regresora es combinación lineal de una, varias o todas las regresoras restantes; el segundo tipo, obedece a una combinación lineal cercana a cero entre dos o más regresoras; se puede detectar el problema de multicolinealidad en  $\mathbf{X}$  mediante su rango.

Definición: El rango de una matriz  $\mathbf{X}_{n \times p}$  es el número de renglones (columnas) linealmente independientes, donde:

$$0 \leq \text{rango}(\mathbf{X}) \leq \min(n, p)$$

Además

$$\text{rango}(\mathbf{X}) = \text{rango}(\mathbf{X}'\mathbf{X})$$

$$\therefore 0 \leq \text{rango}(\mathbf{X}'\mathbf{X}) = r \leq p$$

De acuerdo a lo anterior, si  $\mathbf{X}$  tiene dependencias lineales exactas, entonces  $\mathbf{X}'\mathbf{X}$  es singular, y la expresión (1.16) no puede resolverse para  $\hat{\underline{\beta}}$ . De igual forma sucede bajo multicolinealidad casi perfecta en  $\mathbf{X}$ , pues  $\mathbf{X}'\mathbf{X}$  hereda estas dependencias lineales y las entradas del vector  $\hat{\underline{\beta}}$  por mínimos cuadrados, serán grandes en valor absoluto. Por otro lado, considere el cuadrado de la distancia de  $\hat{\underline{\beta}}$  al valor del parámetro

## 2.1 Multicolinealidad

verdadero  $\underline{\beta}$ , dada por:  $\rho = (\underline{\hat{\beta}} - \underline{\beta})'(\underline{\hat{\beta}} - \underline{\beta})$ ; el valor esperado de  $\rho$ , comúnmente denominado error cuadrático medio (ECM), es:

$$\text{ECM} = E(\rho) = E\left((\underline{\hat{\beta}} - \underline{\beta})'(\underline{\hat{\beta}} - \underline{\beta})\right) = \sum_{i=1}^p E(\hat{\beta}_i - \beta_i)^2 = \sum_{i=1}^p V(\hat{\beta}_i) = \text{Tr}\left((\mathbf{X}'\mathbf{X})^{-1}\right) \sigma^2$$

Ya que  $E(\underline{\hat{\beta}}) = \underline{\beta}$  y  $\Sigma_{\underline{\hat{\beta}}} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ . Ahora, por el teorema de descomposición espectral<sup>1</sup> se puede expresar a  $\mathbf{X}'\mathbf{X}$  como:

$$\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{\Delta}\mathbf{V}'$$

$$\begin{aligned} \text{Así:} \quad \Rightarrow \quad \text{Tr}\left((\mathbf{X}'\mathbf{X})^{-1}\right) &= \text{Tr}\left(\mathbf{V}\mathbf{\Delta}^{-1}\mathbf{V}'\right) \text{ por propiedad de la inversa} \\ &= \text{Tr}\left(\mathbf{\Delta}^{-1}\mathbf{V}'\mathbf{V}\right); \text{ por propiedad de traza} \\ &= \text{Tr}\left(\mathbf{\Delta}^{-1}\right); \text{ por la ortogonalidad de } \mathbf{V} \\ &= \sum_{i=1}^p \frac{1}{\lambda_i} \end{aligned}$$

$$\therefore \text{ECM} = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} \quad (2.1)$$

Cuando existe dependencia lineal cercana entre regresoras, los valores propios del polinomio característico  $|\mathbf{X}'\mathbf{X} - \lambda\mathbf{I}_p|$  serán cercanos a cero, implicando que el ECM sea grande, convirtiendo a  $\underline{\hat{\beta}}$  por mínimos cuadrados en una estimación errónea de  $\underline{\beta}$ . Aunado a esto,  $\Sigma_{\underline{\hat{\beta}}} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$  tendrá entradas grandes, entonces  $\underline{\hat{\beta}}$  será insesgado pero  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$  no serán de varianza mínima; aplicar la RLM con algún tipo de multicolinealidad en  $\mathbf{X}$  hace que  $\underline{\hat{\beta}}$  produzca un “sobre ajuste” del modelo a los datos, llevando a resultados equívocos. Cuando se presenta el problema de multicolinealidad, se requiere de otras técnicas multivariadas tales como el análisis de componentes principales (ACP), regresión sesgada (RS) o mínimos cuadrados parciales (MCP), para obtener dicho estimador.

El problema de multicolinealidad es inherente a 3 causas principales:

<sup>1</sup> Ver Anexo A



## Capítulo 2. El problema de Multicolinealidad

1. Métodos de recolección de información
2. Especificación errónea del modelo de regresión
3. Mayor número de variables regresoras que de observaciones

El último punto, considera a  $X$  como una matriz baja y gruesa, que se puede representar con el siguiente modelo lineal:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}_{n \times (p+1)} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{(p+1) \times 1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1}$$

Este tipo de modelos son muy comunes en investigaciones médicas, químicas, psicológicas, etc., donde hay un pequeño número de unidades muestrales. En el sistema  $\underline{y} = X\underline{\beta}$ , el número de ecuaciones es menor que el número de incógnitas, teniendo un número infinito de soluciones, pues  $\text{rango}(X) = r \leq p+1$ , entonces existen  $(p+1) - r$  columnas linealmente dependientes, implicando multicolinealidad perfecta. Es en este caso en particular, para el cual la técnica de MCP fue desarrollada.

### 2.1.1 Detección de la Multicolinealidad

Existen varias estadísticas para detectar multicolinealidad en  $X$ , aquí se presentan algunas de las más usadas; su aplicación es básica antes de plantear el modelo de regresión para evitar errores de ajuste derivados de este problema.

- **Coeficiente de determinación  $R^2$ , factor de inflación de varianza (FIV) y factor de tolerancia (T)**

Bajo la hipótesis de multicolinealidad en  $X$ , existe una relación lineal entre las regresoras. Se plantea una ecuación auxiliar para cada vector columna de  $X$  de la siguiente manera:

$$\underline{x}_i = \beta_0 + \beta_1 \underline{x}_1 + \dots + \beta_{i-1} \underline{x}_{i-1} + \beta_{i+1} \underline{x}_{i+1} + \dots + \beta_p \underline{x}_p \quad (2.2)$$

## 2.1 Multicolinealidad

Con  $i = 1, 2, \dots, n$ ; para medir el ajuste del modelo (2.2) para la  $i$ -ésima regresora, se calcula el coeficiente de determinación  $R_i^2$ :

$$R_i^2 = \frac{\sum_{i=1}^n (\hat{x}_i - \bar{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2} = \frac{SCR}{SCT} \quad (2.3)$$

Con  $R_i^2 \leq 1$ , y donde:

SCR := Suma de cuadrados debida a la regresión

SCT := Suma de cuadrados totales.

El coeficiente de determinación  $R_i^2$ , representa la proporción de variabilidad explicada por el modelo con respecto a la variabilidad total del sistema. Dicho de esta manera, si  $R_i^2$  es cercano a uno, implica buen ajuste en el modelo (2.2), corroborando correlación perfecta entre la  $i$ -ésima regresora y las restantes. Cuando  $R_i^2$  es cercano a cero, refleja un mal ajuste del modelo y podría concluirse este no tiene sentido, por falta de dependencia lineal entre  $x_i$  con las regresoras restantes.

El procedimiento yace en la obtención del coeficiente de determinación para todos los modelos de regresión de  $x_i$  sobre  $\{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p\}$ , e ir identificando independencia o dependencia entre ellas.

El efecto de la multicolinealidad es “inflar” la varianza de los estimadores de los coeficientes de regresión, una medida usual para diagnosticar multicolinealidad está dada por el factor de inflación de varianza (FIV), definido como:

$$FIV_i = (1 - R_i^2)^{-1} \quad \text{con } i = 1, 2, \dots, p \quad (2.4)$$

El FIV es el factor por el cual la varianza de  $\hat{\beta}$  es incrementada, debido a la dependencia lineal entre regresoras; tomará el valor de 1 en caso de independencia lineal entre  $x_i$  con las variables regresoras restantes y no posee límite superior, en

## Capítulo 2. El problema de Multicolinealidad

caso contrario. La experiencia dicta que si el VIF excede el valor de 5, hay una clara indicación de que  $\underline{\beta}$  será pobremente estimado, debido a un severo problema de multicolinealidad en  $\mathbf{X}$ .

Otro factor para diagnosticar dependencias lineales de regresoras está dado por  $T = 1 - R_1^2$ , conocido como Factor de Tolerancia; tiene valor máximo de 1 cuando la variable  $\underline{x}_i$  no tiene asociación lineal con las restantes y valor mínimo de cero cuando la regresora  $\underline{x}_i$  es combinación lineal perfecta del resto.

Varios autores manejan al FIV y al factor T, como medidas no concluyentes para realizar un diagnóstico; sin embargo su uso es aplaudido por representar una manera de visualizar un sobre ajuste del modelo de regresión planteado sobre  $\underline{y}$ .

- **Análisis de la matriz de correlaciones  $\mathbf{R}_{\mathbf{X}\mathbf{X}}$ .**

El análisis de las entradas de la matriz  $\mathbf{R}_{\mathbf{X}\mathbf{X}}$  permite detectar correlaciones bivariadas fuertes entre las variables regresoras, la desventaja es que no se detectará la dependencia lineal entre más de dos regresoras.

- **Valores propios de la matriz  $\mathbf{X}'\mathbf{X}$**

Si existen dependencias lineales cercanas entre las variables regresoras, uno o más valores propios de  $\mathbf{X}'\mathbf{X}$  serán cercanos a cero; aquellos iguales a cero implican multicolinealidad perfecta, i.e.  $\text{rango}(\mathbf{X}'\mathbf{X}) < p$ . Algunos analistas usualmente examinan el número condición, definido como:

$$\text{Número Condición} := \kappa = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (2.5)$$

Los siguientes criterios son utilizados para  $\kappa$  :

- $\kappa < 100$  No hay un problema serio de multicolinealidad
- $100 \leq \kappa \leq 1000$  Existe multicolinealidad casi perfecta
- $\kappa > 1000$  Existe multicolinealidad perfecta

## 2.1 Multicolinealidad

O bien, es común utilizar el factor  $\sqrt{\kappa}$ ; sus rangos son la raíz cuadrada de los rangos de  $\kappa$ . Para estimar el número de dependencias lineales en  $\mathbf{X}'\mathbf{X}$ , se puede tomar el índice condición, definido como:

$$\kappa_j = \frac{\lambda_{\max}}{\lambda_j} \quad \text{con } j=1,2,\dots,p \quad (2.6)$$

El número de los  $\kappa_j$  mayores a 1000, es el valor estimado de dependencias lineales cercanas en  $\mathbf{X}$ .

## 2.2 Pre-procesamiento de los datos

En el análisis de datos, es necesario calcular diversas estadísticas sobre las variables dada una muestra; por ejemplo, la media y varianza muestral, así como covarianza y correlación muestral; sin embargo algunas de estas estadísticas retienen las unidades de medida como lo son la varianza y la covarianza, característica que complica la comparación de estas estadísticas para variables que estén en diferentes unidades de medida, de hecho, en algunos casos tal comparación no tiene sentido; así lo factible es eliminar las unidades de medida al estandarizar las variables originales, lo que significa transformarlas a otras, con media cero y varianza uno, esto se realiza al centrar y escalar cada variable restando su media muestral y dividiéndola entre su desviación estándar. En notación matricial, se tiene:

$$\text{Si } \mathbf{X}_{n \times p} \Rightarrow \mathbf{X}^* = \left( \mathbf{X} - \left( \mathbf{1} \underline{\mu}' \right) \right) \Psi$$

Donde:

$\underline{\mu}_x :=$  Vector de medias muestrales dado por  $\underline{\mu}_x = \frac{1}{n} \mathbf{X}' \mathbf{1}$ , i.e:

$$\underline{\mu}_x' = \left( \bar{x}_1 \quad \bar{x}_2 \quad \dots \quad \bar{x}_p \right)$$

$\Psi :=$  Matriz diagonal que contiene las desviaciones estándar muestrales de las

$p$  variables de  $\mathbf{X}$ , i.e.,  $\Psi = \text{diag} \left( \sigma_{x_1}^{-1}, \sigma_{x_2}^{-1}, \dots, \sigma_{x_p}^{-1} \right)$  con:

$$\sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$$

## Capítulo 2. El problema de Multicolinealidad

La estandarización otorga igual importancia a cada una de las variables y, puede realizarse como una etapa previa al análisis, con la finalidad de obtener mejores resultados. En el análisis de regresión, si se tiene conocimiento de que algunas variables regresoras tienen mayor “ruido” que otras, puede ser más interesante no estandarizar las variables, para no dar igual importancia a éstas que a las variables regresoras altamente correlacionadas con la variable de respuesta.

Nota: Si las variables están medidas en las mismas unidades, la estandarización es redundante.

# CAPÍTULO 3

## MÉTODOS DE REGRESIÓN POR VARIABLES LATENTES (RVL)

### 3.1 Introducción

Los métodos de regresión por variables latentes (RVL) son conceptualmente muy diferentes al modelo de RLM, ya que trabajan con datos que presentan o un gran número de variables regresoras o multicolinealidad en estas, de manera sencilla, características que los convierten en métodos sumamente útiles.

El fundamento principal de los métodos RVL es suponer que el sistema de variables en consideración, puede ser descrito por un conjunto pequeño de otras, conocidas como “variables artificiales” o “variables latentes”, que resumen óptimamente la información del sistema, antes de resolver el problema de predicción. Las variables latentes por definición son combinaciones lineales de las variables originales, donde los vectores coeficientes o “vectores de pesos”  $\underline{w}_i$  son no triviales y elegidos de tal manera que se cumple algún criterio de optimización; se pretende ajustar a las  $p$  variables regresoras mediante un pequeño subconjunto de ellas, de dimensión  $a < p$ , a costa de perder una pequeña (o insignificante) parte de información. Aunque el supuesto es el mismo para todo método RVL, las variables latentes son construidas siguiendo diferentes criterios de maximización. Se introducirá el concepto de descomposición bilineal de una matriz, que se aplicará a la matriz de datos  $\mathbf{X}$ , con el objetivo de llevar la información a un nuevo sistema de coordenadas definido por las variables latentes; matemáticamente, se trata de descomponer a  $\mathbf{X}$  mediante el producto de dos matrices,  $\mathbf{T}$  (matriz de variables latentes) y  $\mathbf{P}$  (matriz de cargas) más una matriz de residuales  $\mathbf{E}$ .

Para ampliar esta idea, considere los siguientes modelos:

### Capítulo 3. Métodos de Regresión por Variables latentes (RVL)

Sea  $\mathbf{X}_{n \times p}$  y  $\underline{y}_{n \times 1}$ , entonces:

$$\mathbf{X} = \mathbf{T}_a \mathbf{P}_a' + \mathbf{E}_a \quad (3.1)$$

$$\underline{y} = \mathbf{T}_a \underline{q}_a' + \underline{f}_a \quad (3.2)$$

Donde:

$\mathbf{T}_a$  := Matriz de variables latentes de  $\mathbf{X}$ , de orden  $n \times a$ .

$\mathbf{P}_a$  := Matriz de cargas de  $\mathbf{X}$ , de orden  $p \times a$ ; contiene los coeficientes de la regresión de  $\mathbf{X}$  sobre  $\mathbf{T}_a$ , es decir, las cargas son las proyecciones de los datos originales en el nuevo sistema de coordenadas definido por los vectores columna de  $\mathbf{T}_a$ .

$\mathbf{E}_a$  := Matriz de residuales de  $\mathbf{X}$ , de orden  $p \times a$ .

$\mathbf{W}_a$  := Matriz que contiene como columnas a los vectores de pesos no triviales  $\{\underline{w}_i\}_{i=1}^a$  de orden  $p \times a$ . Los vectores de pesos son los ponderadores de las columnas de  $\mathbf{X}$ . Por lo general se normaliza a los  $\underline{w}_i$  para  $i = 1, 2, \dots, a$  ( $\|\underline{w}_i\| = 1$ ).

$\underline{q}_a$  := Vector de cargas de  $\underline{y}$  de orden  $a$ ; contiene los coeficientes de la regresión de  $\underline{y}$  sobre  $\mathbf{T}$ .

$\underline{f}_a$  := Vector de residuales de  $\underline{y}$ , de orden  $n$ .

En la descomposición bilineal de  $\mathbf{X}$ , interactúan las variables latentes y las cargas, pues con cada pareja de vectores  $\underline{t}_i$  y  $\underline{p}_i'$ , desde  $i = 1, 2, \dots, n$ , se pretende explicar al conjunto de regresoras originales mediante aproximaciones de rango uno, dadas por:

$$\widehat{\mathbf{X}}^{(1)} = \underline{t}_1 \underline{p}_1', \widehat{\mathbf{X}}^{(2)} = \underline{t}_2 \underline{p}_2', \dots, \widehat{\mathbf{X}}^{(a)} = \underline{t}_a \underline{p}_a' \quad (3.3)$$

La información no explicada por las primeras  $a$  variables latentes, está contemplada en la matriz de residuales  $\mathbf{E}_{a+1}$ .

La verosimilitud del modelo (3.2) se justifica debido a que en regresión, la variable  $\underline{y}$  se explica por  $\mathbf{X}$  entonces el conjunto  $\{t_1, t_2, \dots, t_a\}$  se supone buen regresor para  $\underline{y}$ , así:

$$\hat{\underline{y}}^{(1)} = t_1 q_1, \hat{\underline{y}}^{(2)} = t_2 q_2, \dots, \hat{\underline{y}}^{(a)} = t_a q_a \quad (3.4)$$

Si el supuesto de dependencia lineal entre  $\mathbf{X}$  y  $\underline{y}$  es correcto, se espera que las aproximaciones de rango uno para  $\underline{y}$  dejen solo variación aleatoria en el vector de residuales  $\underline{f}_a$ . En el modelo (3.1), la descomposición bilineal de  $\mathbf{X}$  es exacta solo si el número de variables latentes es igual al número de vectores columna linealmente independientes de  $\mathbf{X}$ , puesto que  $\{t_1, t_2, \dots, t_a\}$  genera un subespacio del espacio columna de  $\mathbf{X}$ . El número de variables latentes debe ser restringido, para lograr el objetivo de describir a las variables originales sobre un subespacio de baja dimensionalidad, tal que el modelo de regresión sea óptimo ( $a < r$ ), esto es, la información relevante debe ser resumida por las primeras variables latentes, mientras que el “ruido” debe ser modelado por las últimas. La elección del número  $a$  se hace aplicando técnicas de validación de ajuste del modelo a los datos, o bien, reteniendo aquellas variables latentes que expliquen gran cantidad de la varianza global del sistema original. Cabe señalar que, para cualquier método RVL la calidad de predicción aumenta conforme el número de variables latentes lo hace por dejar poca o nula variabilidad sistemática en las matrices de residuales, sin embargo baja la calidad de interpretación<sup>1</sup>.

La figura 3.1, representa la forma matricial de los modelos (3.1) y (3.2). La matriz  $\mathbf{E}_a$  y el vector  $\underline{f}_a$  representan en el caso ideal residuales puramente aleatorios o “ruido blanco”, de lo contrario (sí contienen variación sistemática) será necesario agregar otra(s) componente(s), para poder explicar mejor la información y conseguir mayor calidad predictiva. Si la variable de respuesta es una matriz ( $\mathbf{Y}$ ), dependiendo del modelo RVL, se toman algunas consideraciones extras para la construcción de las variables latentes.

---

<sup>1</sup> Helland, I. (1988)



## Capítulo 3. Métodos de Regresión por Variables latentes (RVL)

$$\begin{array}{c}
 \mathbf{X} \quad \underline{t}_1 \quad \underline{p}'_1 \quad \underline{t}_2 \quad \underline{p}'_2 \quad \dots \quad \underline{t}_a \quad \underline{p}'_a \quad \mathbf{E}_a \\
 \left[ \begin{array}{c} \\ \\ \\ \end{array} \right] = \left( \begin{array}{c} \\ \\ \\ \end{array} \right) + \left( \begin{array}{c} \\ \\ \\ \end{array} \right) + \dots + \left( \begin{array}{c} \\ \\ \\ \end{array} \right) + \left[ \begin{array}{c} \\ \\ \\ \end{array} \right] \\
 \\
 \underline{y} \quad \underline{t}_1 \quad \underline{q}_1 \quad \underline{t}_2 \quad \underline{q}_2 \quad \dots \quad \underline{t}_a \quad \underline{q}_a \quad \underline{f}_a \\
 \left( \begin{array}{c} \\ \\ \\ \end{array} \right) = \left( \begin{array}{c} \\ \\ \\ \end{array} \right) + \left( \begin{array}{c} \\ \\ \\ \end{array} \right) + \dots + \left( \begin{array}{c} \\ \\ \\ \end{array} \right) + \left( \begin{array}{c} \\ \\ \\ \end{array} \right)
 \end{array}$$

Figura 3.1<sup>2</sup>: Descomposición bilineal de  $\mathbf{X}$  y  $\underline{y}$ , por aproximaciones de rango uno.

### 3.2 Análisis de Componentes Principales (ACP)

#### 3.2.1 Introducción

La técnica de análisis de componentes principales es debida a Hotelling (1933), aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados, introducidos por K. Pearson (1901). El ACP es una técnica sugerida para datos multicolineales, pues estudia las relaciones que se presentan dentro del conjunto de variables regresoras y a su vez, las explica por medio del conjunto de variables latentes, mejor conocidas como componentes principales.

Las componentes principales son construidas mutuamente ortogonales y según el orden de importancia que adquieren, en cuanto al porcentaje de variabilidad total que recogen de la muestra; éstas son combinaciones lineales de las variables regresoras originales tal que los vectores de pesos son elegidos al maximizar la cantidad de varianza que describe a cada componente, esto es, se busca que la mayor variabilidad que define a  $\mathbf{X}$ , se concentre en la primera componente principal; la segunda será elegida tal que concentre la mayor variabilidad de  $\mathbf{X}$  después de la primera y así sucesivamente; la idea es reducir la dimensión del espacio generado por las variables regresoras  $\{x_1, x_2, \dots, x_p\}$  del espacio  $p$  a otro de menor dimensión, ya que ACP efectúa un cambio del sistema de coordenadas  $p$ -dimensional a otro  $a$ -dimensional

<sup>2</sup> Fuente: Risvik, H. (2007)

## 3.2 Análisis de Componentes Principales (ACP)

estructurado por las primeras  $a$  componentes principales, así el efecto de reducción de dimensionalidad recae en las componentes principales que queden fuera del análisis por representar un porcentaje bajo o insignificante, de la variabilidad total del sistema; el ACP es una técnica matemática que no requiere el supuesto de alguna función de distribución de los datos para efectuar la reducción de dimensionalidad.

Si la matriz  $X$  es de rango completo por columnas, efectuar un análisis por componentes principales no tiene sentido ya que, se obtendría el conjunto de variables regresoras originales, pero ordenadas de mayor a menor varianza.

Las metas específicas del ACP son:

1. Extraer la información más importante de la matriz de datos  $X$ .
2. Simplificar la descripción de la variable de respuesta manteniendo esa información, para construir el modelo de regresión.

Es útil cuando:

1. El número de variables regresoras excede el número de observaciones (alta dimensionalidad de regresoras).
2. Existe algún grado de multicolinealidad en la matriz de datos  $X$ .

### 3.2.2 Invarianza de escala

El ACP no es invariante bajo transformaciones lineales, ya que aplica el criterio de maximización de varianza y, guiará a las componentes principales hacia las variables regresoras con mayor varianza; si existe una notoria diferencia entre las varianzas de las regresoras originada por las unidades de medida de las variables, el ACP estará dando resultados erróneos. La mejor elección es iniciar el ACP con las variables regresoras estandarizadas, para asegurar que todas las variables tengan igual importancia en el análisis.

El ACP omitiendo la estandarización de las variables originales, sería llevado a cabo solo si las variables independientes están bajo las mismas unidades de medida o bien, en unidades de medida comparables.

### 3.2.3 Fundamento matemático

Sea  $z_1$  la primera componente principal dada por  $z_1 = \underline{a}_1' \underline{x}$  con:

$\underline{a}_1 = (a_{11} \ a_{21} \ \dots \ a_{p1})$  y  $\underline{x}' = (x_1 \ x_2 \ \dots \ x_p)$  un vector aleatorio, entonces:

$$\begin{aligned} V(z_1) &= E \left[ \left( \underline{a}_1' \underline{x} - E(\underline{a}_1' \underline{x}) \right) \left( \underline{a}_1' \underline{x} - E(\underline{a}_1' \underline{x}) \right)' \right] \\ &= \underline{a}_1' E \left[ (\underline{x} - E(\underline{x})) (\underline{x} - E(\underline{x}))' \right] \underline{a}_1 \\ &= \underline{a}_1' \Sigma_{xx} \underline{a}_1 \end{aligned} \quad (3.5)$$

Donde:  $\Sigma_{xx}$  es la matriz de varianzas y covarianzas de  $\underline{x}$ . El vector de pesos  $\underline{a}_1$  debe

ser elegido tal que  $z_1$  posea la varianza máxima posible y  $\sum_{i=1}^p a_{i1}^2 = 1$ , es decir:

$$\text{Max}(V(z_1)) = \text{Max} \left( V(\underline{a}_1' \underline{x}) \right)$$

$$\text{Sujeto a: } \underline{a}_1' \underline{a}_1 = 1$$

Sea  $\phi$  una función que incluye la varianza de la primera componente principal y una penalidad que contiene la restricción  $\underline{a}_1' \underline{a}_1 = 1$  y al multiplicador de Lagrange  $\lambda$ :

$$\phi = \underline{a}_1' \Sigma_{xx} \underline{a}_1 - \lambda (\underline{a}_1' \underline{a}_1 - 1) \quad (3.6)$$

La maximización de  $\phi$  determina el vector de pesos  $\underline{a}_1$  que maximiza  $V(z_1)$ .

Derivando con respecto a  $\underline{a}_1$  se obtiene:

$$\frac{\partial \phi}{\partial \underline{a}_1} = 2 \Sigma_{xx} \underline{a}_1 - 2 \lambda \underline{a}_1$$

Igualando a cero la derivada simbólica anterior se tiene:

### 3.2 Análisis de Componentes Principales (ACP)

$$2\Sigma_{xx} \underline{a}_1 = 2\lambda \underline{a}_1 \Leftrightarrow \Sigma_{xx} \underline{a}_1 = \lambda \underline{a}_1 \quad (3.7)$$

Por definición de vector propio<sup>3</sup>,  $\underline{a}_1$  es el vector propio de  $\Sigma_{xx}$  asociado al valor propio más grande, por ejemplo  $\lambda_1$ , para lograr la maximización de  $\varphi$ . Multiplicando (3.7) por  $\underline{a}_1'$  por la izquierda y de (3.5), resulta:

$$\underline{a}_1' \Sigma_{xx} \underline{a}_1 = \lambda_1 \underline{a}_1' \underline{a}_1 = \lambda_1 = V(z_1) \quad (3.8)$$

Entonces, la varianza de la primera componente principal corresponde al valor propio  $\lambda_1$ . La segunda componente principal  $z_2 = \underline{a}_2' \underline{x}$ , es construida tal que concentre la máxima variabilidad de  $\underline{x}$  después de la primera, sujeta a dos condiciones:  $z_2$  debe ser no correlacionada con  $z_1$  y  $\underline{a}_2$  debe de ser de norma unitaria, esto es:

$$\begin{aligned} \text{Max} \left( V(z_2) \right) &= \text{Max} \left( V \left( \underline{a}_2' \underline{x} \right) \right) \\ \text{Sujeto a: } &\begin{cases} \text{Cov}(z_2, z_1) = 0 \\ \underline{a}_2' \underline{a}_2 = 1 \end{cases} \end{aligned}$$

La primera restricción corresponde a:

$$\begin{aligned} \text{Cov}(z_2, z_1) &= \text{Cov} \left( \underline{a}_2' \underline{x}, \underline{a}_1' \underline{x} \right) \\ &= \underline{a}_2' \Sigma_{xx} \underline{a}_1 \\ &= 0 \end{aligned}$$

Multiplicando (3.7) por  $\underline{a}_2'$  por la izquierda, se tiene:

$$\underline{a}_2' \Sigma_{xx} \underline{a}_1 = \lambda_1 \underline{a}_2' \underline{a}_1 = 0 \Rightarrow \underline{a}_2' \underline{a}_1 = 0 \quad (3.9)$$

---

<sup>3</sup> Ver Anexo A

### Capítulo 3. Métodos de Regresión por Variables Latentes, RVL

Sea  $\phi$  la función a maximizar que contiene las penalidades y los multiplicadores de Lagrange correspondientes, es decir:

$$\phi = \underline{a}_2' \Sigma_{xx} \underline{a}_2 - \lambda \left( \underline{a}_2' \underline{a}_2 - 1 \right) - \gamma \left( \underline{a}_2' \underline{a}_1 \right) \quad (3.10)$$

Obteniendo la derivada simbólica de  $\phi$  con respecto a  $\underline{a}_2$  e igualado a cero:

$$\frac{\partial \phi}{\partial \underline{a}_2} = 2 \Sigma_{xx} \underline{a}_2 - 2\lambda \underline{a}_2 - \gamma \underline{a}_1 = 0$$

Multiplicando por  $\underline{a}_1'$  por la izquierda se obtiene:

$$2 \underline{a}_1' \Sigma_{xx} \underline{a}_2 - 2\lambda \underline{a}_1' \underline{a}_2 - \gamma \underline{a}_1' \underline{a}_1 = 0 \quad (3.11)$$

Pero  $\underline{a}_1' \Sigma_{xx} \underline{a}_2 = 0$ ,  $\underline{a}_1' \underline{a}_2 = 0$  y  $\underline{a}_1' \underline{a}_1 = 1$ , entonces  $\gamma = 0$ . Sustituyendo estos resultados en la ecuación anterior, se tiene:

$$2 \Sigma_{xx} \underline{a}_2 = 2\lambda \underline{a}_2 \quad (3.12)$$

Nuevamente por definición de vector propio, el vector de pesos  $\underline{a}_2$  es el vector propio asociado al valor propio más grande, digamos  $\lambda_2$  de  $\Sigma_{xx}$ . Si se multiplica por  $\underline{a}_2'$  a (3.12) por la izquierda, entonces:

$$\underline{a}_2' \Sigma_{xx} \underline{a}_2 = \lambda_2 = V(z_2) \quad (3.13)$$

Note que la varianza de la segunda componente principal corresponde al segundo valor propio más grande de  $\Sigma_{xx}$ .

Continuando con este argumento, la  $j$ -ésima componente principal será la combinación lineal de el vector propio asociado al  $j$ -ésimo valor propio más grande de  $\Sigma_{xx}$  y el vector aleatorio  $\underline{x}$ .

### 3.2 Análisis de Componentes Principales (ACP)

La matriz  $\Sigma_{xx}$  de orden  $p \times p$  es simétrica y semidefinida positiva, entonces los valores propios asociados a ella son reales y no negativos; si  $\text{rango}(\Sigma_{xx}) = r$ , entonces  $\Sigma_{xx}$  tiene  $r$  valores propios distintos de cero y los vectores propios asociados a éstos forman un conjunto linealmente independiente. Por lo que  $r$ , es el número máximo de componentes principales a construir.

Ahora, considerando la matriz de datos  $\mathbf{X}$  de orden  $n \times p$ , el producto matricial que conforma el conjunto de componentes principales, es:

$$\mathbf{Z} = \mathbf{X}\mathbf{A} \quad (3.14)$$

Donde, por construcción  $\mathbf{A}$  es una matriz ortogonal de orden  $p \times r$ , que contiene como columnas los vectores propios asociados a los valores propios del estimador insesgado de la matriz  $\Sigma_{xx}$ , dado por  $\hat{\Sigma}_{xx} = \mathbf{S}_{xx} = \frac{1}{n-1} \mathbf{X}'\mathbf{X}$ , de manera decreciente.

Técnicamente en el ACP,  $\mathbf{X}$  es factorizada por su descomposición en valores singulares, i.e:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}' \quad (3.15)$$

Donde:

$\mathbf{U}$  y  $\mathbf{V}$  son matrices ortogonales, que contienen como columnas los vectores singulares izquierdos y derechos normalizados, respectivamente.

$\mathbf{D}$  es una matriz diagonal que contiene los valores singulares de  $\mathbf{X}$  ordenados de manera decreciente, esto es,  $\mathbf{D} = \text{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_p}\right)$  si  $\mathbf{X}$  es de rango completo.

Una vez especificado lo anterior, es claro que  $\mathbf{A} = \mathbf{V}$ . Multiplicando a (3.15) por la izquierda por  $\mathbf{V}$  y por (3.14), se obtiene:

$$\mathbf{Z} = \mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{D}\mathbf{V}'\mathbf{V} = \mathbf{U}\mathbf{D} \quad (3.16)$$

De aquí, la descomposición bilineal de  $\mathbf{X}$  está dada por:

### Capítulo 3. Métodos de Regresión por Variables Latentes, RVL

$$\mathbf{X} = \mathbf{Z}\mathbf{V}' \quad (3.17)$$

Si el número de componentes principales es menor a  $r$ , la descomposición bilineal de  $\mathbf{X}$  no es exacta, dejando “ruido” en una matriz de residuales, por ejemplo  $\mathbf{E}$ , la cual concentra la variabilidad de  $\mathbf{X}$  que representan las componentes principales que han quedado fuera del análisis, por lo tanto se plantea el siguiente modelo:

$$\mathbf{X} = \mathbf{Z}_a \mathbf{V}_a' + \mathbf{E}_a \quad (3.18)$$

La matriz  $\mathbf{Z}_a$  es definida como matriz de variables latentes y recoge la información de las variables originales en el nuevo sistema  $a$ -dimensional de ejes ortogonales; la matriz  $\mathbf{V}_a$  es la matriz de cargas y recoge los cosenos directores de los nuevos ejes respecto del sistema de ejes original.

$$\begin{array}{ccccccc} \mathbf{X} & \mathbf{z}_1 & \mathbf{v}'_1 & \mathbf{z}_2 & \mathbf{v}'_2 & \mathbf{z}_a & \mathbf{v}'_a & \mathbf{E}_a \\ \left[ \begin{array}{c} \\ \\ \\ \end{array} \right] & = & \left( \begin{array}{c} \\ \\ \\ \end{array} \right) & + & \left( \begin{array}{c} \\ \\ \\ \end{array} \right) & + \dots + & \left( \begin{array}{c} \\ \\ \\ \end{array} \right) & + & \left[ \begin{array}{c} \\ \\ \\ \end{array} \right] \end{array}$$

Figura 3.2<sup>4</sup>: Al producto  $\mathbf{Z}_a \mathbf{V}_a'$  se le conoce como la parte estructural de la descomposición bilineal; la matriz de residuales  $\mathbf{E}_a$  es conocida como la parte “ruido” y, concentra la variabilidad de las componentes principales que han quedado fuera del análisis.

#### 3.2.4 Selección del número de componentes principales

El criterio para el número de componentes principales a retener, está basado en la varianza del sistema explicada por cada una de ellas, pues:

$$\text{Var}(\mathbf{Z}) = \text{Var}(\mathbf{X}\mathbf{V}) = \mathbf{V}\mathbf{S}_{xx}\mathbf{V}' = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_r \end{pmatrix} = \mathbf{\Lambda} \quad (3.19)$$

<sup>4</sup> Fuente: Elaboración propia

### 3.2 Análisis de Componentes Principales (ACP)

Derivado de la descomposición espectral de la matriz  $\mathbf{S}_{xx}$ . Existe relación entre la varianza de las componentes principales y la varianza de las variables regresoras originales, dada por:

$$\begin{aligned}
 \sum_{i=1}^r \text{Var}(z_i) &= \sum_{i=1}^r \lambda_i \\
 &= \text{tr}(\mathbf{\Lambda}) \\
 &= \text{tr}(\mathbf{V} \mathbf{\Sigma}_{xx} \mathbf{V}') \\
 &= \text{tr}(\mathbf{\Sigma}_{xx}) \text{ por propiedad de traza} \\
 &= \sum_{i=1}^r \text{Var}(x_i) \tag{3.20}
 \end{aligned}$$

Entonces, la  $j$ -ésima componente principal aporta el  $\left( \frac{\lambda_j}{\sum_{i=1}^p \lambda_i} \right) \times 100\%$  de la variabilidad total de  $\mathbf{X}$ , llamado “porcentaje de variabilidad explicada”.

El juicio a seguir es: *“retener aquel número de componentes principales tal que la suma de sus varianzas reflejen al menos el 85% de la variabilidad total del sistema”*.

Si las variables originales fueron estandarizadas previamente, entonces la matriz  $\mathbf{S}_{xx}$  corresponde a la matriz de correlaciones  $\mathbf{R}_{xx}$ , dada por:

$$\mathbf{R}_{xx} = \begin{pmatrix} 1 & \rho_{x_2 x_1} & \cdots & \rho_{x_p x_1} \\ \rho_{x_1 x_2} & 1 & \cdots & \rho_{x_p x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{x_1 x_p} & \rho_{x_2 x_p} & \cdots & 1 \end{pmatrix}$$

Y el ACP estará obteniendo los vectores de pesos como los vectores propios asociados a los valores propios de la matriz  $\mathbf{R}_{xx}$ , en orden decreciente.



### 3.2.5 Algoritmo NIPALS ACP

Existen diversos métodos que permiten ejecutar el ACP de manera sencilla, uno de ellos es el algoritmo Nonlinear Iterative Partial Least Squares<sup>5</sup> (NIPALS), diseñado por Herman Wold (1966). Este algoritmo calcula las componentes principales de forma iterativa, lo cual permite ahorrar tiempo de cálculo al predeterminar el número de componentes principales a obtener.

Los pasos de NIPALS ACP son mostrados a continuación:

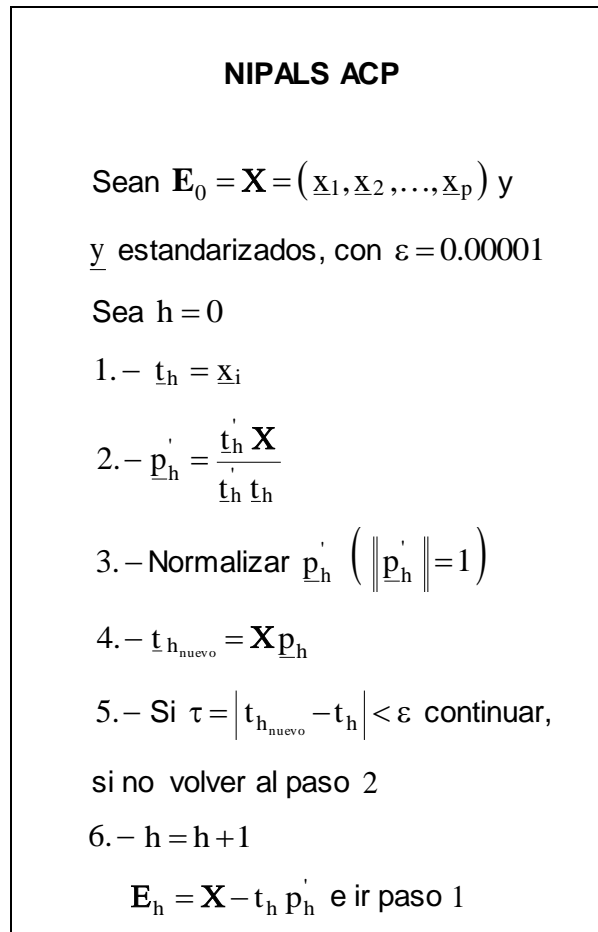


Figura 3.3<sup>6</sup>: Algoritmo NIPALS ACP

<sup>5</sup> Wold, H. (1966)

<sup>6</sup> Wold, H. (1966)

### 3.2 Análisis de Componentes Principales (ACP)

El número máximo de iteraciones de NIPALS es el rango de  $\mathbf{X}$  ( $\text{rango}(\mathbf{X}'\mathbf{X}) = r$ ). De los pasos 2 y 4,  $\underline{t}_h$  y  $\underline{p}_h$  son escalares, puesto que  $\underline{t}_h$  y  $\underline{p}_h$  son vectores de orden  $n$  y  $p$ , respectivamente; substituyendo dichos valores por un valor constante, por ejemplo  $\kappa_i$ , para la primera iteración se tienen las expresiones:

$$\kappa_1 \underline{p}_h' = \underline{t}_h' \mathbf{X} \quad (3.21)$$

$$\kappa_2 \underline{t}_h = \mathbf{X} \underline{p}_h \quad (3.22)$$

Despejando  $\underline{t}_h$  de (3.22) y substituyendo en (3.21) se obtiene:

$$\begin{aligned} \underline{t}_h &= \frac{\mathbf{X} \underline{p}_h}{\kappa_2} \\ \Rightarrow \kappa_3 \underline{p}_h' &= (\mathbf{X} \underline{p}_h)' \mathbf{X} \end{aligned}$$

Con  $\kappa_3 = \kappa_1 \kappa_2$ . Aplicando la transposición:

$$\kappa_3 \underline{p}_h = \mathbf{X}' \mathbf{X} \underline{p}_h$$

De acuerdo con la definición de vector propio, encontramos que en NIPALS ACP las cargas son los vectores propios de  $\mathbf{X}'\mathbf{X}$ . De manera similar, despejando  $\underline{p}_h$  de (3.21)

obtenemos:  $\underline{p}_h' = \frac{\underline{t}_h' \mathbf{X}}{\kappa_1}$ ; aplicando la transposición se tiene que  $\underline{p}_h = \frac{\mathbf{X}' \underline{t}_h}{\kappa_1}$  y

substituyendo en (3.22), se obtiene:

$$\kappa_4 \underline{t}_h = \mathbf{X} \mathbf{X}' \underline{t}_h \quad (3.23)$$

Así, siguiendo NIPALS, las componentes principales son vectores propios de la matriz  $\mathbf{X} \mathbf{X}'$ . Después de la primera iteración, NIPALS sustrae la proporción de variabilidad del sistema que representa la primera componente principal, entonces para la segunda iteración, se obtiene la segunda componente principal, que representa la mayor variabilidad después de la primera, y así sucesivamente. La actualización de la matriz

### Capítulo 3. Métodos de Regresión por Variables Latentes, RVL

de datos (paso 6 de NIPALS), permite construir cada elemento de la descomposición bilineal de  $\mathbf{X}$ . Cuando la descomposición bilineal de la matriz  $\mathbf{X}$  es exacta, la variabilidad total del sistema ha sido capturada por las componentes principales y NIPALS, termina cuando la matriz de residuales es nula (al  $r$ -ésimo paso).

#### 3.2.6 Regresión por Componentes Principales (RCP)

La regresión por componentes principales ajusta un modelo predictivo sobre el sistema de coordenadas generado por las componentes principales retenidas, llevando la dependencia lineal entre  $\mathbf{X}$  y  $\underline{y}$  a través de éstas, es decir, toda la variabilidad que describe a  $\mathbf{X}$  es utilizada para predecir  $\underline{y}$ . Por (3.18), se tiene que el modelo de regresión para  $\underline{y}$  es:

$$\begin{aligned}\underline{y}_{\text{RCP}}^a &= (\mathbf{Z}_a \mathbf{V}_a' + \mathbf{E}_a) \underline{\beta}_{\text{MCO}} + \underline{f} \\ &= \mathbf{Z}_a \mathbf{V}_a' \underline{\beta}_{\text{MCO}} + (\mathbf{E}_a \underline{\beta}_{\text{MCO}} + \underline{f}) \\ &= \mathbf{Z}_a \underline{\beta}_{\text{RCP}}^a + \underline{f}^*\end{aligned}$$

Mediante la RCP, el estimador para  $\beta$ ,  $\hat{\underline{\beta}}_{\text{RCP}}^a$ , debe minimizar  $\left\| \underline{y} - \hat{\underline{y}}_{\text{RCP}}^a \right\|^2$ , esto es:

$$\hat{\underline{\beta}}_{\text{RCP}}^a = \mathbf{Z}_a (\mathbf{Z}_a' \mathbf{Z}_a)^{-1} \mathbf{Z}_a' \underline{y}$$

# CAPÍTULO 4

## MÍNIMOS CUADRADOS PARCIALES (MCP)

### 4.1 Introducción

La técnica de mínimos cuadrados parciales (MCP), fue desarrollada por Herman Wold y colaboradores en la década de los sesenta como una técnica económica; su hijo Svante Wold ha realizado diversas modificaciones, consolidándola como herramienta estándar, desde 1980 hasta la fecha, de un área de la química analítica, la quimiometría<sup>1</sup>. En otros campos de investigación como la mercadotecnia, la ingeniería y la psicología, ha tenido gran aceptación. Las ideas motivadoras para la creación de MCP fueron heurísticas y durante mucho tiempo fue conocido como un método de optimización empírico, por lo que se han desarrollado diversos artículos tratando de encontrar sus propiedades estadísticas; este interés se debe a su buen funcionamiento, que la ha caracterizado como uno de los métodos mayormente implementados. Se puede resumir el trabajo de Wold en el algoritmo Non-linear Partial Least Squares, donde en un principio Wold plasmó la técnica de ACP, después siguiendo otro criterio de maximización, insertó al algoritmo nuevos supuestos y condiciones de tal forma que obtuvo NIPALS MCP<sup>2</sup>; varios autores han realizado modificaciones a este algoritmo, para dar origen a otras técnicas tales como MCP-2WA, MCP-SVD, O-MCP, etc., es por esto que el método de MCP, no es un algoritmo sino una clase de algoritmos, sin embargo existe dificultad en la comparación entre dos o más conjuntos de variables latentes obtenidas de diferentes versiones, como se detallará más adelante. De manera general, MCP combina algunas características de ACP y de RLM, para dar origen a una técnica que en la práctica resulta más potente y menos restrictiva que las anteriores por si solas; por lo cual se le considera de “suave

---

<sup>1</sup> “Quimiometría es la ciencia que permite relacionar las medidas realizadas en un sistema químico o proceso con el estado del mismo, mediante la aplicación de métodos matemáticos o estadísticos”. ICS (International Chemometrics Society)

<sup>2</sup> H. Wold diseñó Non-Linear Partial Least Squares (NIPALS) versión para ACP y versión MCP; para fines prácticos, en este trabajo se nombrará NIPALS ACP al algoritmo para el ACP y NIPALS MCP1 y NIPALS MCP2, para la versión de MCP univariada y multivariada, respectivamente.

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

modelado”. En palabras de Wold, MCP es “distribution-free” (de distribución libre), ya que la especificación del modelo no incluye un modelo probabilístico o de máxima verosimilitud. MCP, es una técnica lineal de correlación multivariada y de reducción de dimensionalidad. Pertenece a los métodos RVL por llevar la dependencia lineal entre  $\underline{X}$  y  $\underline{y}$  a través de variables latentes, pues realiza una transformación ortogonal a las variables regresoras originales, para obtener el conjunto de variables latentes tales que éstas expliquen, tanto como sea posible la covarianza entre las variables independientes y las dependientes, es decir, la construcción de las variables latentes se ve influenciada por los datos dados en  $\underline{X}$  y en  $\underline{y}$ ; las variables latentes explican la estructura común entre ambos conjuntos, esto trae consigo una maximización en el ajuste del modelo de regresión, ya que al tomar el papel de variables regresoras, siguiendo la filosofía de reducir la dimensionalidad de los datos al retener pocas variables latentes, poseen gran capacidad predictiva para  $\underline{y}$ . Tal capacidad y flexibilidad han convertido a MCP en una técnica ampliamente usada y recomendada cuando la meta principal es la predicción. Por otro lado, la calidad de interpretación está directamente ligada al número de variables latentes retenidas; el criterio de retención se basa usualmente en técnicas de validación como validación cruzada, o bien, considerando la varianza de  $\underline{X}$  y de  $\underline{y}$  explicada por cada variable latente, tema que se toca en la sección 4.4. Las metas específicas de MCP son:

1. Extraer la información de la matriz de datos  $\underline{X}$  que esté altamente correlacionada con  $\underline{y}$ .
2. Reducir la dimensionalidad de  $\underline{X}$  manteniendo esa información.
3. Simplificar la descripción de las variables independientes
4. Analizar la estructura común entre  $\underline{X}$  y  $\underline{y}$

Es útil cuando:

1. El número de variables regresoras excede el número observaciones
2. Existe algún grado de multicolinealidad en la matriz de datos  $\underline{X}$ .

En Quimiometría, MCP es una herramienta estándar para modelar las relaciones lineales entre variables multivariadas, obtenidas del cálculo de concentraciones u otras propiedades, a partir de datos instrumentales. Esta es un área de intensa actividad, por sus amplias aplicaciones en la industria química, de procesos y en estudios

## 4.1 Introducción

ambientales. Cuando los datos son obtenidos por métodos instrumentales en química analítica, es frecuente encontrar un mayor número de variables independientes que muestras, por ejemplo, los espectrogramas son usados para estimar la cantidad de diferentes componentes en una muestra química, los cuales pueden ser cientos y altamente multicolineales, entonces la regresión por MCP (RMCP) es implementada para predecir muestras químicas futuras. En la literatura, el método de MCP está dividido en dos casos, el caso univariado denotado por MCP1, que contempla una sola variable dependiente y el caso multivariado MCP2, con más de una variable dependiente. En este trabajo se desarrollan ambos algoritmos. La complejidad que representa el caso multivariado lleva a una generalización del univariado bajo supuestos y condiciones adicionales, es por ello que no se presentan ambos casos en un solo modelo general. A continuación se desarrolla el caso MCP1 y su modelo predictivo final, RMCP1 (Regresión por Mínimos Cuadrados Parciales Univariado), posteriormente MCP2 y RMCP2; se discutirán las características, propiedades e interpretación de las variables latentes en ambos casos. Los métodos MCP1 y MCP2, como cualquier método de regresión, son correlativos y no causales, lo que significa que las relaciones determinadas serán aplicables únicamente a muestras similares (i.e. cuyas propiedades estén dentro del mismo rango), a las usadas para la creación del modelo.

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

El modelo a estudiar en el caso univariado de MCP es  $\underline{y} = \underline{X}\underline{\beta} + \underline{f}$  con  $\underline{y}_{n \times 1}$  y  $\underline{X}_{n \times p}$ .

Como todo método RVL, la relación lineal entre  $\underline{X}$  y la variable de respuesta es llevada a través del conjunto de variables latentes, esto significa que  $\underline{y}$  también es explicada por este conjunto, dando origen a los modelos siguientes:

$$\underline{X} = \underline{T}_a \underline{P}_a' + \underline{E}_a \quad (4.1)$$

$$\underline{y} = \underline{T}_a \underline{q}_a + \underline{f}_a \quad (4.2)$$

La técnica MCP1 diseñada por Wold, construye de manera iterativa el conjunto de variables latentes  $(\underline{t}_1, \dots, \underline{t}_a) \in \mathfrak{R}^{n \times a}$  tal que sea mutuamente ortogonal y que explique tanto como sea posible la covarianza entre  $\underline{X}$  y  $\underline{y}$ , para almacenarlas en la matriz  $\underline{T}_a$  como columnas.

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

### 4.2.1 Invarianza de escala

La técnica de MCP no es invariante bajo cambios de escala por lo que se remueve la escala de las variables originales antes de comenzar el análisis al aplicar la estandarización como en la sección 2.2. Para detallar este punto, como MCP elige a aquellos vectores de pesos que maximicen la covarianza entre  $\underline{X}$  y  $\underline{y}$ , y la covarianza es una medida de asociación lineal que retiene las unidades de medida de las variables, entonces si existen algunas variables regresoras con covarianza grande con la variable de respuesta, originado por las diferencias de escala, estas variables dominarán los vectores de pesos y a través de ellos, los resultados de MCP, situación que puede no ser benéfica al dejar conclusiones alejadas de la realidad.

El estandarizar los datos implica dar igual importancia a las variables de manera a priori, esto brinda estabilidad numérica al conjunto de vectores de pesos. Desde un punto de vista teórico, realizar el análisis con los datos estandarizados siempre puede efectuarse, sin embargo existen ciertas circunstancias prácticas en las cuales podría la estandarización resultar contraproducente, por tanto se deja a decisión del analista hacerlo o no. Cabe señalar que si la estandarización fue hecha, para interpretar los resultados se deben regresar las variables a sus unidades de medida originales, con el propósito de dar contexto a los resultados.

En este trabajo, para el desarrollo teórico de MCP, se supondrá de ahora en adelante que  $\underline{X}$  y  $\underline{y}$  han sido estandarizadas como en sección 2.2.

### 4.2.2 Planteamiento matemático

Sea  $\underline{x} = (x_1 \ x_2 \ \dots \ x_p)'$  un vector aleatorio y  $\underline{w}_1 = (w_{11} \ w_{21} \ \dots \ w_{p1})'$  el vector de pesos. La primera variable latente  $t_1$ , definida como  $t_1 = \underline{x}' \underline{w}_1$ , debe tener covarianza máxima con la variable de respuesta  $y$ ; para este fin, se elige aquel vector de pesos  $\underline{w}_1 \neq 0$ , sujeto a  $\underline{w}_1' \underline{w}_1 = 1$  tal que la maximización se cumpla.

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

Se define al vector de covarianzas de  $\underline{x}$  y  $y$ , como:

$$\underline{\sigma}_{Xy} = \left( \text{Cov}(x_1, y) \quad \text{Cov}(x_2, y) \quad \cdots \quad \text{Cov}(x_p, y) \right)' \quad (4.3)$$

Note que  $\underline{\sigma}_{Xy} \neq \underline{0}$ , debido a que el análisis de regresión establece dependencia de  $y$  sobre  $\underline{x}$ . El criterio de maximización es sintetizado en la siguiente expresión:

$$\begin{aligned} \text{Max}(\text{Cov}(t_1, y))^2 &= \text{Max}(\text{Cov}(\underline{x}' \underline{w}_1, y))^2 \\ &= \text{Max}(\underline{w}_1' \underline{\sigma}_{Xy})^2 \\ &= \text{Max}(\underline{w}_1' \underline{\sigma}_{Xy} \underline{\sigma}_{Xy}' \underline{w}_1) \end{aligned} \quad (4.4)$$

Sea  $\varphi$  la función a maximizar que contiene la penalidad  $\underline{w}_1' \underline{w}_1 = 1$ , utilizando el multiplicador de Lagrange  $\lambda$ , se tiene:

$$\varphi = \underline{w}_1' \underline{\sigma}_{Xy} \underline{\sigma}_{Xy}' \underline{w}_1 - \lambda (\underline{w}_1' \underline{w}_1 - 1)$$

Derivando con respecto a  $\underline{w}_1$  obtenemos:

$$\frac{\partial \varphi}{\partial \underline{w}_1} = 2 \underline{\sigma}_{Xy} \underline{\sigma}_{Xy}' \underline{w}_1 - 2\lambda \underline{w}_1$$

Igualando a cero:

$$\begin{aligned} 2 \underline{\sigma}_{Xy} \underline{\sigma}_{Xy}' \underline{w}_1 - 2\lambda \underline{w}_1 &= 0 \\ \Leftrightarrow \underline{\sigma}_{Xy} \underline{\sigma}_{Xy}' \underline{w}_1 &= \lambda \underline{w}_1 \end{aligned} \quad (4.5)$$

Por definición de vector propio y valor propio,  $\underline{w}_1$  es vector propio de  $\underline{\sigma}_{Xy} \underline{\sigma}_{Xy}'$  asociado al valor propio más grande, digamos  $\lambda_1$ , para lograr maximizar la función.

Por otro lado, multiplicando por la izquierda por  $\underline{w}_1'$  y usando  $\underline{w}_1' \underline{w}_1 = 1$ , se obtiene:

$$\underline{w}_1' \underline{\sigma}_{Xy} \underline{\sigma}_{Xy}' \underline{w}_1 = \lambda_1 \underline{w}_1' \underline{w}_1 = \lambda_1$$



## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

Multiplicando por la izquierda a (4.5) por  $\underline{\sigma}'_{Xy}$ , resulta:

$$\begin{aligned} \underline{\sigma}'_{Xy} \underline{\sigma}_{Xy} \underline{\sigma}'_{Xy} \underline{w}_1 &= \lambda_1 \underline{\sigma}'_{Xy} \underline{w}_1 \\ \Rightarrow \left( \underline{\sigma}'_{Xy} \underline{\sigma}_{Xy} - \lambda_1 \right) \underline{\sigma}'_{Xy} \underline{w}_1 &= 0 \end{aligned}$$

En esta ecuación  $\underline{\sigma}'_{Xy} \underline{\sigma}_{Xy} - \lambda_1 = 0$  ó  $\underline{\sigma}'_{Xy} \underline{w}_1 = 0$ , sin embargo  $\underline{\sigma}'_{Xy} \underline{w}_1 \neq 0$  por ser la función que se desea maximizar, entonces:

$$\underline{\sigma}'_{Xy} \underline{\sigma}_{Xy} = \left\| \underline{\sigma}_{Xy} \right\|^2 = \lambda_1$$

Por lo tanto, el vector  $\underline{w}_1$  que maximiza (4.4) es:

$$\underline{w}_1 = \frac{\underline{\sigma}_{Xy}}{\left\| \underline{\sigma}_{Xy} \right\|} \quad (4.6)$$

Ahora, considerando la matriz de datos  $\mathbf{X}$  de orden  $n \times p$  y las observaciones de la variable de respuesta dadas en  $\underline{y}$ , entonces el estimador insesgado para  $\underline{\sigma}_{Xy}$ , con  $\mathbf{X}$  y  $\underline{y}$  centradas, es:

$$\underline{\hat{\sigma}}_{Xy} = \underline{s}_{Xy} = \frac{\mathbf{X}' \underline{y}}{n-1}$$

Sustituyendo lo anterior en (4.6), se tiene:

$$\underline{w}_1 = \frac{\mathbf{X}' \underline{y}}{\left\| \mathbf{X}' \underline{y} \right\|} \quad (4.7)$$

Ya que:  $\left\| \underline{s}_{Xy} \right\| = (n-1)^{-1} \left\| \mathbf{X}' \underline{y} \right\| \Rightarrow \underline{w}_1 = \frac{\underline{s}_{Xy}}{\left\| \underline{s}_{Xy} \right\|} = \frac{(n-1)^{-1} \mathbf{X}' \underline{y}}{(n-1)^{-1} \left\| \mathbf{X}' \underline{y} \right\|}$

Por tanto, la primera variable latente de MCP1 queda definida como:

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

$$\hat{t}_1 = \mathbf{X} \underline{w}_1 = \mathbf{X} \left( \frac{\mathbf{X}' \underline{y}}{\|\mathbf{X}' \underline{y}\|} \right) \quad (4.8)$$

Es válido calcular a  $\underline{w}_1$  como  $\underline{w}_1 = (\mathbf{n} - 1)^{-1} \mathbf{X}' \underline{y}$  seguido de su normalización, esto nos permite escribir matricialmente lo siguiente:

$$\underline{w}_1 = (\mathbf{n} - 1)^{-1} \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = (\mathbf{n} - 1)^{-1} \begin{pmatrix} \sum_{i=1}^n x_{i1} y_i \\ \sum_{i=1}^n x_{i2} y_i \\ \vdots \\ \sum_{i=1}^n x_{ip} y_i \end{pmatrix}$$

Debido a la estandarización de las variables se tiene que el  $j$ -ésimo elemento de  $\underline{w}_1$ , digamos  $w_{j1}$ , es el estimador del coeficiente de regresión lineal simple  $\hat{\beta}_1^j$ , del modelo  $\underline{y} = \beta_0 + \underline{x}_j \beta_1$  (antes de continuar se denota a  $\hat{\beta}_1^j$  como el estimador coeficiente de regresión  $\beta_1$  del modelo  $\underline{y} = \hat{\beta}_0 + \underline{x}_j \hat{\beta}_1$ ), para  $j = 1, 2, \dots, p$ , pues:

$$w_{j1} = (\mathbf{n} - 1)^{-1} \sum_{i=1}^n x_{ij} y_i \left( \frac{\sum_{i=1}^n (x_{ij})^2}{\sum_{i=1}^n (x_{ij})^2} \right) = (\mathbf{n} - 1)^{-1} \sum_{i=1}^n \left( \frac{x_{ij} y_i}{\sum_{i=1}^n (x_{ij})^2} \right) \left( \sum_{i=1}^n (x_{ij})^2 \right) = \hat{\beta}_1^j$$

Ya que se suponen a las variables independientes y dependientes estandarizadas. Para corroborar ésto, considere a  $\underline{x}_j^*$  como la  $j$ -ésima variable dependiente antes de la estandarización, entonces:

$$\sum_{i=1}^n (x_{ij})^2 = \sum_{i=1}^n \left( \frac{x_{ij}^* - \bar{x}_j}{\sqrt{\frac{\sum_{i=1}^n (x_{ij}^* - \bar{x}_j)^2}{\mathbf{n} - 1}}} \right)^2 = (\mathbf{n} - 1) \frac{\sum_{i=1}^n (x_{ij}^* - \bar{x}_j)^2}{\sum_{i=1}^n (x_{ij}^* - \bar{x}_j)^2} = \mathbf{n} - 1$$

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

Y normalizar  $\underline{w}_1$  ( $\|\underline{w}_1\|=1$ ), entonces  $w_{j1} \propto \hat{\beta}_1^j$ . De esta manera, la variable latente  $t_1$  resulta ser una combinación lineal de las columnas de  $\mathbf{X}$ , donde los factores son proporcionales a los  $\hat{\beta}_1^j$ , así MCP1 encuentra las columnas de  $\mathbf{X}$  mayormente correlacionadas con  $\underline{y}$  y,  $t_1$  puede ser entendido como la mejor combinación lineal de las columnas de  $\mathbf{X}$  para el propósito de predecir  $\underline{y}$ . Además, si la estandarización de las variables originales fue hecha, se cumple que el vector de covarianzas muestral corresponde al vector de correlaciones muestrales, por lo tanto, se puede decir que la primera variable latente maximiza correlación y covarianza al mismo tiempo; entonces,  $t_1$  incluye variación sistemática en  $\mathbf{X}$  fuertemente correlacionada con  $\underline{y}$ .

Con el fin de reducir los cálculos, se normaliza a  $t_1$ , i.e.:

$$\underline{t}_1 \leftarrow \frac{t_1}{\|\underline{t}_1\|} \quad (4.9)$$

La variable latente  $t_1$  debe ser removida de los datos para comenzar las descomposiciones bilineales de  $\mathbf{X}$  y de  $\underline{y}$ , dejando en las matrices de residuales la información no explicada por  $t_1$ , como sigue:

$$\underline{y} = t_1 q_1 + \underline{f}_1 \Rightarrow \underline{f}_1 = \underline{y} - t_1 q_1 \quad (4.10)$$

En donde, se proyecta  $\underline{y}$  sobre  $\underline{t}_1$ , tal que  $\|\underline{y} - t_1 q_1\|^2$  sea mínima, por tanto la carga para  $\underline{y}$  es  $q_1 = \underline{t}_1' \underline{y}$  (por ser  $\underline{t}_1$  de norma unitaria). Para el desarrollo teórico de MCP, se omite el símbolo “^” o bien, de la palabra estimación en el cálculo de los vectores de cargas, pues así se presenta en la literatura. El vector  $\underline{f}_1$  es el complemento ortogonal a la proyección y contempla la cantidad de variabilidad de  $\underline{y}$  no explicada por  $t_1$ . Equivalentemente se descompone a  $\mathbf{X}$ , mediante el producto de dos vectores, como sigue:

$$\mathbf{X} = t_1 \underline{p}_1' + \mathbf{E}_1 \Rightarrow \mathbf{E}_1 = \mathbf{X} - t_1 \underline{p}_1' \quad (4.11)$$

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

Tal que  $\underline{p}'_1 = \underline{t}'_1 \mathbf{X}$ . Por la especificación del vector de pesos  $\underline{w}_1$ , la descomposición bilineal de  $\mathbf{X}$  está guiada por la variabilidad en  $\underline{y}$ ; en (4.11) se extrae la variación de  $\mathbf{X}$  directamente correlacionada con  $\underline{y}$ , y la variación restante está contemplada en la matriz de residuales  $\mathbf{E}_1$ . Es importante señalar que, si todas las variables regresoras presentan dependencia lineal con la variable de respuesta, entonces todas las columnas de  $\mathbf{X}$  son relevantes para  $\underline{y}$ , y el método de MCP1 no requiere calcular más de una variable latente;  $\underline{t}_1$  explicará “casi” perfectamente a  $\underline{y}$  y el producto  $\underline{t}_1 q_1$ , dejará solo “ruido blanco” o residuales puramente aleatorios en  $\underline{f}_1$ . El cálculo de variables latentes adicionales, proviene de un error dado en la relación funcional entre  $\mathbf{X}$  y  $\underline{y}$ ; si una o varias columnas de  $\mathbf{X}$  no están correlacionadas con  $\underline{y}$  o, lo hacen de forma no lineal, la ponderación de las columnas de  $\mathbf{X}$  que realiza  $\underline{t}_1$  tendrá perturbación y las proyecciones de  $\mathbf{X}$  y de  $\underline{y}$  sobre  $\underline{t}_1$  la heredarán, así los residuales  $\underline{f}_1$  y  $\mathbf{E}_1$  contendrán variación sistemática, requiriendo calcular más de una variable latente, para terminar ambas descomposiciones bilineales. Por esta razón, el algoritmo NIPALS MCP1 se plantea iterativo hasta la  $h$ -ésima iteración, tal que  $\mathbf{E}_{h+1}$  no contribuya significativamente a la descripción de  $\underline{f}_{h+1}$ , en otras palabras, hasta que  $\mathbf{E}_{h+1}$  y  $\underline{f}_{h+1}$  estén no correlacionadas entre sí. Si existe alta dimensionalidad de variables regresoras, será muy probable que suceda esta problemática.

Si existe perturbación en la relación funcional entre  $\mathbf{X}$  y  $\underline{y}$ , para la información no explicada por  $\underline{t}_1$  y contenida en los residuales  $\mathbf{E}_1$  y  $\underline{f}_1$ , se construirá la segunda variable latente  $\underline{t}_2$ , siguiendo el mismo criterio de maximización que para la primera, pero ahora, sobre los residuales; el vector de pesos  $\underline{w}_2$  será elegido como

$$\underline{w}_2 = \frac{\mathbf{E}'_1 \underline{f}_1}{\|\mathbf{E}'_1 \underline{f}_1\|}, \text{ por realizar el mismo planteamiento matemático que para } \underline{w}_1. \text{ Los}$$

pasos subsecuentes son realizados al proyectar los residuales  $\mathbf{E}_{j-1}$  y  $\underline{f}_{j-1}$  sobre la  $j$ -ésima variable latente.

La técnica de MCP1, es sintetizada en algoritmo NIPALS MCP1.

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

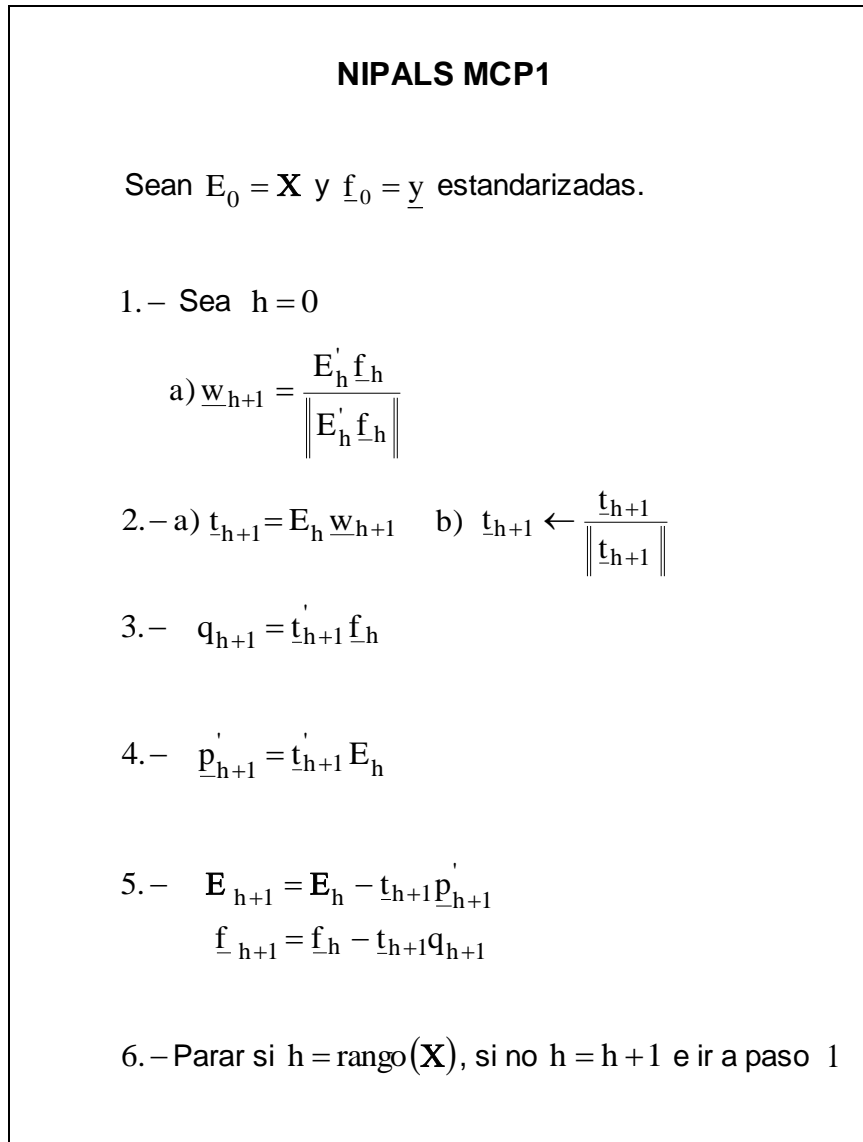


Figura 4.1<sup>3</sup>: Algoritmo NIPALS MCP1

La construcción de las cargas  $\underline{p}'_h$  y  $\underline{q}_h$  en cada iteración, dio origen al término “Mínimos Cuadrados Parciales”. El acrónimo PLS (Partial Least Squares, en inglés) se ha interpretado también como proyección a estructuras latentes<sup>4</sup> (Projection to Latents Structures), con base en las proyecciones efectuadas para la obtención de las cargas en cada iteración.

Por otro lado, (4.5) permite tomar directamente los vectores propios asociados al valor propio  $\lambda_{h_1}$  más grande, de las matrices  $\mathbf{E}'_h \underline{f}_h \underline{f}'_h \mathbf{E}_h$  para  $h = 1, 2, \dots, r$ , puesto que el criterio de maximización debe cumplirse para los residuales  $\mathbf{E}_{h-1}$  y  $\underline{f}_{h-1}$ , es decir:

<sup>3</sup> Ergon, R. (2009); Helland, I. (1988)

<sup>4</sup> Abdi, H. (2010)

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

$$\text{Max}\left(\text{Cov}\left(\underline{w}_h' \mathbf{E}_{h-1}, \underline{f}_{h-1}\right)^2\right) = \text{Max}\left(\underline{w}_h' \underline{\sigma}_{\mathbf{E}_{h-1} \underline{f}_{h-1}} \underline{w}_h\right) \quad (4.12)$$

$$\text{Sujeto a: } \|\underline{w}_h\| = 1$$

Efectuando el mismo desarrollo matemático que para  $\underline{w}_1$ , y sustituyendo la siguiente expresión:

$$\widehat{\underline{\sigma}}_{\mathbf{E}_{h-1} \underline{f}_{h-1}} = \underline{s}_{\mathbf{E}_{h-1} \underline{f}_{h-1}} = \frac{1}{n-1} \mathbf{E}_{h-1}' \underline{f}_{h-1}$$

Se tiene:  $\widehat{\underline{\sigma}}_{\mathbf{E}_{h-1} \underline{f}_{h-1}} \widehat{\underline{\sigma}}_{\mathbf{E}_{h-1} \underline{f}_{h-1}}' \underline{w}_h = \lambda_{h_1} \underline{w}_h$  tal que  $\lambda_{h_1} > \lambda_{h_2} > \dots > \lambda_{h_r}$ , así

$$\mathbf{E}_{h-1}' \underline{f}_{h-1} \underline{f}_{h-1}' \mathbf{E}_{h-1} \underline{w}_h = \lambda_{h_1} \underline{w}_h \quad (4.13)$$

Pues  $(n-1)^2 \lambda_{h_1} > (n-1)^2 \lambda_{h_2} > \dots > (n-1)^2 \lambda_{h_r}$ . Es evidente que estamos bajo un marco de vectores singulares derechos, asociados a los primeros valores singulares del producto  $\mathbf{E}_{h-1}' \underline{f}_{h-1}$ , de manera iterativa. Entonces, podemos decir que MCP1 es lo más parecido a realizar la descomposición en valores singulares sobre  $\mathbf{E}_{h-1}' \underline{f}_{h-1}$  iterativamente. Para la primera iteración, se tiene:

$$\mathbf{X}' \underline{y} = \mathbf{U} \Delta \mathbf{V}'$$

De aquí  $\underline{w}_1 = \underline{v}_1$ . A diferencia de ACP, donde el conjunto de vectores de pesos  $\{\underline{w}_1, \dots, \underline{w}_a\}$  es el conjunto de vectores singulares derechos de  $\mathbf{X}$ , y por tanto, se requiere ejecutar la descomposición en valores singulares de  $\mathbf{X}$  solo una vez, para obtener el conjunto de componentes principales, en MCP1 y como se detalla en MCP2, se necesita ejecutar esta descomposición varias veces, puesto que el fundamento es construir las variables latentes  $\{t_i\}_{i=1}^a$  mutuamente ortogonales en  $\mathfrak{R}^n$ ; si  $\mathbf{T} = \mathbf{X} \mathbf{V}$ , entonces  $\mathbf{T}$  no necesariamente será ortogonal, ya que al multiplicar por la izquierda por  $\mathbf{X}$  a (4.5), se tiene:

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

$$\mathbf{X}\mathbf{X}'\mathbf{y}\mathbf{y}'\mathbf{X}\mathbf{w} = \lambda\mathbf{w} \Rightarrow \mathbf{X}\mathbf{X}'\mathbf{y}\mathbf{y}'\mathbf{t} = \lambda\mathbf{t} \text{ para algún vector } \mathbf{t}$$

Por la definición de variable latente. La matriz  $\mathbf{X}\mathbf{X}'\mathbf{y}\mathbf{y}'$  no necesariamente es simétrica, entonces no necesariamente es diagonalizable ortonormal<sup>5</sup> y el conjunto  $\{\mathbf{t}_i\}_{i=1}^a$  no necesariamente será mutuamente ortonormal. Es por esta razón que MCP construye iterativamente el conjunto  $\{\mathbf{t}_i\}_{i=1}^a$ .

Finalmente, todas las variables calculadas en NIPALS MCP1 serán almacenadas como columnas de las siguientes matrices:

$$\mathbf{T}_a = (\mathbf{t}_1 \quad \mathbf{t}_2 \quad \cdots \quad \mathbf{t}_a) := \text{Matriz de variables latentes de } \mathbf{X}$$

$$\mathbf{P}_a = (\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_a) := \text{Matriz de cargas de } \mathbf{X}$$

$$\mathbf{q}_a = (q_1, q_2, \dots, q_a) := \text{Vector de cargas de } \mathbf{y}$$

$$\mathbf{W}_a = (\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_a) := \text{Matriz de vectores de pesos de } \mathbf{X}$$

### 4.2.3 Regresión por Mínimos Cuadrados Parciales Univariado, RMCP1

La meta de MCP1 es describir óptimamente a la variable de respuesta mediante un número reducido de variables latentes, al efectuar un ajuste a  $\mathbf{y}$  por mínimos cuadrados sobre el subespacio generado por dichas variables. Al igual que en ACP, existen diversos criterios para determinar el número de variables latentes que serán utilizadas en el modelo de RMCP1, ligados a la capacidad predictiva; éstos serán vistos en la sección 4.4.

En la RMCP1, se desea minimizar la suma de cuadrados de los residuales, dada por:

$$\text{Min} \left( \mathbf{y} - \hat{\mathbf{y}}_{\text{MCP1}}^a \right)' \left( \mathbf{y} - \hat{\mathbf{y}}_{\text{MCP1}}^a \right) = \text{Min} \left( \mathbf{y} - \mathbf{T}_a \hat{\underline{\beta}}_{\text{MCP1}}^a \right)' \left( \mathbf{y} - \mathbf{T}_a \hat{\underline{\beta}}_{\text{MCP1}}^a \right) \quad (4.14)$$

---

<sup>5</sup> Ver Anexo A

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

NIPALS MCP1 efectúa la regresión de  $\underline{y}$  sobre cada una de las variables latentes (paso 3 del algoritmo) de manera iterativa, construyendo el modelo  $\underline{y} = \mathbf{T}_a \underline{q}_a + \underline{f}_a$  para el cual se espera que  $\|\underline{f}_a\|$  sea pequeña. Es decir, las aproximaciones de rango uno para  $\underline{y}$  fueron efectuadas en cada iteración de tal forma que en la  $a$ -ésima iteración,  $\underline{f}_a$  solo contenga residuales puramente aleatorios o “ruido blanco”, lo cual nos permite escribir:

$$\hat{\underline{y}}_{\text{MCP1}}^a = \mathbf{T}_a \underline{q}_a \quad (4.15)$$

Que denota los valores ajustados para  $\underline{y}$  por MCP1, utilizando como variables regresoras a las primeras  $a$  variables latentes.

Para fines teóricos, se construye el modelo de regresión para  $\underline{y}$  derivado del modelo original  $\underline{y}_{\text{MCO}} = \mathbf{X}\beta_{\text{MCO}} + \varepsilon$ , donde explícitamente intervienen las variables regresoras originales. Antes de continuar, se debe encontrar la relación que existe entre la matriz de pesos  $\mathbf{W}_a$  y la matriz  $\mathbf{X}$ , ésta no es clara ya que los vectores de pesos  $\underline{w}_i$ 's fueron construidos como combinaciones lineales de las matrices de residuales de  $\mathbf{X}$  y de  $\underline{y}$  de pasos anteriores. Entonces sea  $a$  el número para el cual las entradas de  $\mathbf{E}_a$  son casi nulas (despreciables), entonces podemos escribir  $\mathbf{X} = \mathbf{T}_a \mathbf{P}_a'$ ; multiplicando por la derecha por  $\mathbf{W}_a$  se tiene:

$$\mathbf{X}\mathbf{W}_a = \mathbf{T}_a \mathbf{P}_a' \mathbf{W}_a \quad (4.16)$$

$$\mathbf{T}_a = \mathbf{X}\mathbf{W}_a (\mathbf{P}_a' \mathbf{W}_a)^{-1} = \mathbf{X}\mathbf{M}_a \quad (4.17)$$

Con  $\mathbf{M}_a = \mathbf{W}_a (\mathbf{P}_a' \mathbf{W}_a)^{-1}$ . Se asumirá que  $(\mathbf{P}_a' \mathbf{W}_a)^{-1}$  existe (Ver sección 4.2.5, propiedad 3). De aquí, el modelo de regresión por MCP1 que incluye a  $\mathbf{X}$  explícitamente, es:



## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

$$\hat{\underline{y}}_{\text{MCP1}}^a = \mathbf{X} \mathbf{W}_a (\mathbf{P}_a' \mathbf{W}_a)^{-1} \underline{q}_a = \mathbf{X} \hat{\underline{\beta}}_{\text{MCP1}}^a \quad (4.18)$$

Entonces, el estimador del coeficiente de regresión  $\underline{\beta}$  por MCP1 es:

$$\hat{\underline{\beta}}_{\text{MCP1}}^a = \mathbf{W}_a (\mathbf{P}_a' \mathbf{W}_a)^{-1} \underline{q}_a \quad (4.19)$$

### 4.2.4 Predicción

Usando el modelo de regresión (4.15), sea  $\underline{x}_0 = (x_{01} \ x_{02} \ \dots \ x_{0p})'$  una nueva observación de las  $p$  variables de  $\mathbf{X}$ , entonces:

$$\hat{\underline{y}}_{\text{MCP1}}^a = \bar{y}^* + \left( \left( (\underline{x}_0 - \underline{\mu}_x^*)' \Psi \right) \hat{\underline{\beta}}_{\text{MCP1}}^a \right) \sigma_{y^*} \quad (4.20)$$

Con:  $\bar{y}^* = \frac{\underline{y}' \underline{y}}{n}$ ,  $\sigma_{y^*} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}$  son la media muestral y la desviación estándar

muestral de la variable de respuesta original (sin estandarizar); Asimismo  $\underline{\mu}_x^* = \frac{1}{n} \mathbf{X}' \underline{1}_n$

y  $\Psi = \text{diag}(\sigma_{x_1^*}^{-1}, \sigma_{x_2^*}^{-1}, \dots, \sigma_{x_p^*}^{-1})$  con  $\sigma_{x_j^*} = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \mu_j)^2}{n-1}}$ , son el vector de medias muestrales y la matriz diagonal con las desviaciones estándar muestrales de las  $p$  variables regresoras originales.

Es decir, se inicia el proceso de predicción con la nueva unidad de observación estandarizada; (4.20) da los valores ajustados de la variable de respuesta en su unidad de medida original.

### 4.2.5 Propiedades básicas de MCP1

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

**Propiedad 1.** – a) El conjunto de vectores de pesos  $\{\underline{w}_1, \dots, \underline{w}_a\}$  es ortonormal y b) generan el subespacio de Krilov dado por:

$$\mathbf{K}_i(\mathbf{X}'\mathbf{X}, \mathbf{X}'\underline{y}) = \text{gen}\{\mathbf{X}'\underline{y}, \mathbf{X}'\mathbf{X}\mathbf{X}'\underline{y}, \dots, (\mathbf{X}'\mathbf{X})^{i-1}\mathbf{X}'\underline{y}\}$$

**Prueba a)** La ortogonalidad de los vectores de pesos se demostrará por inducción.

Para  $i = 1$ , sean  $\underline{w}_1$  y  $\underline{w}_2$  como en NIPALS MCP1, entonces:

$$\begin{aligned} \underline{w}_1' \underline{w}_2 &\propto \underline{w}_1'(\mathbf{E}_1' \underline{f}_1) = \alpha_1 (\underline{w}_1' (\mathbf{E}_1' \underline{f}_1)) \\ &= \alpha_1 \underline{w}_1' (\mathbf{X} - \underline{t}_1 \underline{p}_1)' \underline{f}_1 \\ &= \alpha_1 \underline{w}_1' (\mathbf{X} - \underline{t}_1 (\underline{t}_1' \mathbf{X}))' \underline{f}_1 \\ &= \alpha_1 \underline{w}_1' \mathbf{X}' (\mathbf{I} - \underline{t}_1 \underline{t}_1')' \underline{f}_1 \\ &= \alpha_1 \underline{t}_1' (\mathbf{I} - \underline{t}_1 \underline{t}_1')' \underline{f}_1 \\ &= \alpha_1 (\underline{t}_1' - \underline{t}_1')' \underline{f}_1 \\ &= \underline{0} \end{aligned}$$

Se cumple para  $i = 1$ . Supongamos que se cumple el supuesto de ortogonalidad para  $i = 1, 2, \dots, n$ . Por demostrar que cumple para  $n + 1$ , entonces:

$$\begin{aligned} \underline{w}_n' \underline{w}_{n+1} &\propto \underline{w}_n'(\mathbf{E}_n' \underline{f}_n) = \alpha_n (\underline{w}_n' (\mathbf{E}_n' \underline{f}_n)) \\ &= \alpha_n \underline{w}_n' (\mathbf{E}_{n-1} - \underline{t}_n \underline{p}_n)' \underline{f}_n \\ &= \alpha_n \underline{w}_n' \mathbf{E}_{n-1}' (\mathbf{I} - \underline{t}_n \underline{t}_n')' \underline{f}_n \\ &= \alpha_n \underline{t}_n' (\mathbf{I} - \underline{t}_n \underline{t}_n')' \underline{f}_n \\ &= \alpha_n (\underline{t}_n' - \underline{t}_n')' \underline{f}_n \\ &= \underline{0} \end{aligned}$$

Pues  $\underline{p}_n' = \underline{t}_n' \mathbf{E}_{n-1}$  y  $\underline{t}_n' = \mathbf{E}_{n-1}' \underline{w}_n$  por definición de la  $n$ -ésima variable latente. ■

**Prueba b)** Se hará por inducción. Se introduce la notación que definen los subespacios de Krilov, esto es:  $\mathbf{K}_i(\mathbf{A}, \mathbf{b}) = \text{gen}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{i-1}\mathbf{b}\}$

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

Primero, según lo que se planteó en NIPALS MCP1, se tienen las siguientes afirmaciones:

$$\text{De (4.7) se tiene que: } \underline{w}_1 = \alpha \mathbf{X}' \underline{y} \Rightarrow \underline{w}_1 \in \text{gen} \{ \mathbf{X}' \underline{y} \}$$

$$\text{De (4.8) se tiene que: } \underline{t}_1 = \delta \mathbf{X} \mathbf{X}' \underline{y} \Rightarrow \underline{t}_1 \in \text{gen} \{ \mathbf{X} \mathbf{X}' \underline{y} \}$$

De (4.8) y (4.9) resulta que:

$$\underline{p}_1 = \gamma (\mathbf{X}' \mathbf{X}) \mathbf{X}' \underline{y} \Rightarrow \underline{p}_1 \in \text{gen} \{ (\mathbf{X}' \mathbf{X}) \mathbf{X}' \underline{y} \}$$

Entonces, de (4.7) se cumple para  $i = 1$ . Suponiendo que se cumplen las siguientes afirmaciones para  $i = 1, 2, \dots, n$ :

$$\underline{w}_n \in \text{gen} \{ \mathbf{X}' \underline{y}, \dots, (\mathbf{X}' \mathbf{X})^{n-1} \mathbf{X}' \underline{y} \} \quad (\text{i})$$

$$\underline{t}_n \in \text{gen} \{ \mathbf{X} \mathbf{X}' \underline{y}, \dots, (\mathbf{X} \mathbf{X}')^{n-1} \mathbf{X} \mathbf{X}' \underline{y} \} \quad (\text{ii})$$

$$\underline{p}_n \in \text{gen} \{ (\mathbf{X}' \mathbf{X}) \mathbf{X}' \underline{y}, \dots, (\mathbf{X}' \mathbf{X})^n \mathbf{X}' \underline{y} \} \quad (\text{iii})$$

Por demostrar que se cumple para  $n + 1$ , entonces:

$$\begin{aligned} \underline{w}_{n+1} &= \alpha_{n+1} (\mathbf{E}'_n \underline{f}_n) \\ &= \alpha_{n+1} (\mathbf{E}'_{n-1} - \underline{p}_n \underline{t}'_n) (\underline{f}_{n-1} - \underline{t}_n \underline{q}_n) \\ &= \alpha_{n+1} (\mathbf{E}'_{n-1} \underline{f}_{n-1} - \mathbf{E}'_{n-1} \underline{t}_n \underline{q}_n - \underline{p}_n \underline{t}'_n \underline{f}_{n-1} + \underline{p}_n \underline{t}'_n \underline{t}_n \underline{q}_n) \\ &= \alpha_{n+1} (\mathbf{E}'_{n-1} \underline{f}_{n-1} - \underline{p}_n \underline{q}_n - \underline{p}_n \underline{q}_n + \underline{p}_n \underline{q}_n) \\ &= \alpha_{n+1} (\mathbf{E}'_{n-1} \underline{f}_{n-1} - \underline{p}_n \underline{q}_n) \\ &= \alpha^* \underline{w}_n - \beta^* \underline{p}_n \end{aligned}$$

Por (i) y (iii), se tiene que:

$$\underline{w}_{n+1} \in \text{gen} \{ \mathbf{X}' \underline{y}, \dots, (\mathbf{X}' \mathbf{X})^n \mathbf{X}' \underline{y} \} = \mathbf{K}_{n+1} (\mathbf{X}' \mathbf{X}, \mathbf{X}' \underline{y}) \quad \blacksquare$$

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

Entonces, los vectores  $\{\underline{w}_1, \dots, \underline{w}_a\}$  forman una base ortonormal del subespacio de Krilov  $\mathbf{K}_a(\mathbf{X}'\mathbf{X}, \mathbf{X}'\underline{y})$ , pues los vectores de pesos tienen norma unitaria.

**Propiedad 2.** – a) El conjunto de variables latentes  $\{t_1, \dots, t_a\}$  es ortonormal y b) generan el subespacio de Krilov dado por:

$$\mathbf{K}_i(\mathbf{X}\mathbf{X}', \mathbf{X}\mathbf{X}'\underline{y}) = \text{gen}\{\mathbf{X}\mathbf{X}'\underline{y}, \mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\underline{y}, \dots, (\mathbf{X}\mathbf{X}')^{i-1}\mathbf{X}\mathbf{X}'\underline{y}\}.$$

**Prueba a)** La ortogonalidad de las variables latentes se demostrará por inducción.

Para  $i = 1$ , se tiene que:

$$\begin{aligned} t_1' t_2 &\propto t_1'(\mathbf{E}_1 \underline{w}_2) = \alpha_1 \left( (t_1' \mathbf{E}_1) \underline{w}_2 \right) \\ &= \alpha_1 \left( t_1' (\mathbf{X} - t_1 \underline{p}_1) \underline{w}_2 \right) \\ &= \alpha_1 \left( t_1' (\mathbf{I} - t_1 t_1') \mathbf{X} \underline{w}_2 \right) \\ &= \alpha_1 \left( (t_1' - t_1') \mathbf{X} \underline{w}_2 \right) \\ &= 0 \end{aligned}$$

Se cumple para  $i = 1$ . Supongamos que se cumple el supuesto de ortogonalidad para  $i = 1, \dots, n$ . Por demostrar que se cumple para  $n + 1$ , entonces:

$$\begin{aligned} t_n' t_{n+1} &\propto t_n'(\mathbf{E}_n \underline{w}_{n+1}) = \alpha_n \left( (t_n' \mathbf{E}_n) \underline{w}_{n+1} \right) \\ &= \alpha_n \left( t_n' (\mathbf{X} - t_1 \underline{p}_1 - t_2 \underline{p}_2 - \dots - t_n \underline{p}_n) \underline{w}_{n+1} \right) \\ &= \alpha_n \left( t_n' (\mathbf{I} - t_1 t_1' - t_2 t_2' - \dots - t_n t_n') \mathbf{X} \underline{w}_{n+1} \right) \\ &= \alpha_n \left( (t_n' - t_n' t_1 t_1' - t_n' t_2 t_2' - \dots - t_n' t_n t_n') \mathbf{X} \underline{w}_{n+1} \right) \\ &= \alpha_n \left( (t_n' - t_n') \mathbf{X} \underline{w}_{n+1} \right) \\ &= 0 \end{aligned}$$

Ya que:

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

$$\begin{aligned}
 \underline{p}'_n &= \underline{t}'_n \mathbf{E}_{n-1} = \underline{t}'_n \left( \mathbf{X} - \underline{t}_1 \underline{p}'_1 - \dots - \underline{t}_{n-1} \underline{p}'_{n-1} \right) \\
 &= \underline{t}'_n \mathbf{X} - \left( \underline{t}'_n \underline{t}_1 \right) \underline{p}'_1 - \dots - \left( \underline{t}'_n \underline{t}_{n-1} \right) \underline{p}'_{n-1} \\
 &= \underline{t}'_n \mathbf{X}
 \end{aligned}$$

Como las variables latentes en cada iteración son normalizadas, entonces  $\{\underline{t}_1, \dots, \underline{t}_a\}$  es un conjunto ortonormal. ■

Esta propiedad trae consigo el siguiente resultado para las cargas de  $\underline{y}$ :

$$\begin{aligned}
 \underline{q}_m &= \underline{t}'_m \underline{f}_{m-1} = \underline{t}'_m \left( \underline{y} - \underline{t}_1 \underline{q}'_1 - \dots - \underline{t}_{m-1} \underline{q}'_{m-1} \right) \\
 &= \underline{t}'_m \underline{y} - \left( \underline{t}'_m \underline{t}_1 \right) \underline{q}'_1 - \dots - \left( \underline{t}'_m \underline{t}_{m-1} \right) \underline{q}'_{m-1} \\
 &= \underline{t}'_m \underline{y}
 \end{aligned}$$

Lo anterior ha dado lugar a una versión modificada de MCP1, donde la actualización del vector de residuales  $\underline{f}_h$  es omitida, acción que no daña la estructura del algoritmo.

Entonces  $\{\underline{t}_1, \underline{t}_2, \dots, \underline{t}_a\}$  forma una base ortonormal del subespacio  $a$ -dimensional del espacio columna de  $\mathbf{X}$ . Si  $a = \text{rango}(\mathbf{X})$ , entonces  $\text{gen}\{\underline{t}_1, \underline{t}_2, \dots, \underline{t}_a\}$  es el espacio columna de  $\mathbf{X}$ .

**Prueba b)** Se hará por inducción. De (4.8), la afirmación se cumple para  $i=1$ . Supongamos que se cumple para  $i=1, \dots, n$ . Por demostrar que se cumple para  $n+1$ , entonces:

$$\begin{aligned}
 \underline{t}_{n+1} &= \delta_{n+1} \left( \mathbf{E}_n \underline{w}_{n+1} \right) \\
 &= \delta_{n+1} \left( \left( \mathbf{X} - \underline{t}_1 \underline{p}'_1 - \underline{t}_2 \underline{p}'_2 - \dots - \underline{t}_n \underline{p}'_n \right) \underline{w}_{n+1} \right) \\
 &= \delta_{n+1} \left( \left( \mathbf{X} - \sum_{i=1}^n \underline{t}_i \underline{p}'_i \right) \underline{w}_{n+1} \right) \\
 &= \delta_{n+1} \left( \mathbf{X} \underline{w}_{n+1} - \sum_{i=1}^n \underline{t}_i \left( \underline{p}'_i \underline{w}_{n+1} \right) \right) \\
 &= \delta_{n+1} \mathbf{X} \underline{w}_{n+1} - \beta^* \underline{t}_n \tag{iv}
 \end{aligned}$$

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

Pues  $\underline{p}'_{i, w_{n+1}}$  es un escalar para  $i = 1, 2, \dots, n$ .

$$\therefore \underline{t}_{n+1} \in \text{gen} \left\{ \underline{X}\underline{X}'\underline{y}, \dots, (\underline{X}\underline{X}')^n \underline{X}\underline{X}'\underline{y} \right\} = \mathbf{K}_{n+1}(\underline{X}\underline{X}', \underline{X}\underline{X}'\underline{y}) \quad \blacksquare$$

**Propiedad 3.** – El vector de pesos  $\underline{w}_m$  es ortogonal a la carga  $\underline{p}'_n$  para  $m < n$

**Prueba:** Según lo definido en NIPALS MCP1, para  $m < n$ , se tiene que:

$$\underline{p}'_n = \underline{t}'_n \mathbf{E}_{n-1} \text{ y } \underline{w}_m \propto \mathbf{E}'_{m-1} \underline{f}_{m-1},$$

Entonces:

$$\begin{aligned} \underline{p}'_n \underline{w}_m &= \underline{t}'_n \mathbf{E}_{n-1} \underline{w}_m \\ &= \underline{t}'_n \left( \mathbf{E}_{n-2} - \underline{t}_{n-1} \underline{p}'_{n-1} \right) \underline{w}_m \\ &= \underline{t}'_n \left( \mathbf{E}_{n-2} - \underline{t}_{n-1} \underline{t}'_{n-1} \mathbf{E}_{n-2} \right) \underline{w}_m \\ &= \underline{t}'_n \left( \mathbf{I}_{n-2} - \underline{t}_{n-1} \underline{t}'_{n-1} \right) \mathbf{E}_{n-2} \underline{w}_m \\ &= \underline{t}'_n \left( \mathbf{I} - \underline{t}_{n-1} \underline{t}'_{n-1} \right) \left( \mathbf{I} - \underline{t}_{n-2} \underline{t}'_{n-2} \right) \cdots \left( \mathbf{I} - \underline{t}_m \underline{t}'_m \right) \mathbf{E}_{m-1} \underline{w}_m \\ &= \underline{t}'_n \underline{t}_m \\ &= 0 \end{aligned}$$

La propiedad 3, garantiza la inversión del producto  $\mathbf{P}'_a \mathbf{W}_a$  por ser una matriz triangular superior. Esta matriz, juega un papel importante en el modelo de regresión por MCP1, pues da la relación entre la matriz de datos  $\mathbf{X}$  y la matriz de vectores de pesos  $\mathbf{W}_a$ .

Durante mucho tiempo, las propiedades matemáticas de NIPALS MCP1 fueron desconocidas, pues se había propuesto como una versión modificada de NIPALS ACP, argumentando que “si se pretendía minimizar la suma de cuadrados de los residuales, proyectando la variable de respuesta en un subespacio del espacio columna de  $\mathbf{X}$ , entonces la información de  $\underline{y}$  tendría que influenciar la elección de las direcciones hacia donde las componentes principales estuvieran dirigidas”, es evidente que el argumento es heurístico. Sin embargo, los subespacios de Krilov mencionados

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

en las propiedades 1 y 2, son la clave para el entendimiento de MCP como un método matemático riguroso más que un simple algoritmo, debido a la equivalencia matemática de MCP1 con el método de Bidiagonalización<sup>6</sup> de Lanczos (LBD, por sus siglas en inglés) de análisis numérico, el cual aproxima la solución al sistema de ecuaciones  $\underline{y} = \mathbf{X}\underline{\beta}$ , al descomponer a  $\mathbf{X}$  de manera iterativa a la siguiente estructura:

$$\mathbf{X} = \mathbf{U}_a \mathbf{B}_a \mathbf{V}_a$$

Donde:  $\mathbf{U}_a$  y  $\mathbf{V}_a$  son matrices ortogonales y  $\mathbf{B}_a$  es una matriz bidiagonal superior, esto es:

$$\mathbf{B}_a = \begin{pmatrix} \alpha_1 & \beta_1 & & & & \\ & \alpha_2 & \beta_2 & & & \\ & & \ddots & \ddots & & \\ & & & \alpha_{a-1} & \beta_{a-1} & \\ & & & & & \alpha_a \end{pmatrix}$$

La columnas de  $\mathbf{V}_a$  y  $\mathbf{U}_a$  son ortonormales y generan los mismos subespacios de Krilov que  $\mathbf{W}_a$  y  $\mathbf{T}_a$ , respectivamente<sup>7</sup>. (No es el propósito de este trabajo desarrollar la prueba). Entonces:

$$\mathbf{B}_a = \mathbf{P}'_a \mathbf{W}_a$$

El método LBD, aproxima a la solución  $\underline{\hat{\beta}}$ , tal que se minimice  $\|\underline{y} - \mathbf{X}\underline{\hat{\beta}}\|^2$ . Utilizando la descomposición de  $\mathbf{X}$ ,  $\underline{y} = \mathbf{X}\underline{\beta}$  tiene como solución  $\underline{\hat{\beta}}_{\text{LBD}} = \mathbf{V}'_a \mathbf{B}_a^{-1} \mathbf{U}'_a$ . Dado que generan los mismos subespacios, en los cuales se proyecta la variable de respuesta  $\underline{y}$ , los métodos de MCP1 y LBD producen el mismo vector de coeficientes de regresión, i.e.,  $\underline{\hat{\beta}}_{\text{MCP}} = \underline{\hat{\beta}}_{\text{LBD}}$ .

Para resolver problemas de mínimos cuadrados, LBD es matemáticamente equivalente (más no numéricamente) al método del gradiente conjugado, aplicado a

<sup>6</sup> La conexión entre MCP y LBD es desarrollada en Phatak y De Hoog. (2002)

<sup>7</sup> Eldén, L. (2004)

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

las ecuaciones normales  $\mathbf{X}'\mathbf{X}\underline{\beta} = \mathbf{X}'\underline{y}$ , que iterativamente aproxima a la matriz  $(\mathbf{X}'\mathbf{X})^{-1}$  por medio de direcciones conjugadas, las cuales son especificadas por los datos.

Se puede concluir que existen algoritmos numéricos,<sup>8</sup> mayormente entendidos como métodos matemáticos rigurosos que dan los mismos resultados que NIPALS MCP.

### Propiedad 4. – Propiedades de reducción del estimador $\hat{\underline{\beta}}$ por MCP1

Esta propiedad se refiere a la calidad del estimador  $\hat{\underline{\beta}}$  por MCP1, comparado al que produce la técnica de RLM. En el capítulo 2, se obtuvo el ECM del estimador  $\hat{\underline{\beta}}_{\text{MCO}}$ , como:

$$\text{ECM} = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}$$

A partir de esto, un decremento o incremento del ECM de  $\hat{\underline{\beta}}_{\text{MCO}}$ , está directamente ligado a los valores propios de la matriz de varianzas y covarianzas de la matriz  $\mathbf{X}$ ; entonces el ECM puede ser reducido al modificar las direcciones de  $\hat{\underline{\beta}}_{\text{MCO}}$  que son responsables de la varianza grande<sup>9</sup> de  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ ; se puede reducir al estimador por mínimos cuadrados ordinarios mediante una función de reducción, como sigue:

$$\hat{\underline{\beta}}_{\text{Reducido}} = \sum_{i=1}^p f(\lambda_i) \frac{1}{\sqrt{\lambda_i}} \mathbf{v}_i \mathbf{u}_i' \mathbf{y}_i$$

A los  $f(\lambda_i)$  se les conoce como factores de reducción<sup>10</sup>. Por definición, en la RLM se tiene que  $f(\lambda_i) = 1, \forall i$ . Los factores de reducción pueden ser determinísticos si no dependen de manera lineal de la respuesta  $\underline{y}$ , o estocásticos, si lo hacen de manera lineal o no lineal. Cuando los factores de reducción son determinísticos, los valores de los factores de reducción pueden oscilar dentro de valores mayores o menores a uno,

<sup>8</sup> Gradiente Conjugado y Bidiagonalización de Lanczos.

<sup>9</sup> Ver capítulo 2

<sup>10</sup> Frank and Friedman. (1993)



## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

cuando ésto pasa, se introduce un sesgo en el proceso de estimación del parámetro, es decir, para valores mayores a uno, existirá sesgo en el ECM y la varianza de  $\hat{\underline{\beta}}_{\text{MCO}}$  será incrementada simultáneamente; para valores menores a uno, se introducirá sesgo pero habrá una reducción en la varianza del estimador, trayendo como consecuencia inmediata la reducción del ECM de  $\hat{\underline{\beta}}_{\text{MCO}}$ <sup>11</sup>. Aunque se introduce sesgo en el ECM, se espera que el incremento en éste sea más pequeño comparado con el decremento de la varianza de  $\hat{\underline{\beta}}_{\text{MCO}}$ .

Es preciso, dar la forma explícita de los factores de reducción obtenidos por MCP1, para encontrar propiedades del estimador que ofrece esta técnica. Por un lado, utilizando las primeras  $a$  variables latentes en RMCP1, se estableció en (4.19) que:

$$\hat{\underline{\beta}}_{\text{MCP1}}^a = \mathbf{W}_a (\mathbf{P}'_a \mathbf{W}_a)^{-1} \underline{\mathbf{q}}_a$$

Por la propiedad 2, se sabe que las columnas de  $\mathbf{W}_a$  generan el subespacio de Krilov  $\{\mathbf{X}'\underline{\mathbf{y}}, \dots, (\mathbf{X}'\mathbf{X})^{r-1}\mathbf{X}'\underline{\mathbf{y}}\}$ . Escribiendo a  $\hat{\underline{\beta}}_{\text{MCP1}}^a$ , como combinación lineal de los elementos de este subespacio, se tiene:

$$\hat{\underline{\beta}}_{\text{MCP1}}^a = \sum_{i=1}^a \tilde{\gamma}_i (\mathbf{X}'\mathbf{X})^{i-1} \mathbf{X}'\underline{\mathbf{y}}$$

Con  $\tilde{\gamma}_i$  tal que  $\hat{\underline{\beta}}_{\text{MCP1}}^a$  minimiza (4.14). Considerando la descomposición en valores singulares de  $\mathbf{X}$ , se cumplen las siguientes igualdades:

$$\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{D}^2\mathbf{V}' \text{ y } \mathbf{X}'\underline{\mathbf{y}} = \mathbf{V}\mathbf{D}\mathbf{U}'\underline{\mathbf{y}}$$

Entonces:

$$\begin{aligned} \hat{\underline{\beta}}_{\text{MCP1}}^a &= \tilde{\gamma}_1 (\mathbf{X}'\underline{\mathbf{y}}) + \tilde{\gamma}_2 ((\mathbf{X}'\mathbf{X})\mathbf{X}'\underline{\mathbf{y}}) + \dots + \tilde{\gamma}_a ((\mathbf{X}'\mathbf{X})^{a-1}\mathbf{X}'\underline{\mathbf{y}}) \\ &= \tilde{\gamma}_1 (\mathbf{V}\mathbf{D}\mathbf{U}'\underline{\mathbf{y}}) + \tilde{\gamma}_2 (\mathbf{V}\mathbf{D}^3\mathbf{U}'\underline{\mathbf{y}}) + \dots + \tilde{\gamma}_a (\mathbf{V}\mathbf{D}^{2a-1}\mathbf{U}'\underline{\mathbf{y}}) \\ &= \sum_{i=1}^a \tilde{\gamma}_i \mathbf{D}^{2i} (\mathbf{V}\mathbf{D}^{-1}\mathbf{U}'\underline{\mathbf{y}}) \end{aligned} \quad (\text{i})$$

<sup>11</sup> Butler and Denham. (2000) y Krämer, N. (2007)

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

Dado que,  $\mathbf{D} = \text{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_p}\right)$  con  $\sqrt{\lambda_1} > \sqrt{\lambda_2} > \dots > \sqrt{\lambda_p}$ . Entonces se cumple que  $\mathbf{D}^{2i} = \text{diag}\left(\lambda_1^i, \lambda_2^i, \dots, \lambda_p^i\right) = \Delta^i$ . Entonces, se puede escribir (ii) como:

$$\begin{aligned}\hat{\beta}_{\text{MCP1}}^a &= \sum_{i=1}^a \tilde{\gamma}_i \mathbf{D}^{2i} (\mathbf{V}\mathbf{D}^{-1}\mathbf{U}'\underline{y}) \\ &= \sum_{i=1}^a \tilde{\gamma}_i \Delta^i (\mathbf{V}\mathbf{D}^{-1}\mathbf{U}'\underline{y}) \\ &= \sum_{i=1}^a \tilde{\gamma}_i \Delta^i \hat{\beta}_{\text{MCO}} \\ &= \sum_{i=1}^a f^{(i)}(\lambda) \hat{\beta}_{\text{MCO}}\end{aligned}\quad (\text{ii})$$

Por ejemplo, si  $a=1$ , entonces  $f^{(1)}(\lambda_i) = \tilde{\gamma}_1 \lambda_i$  para  $i=1,2,\dots,p$ , con  $\tilde{\gamma}_1$  que minimice (4.14) y donde  $f^{(1)}(\lambda_i) = \tilde{\gamma}_1 \lambda_i$  pondera a la  $i$ -ésima entrada del vector  $\hat{\beta}_{\text{MCO}}$ . Se puede demostrar que los  $f^{(a)}(\lambda_i)$  que corresponden al modelo de RMCP1, están dados por<sup>12</sup>:

$$f^{(a)}(\lambda_i) = 1 - \prod_{j=1}^r \left(1 - \frac{\lambda_i}{\mu_j^{(a)}}\right) \quad (\text{iii})$$

Donde:  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$  son los valores propios de la matriz  $\mathbf{X}'\mathbf{X}$ , y  $\mu_1^{(a)} \geq \mu_2^{(a)} \geq \dots \geq \mu_r^{(a)}$  corresponden a los valores propios de la matriz  $\mathbf{W}_a'\mathbf{X}'\mathbf{X}\mathbf{W}_a$ . La expresión (iii), nos dice que los factores de reducción dependen de manera no lineal de la respuesta  $\underline{y}$ , por tanto no es clara la forma en la que función de reducción influencia el ECM del estimador  $\hat{\beta}_{\text{MCP1}}$ . El comportamiento de los  $f^{(a)}(\lambda_i)$  de MCP1, ha sido analizado por varios autores<sup>13</sup>, sobre diversos conjuntos de datos que ejemplifican la “reducción” que realiza MCP sobre el estimador  $\hat{\beta}_{\text{MCO}}$ ; sin embargo, en

<sup>12</sup> Lingjaerde, O. and Christophersen, N. (2000)

<sup>13</sup> Lingjaerde and Christophersen. (2000), Goutis, C. (1996) y Butler and Denham (2000)

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

ciertas circunstancias el comportamiento de los  $f^{(a)}(\lambda_i)$  para algunos valores de  $i$ , es irregular al tomar valores mayores a uno e incluso negativos. Frank and Friedman conjeturan que  $\hat{\beta}_{MCP1}^a$  pondera la  $a$ -ésima entrada del vector  $\hat{\beta}_{MCO}$  por un factor mayor a uno; Krämer, Nicole (2007) argumenta que este hecho no necesariamente tiene que incrementar el ECM de  $\hat{\beta}_{MCP1}$ , y da una alternativa para truncar los valores de los factores de reducción con comportamiento irregular. Los estudios realizados sobre este comportamiento muestran que, mientras más variables latentes sean agregadas al modelo de RMCP1, los valores de los factores de reducción convergen uno. Referente a esta situación, Goutis, C. (1996) prueba geoméricamente que el estimador de los coeficientes de regresión por MCP1 es reducido comparado con el de RLM, en el sentido de que MCP1 produce una reducción en la magnitud del vector  $\hat{\beta}_{MCO}$ , esto es:

$$\|\hat{\beta}_{MCP1}^1\| \leq \|\hat{\beta}_{MCP1}^2\| \leq \dots \leq \|\hat{\beta}_{MCO}\|$$

De hecho, se cumple que:

$$\|\hat{\beta}_{MCP1}^p\| = \|\hat{\beta}_{MCO}\| \quad (\text{iv})$$

Para probar (iv), considere lo siguiente: Sea  $\mathbf{X}_{n \times p}$  con rango  $(\mathbf{X}) = p$ . La matriz  $\mathbf{W}_p$  es cuadrada y ortonormal, por lo tanto:

$$\begin{aligned} \hat{\beta}_{MCP1}^a &= \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1} \underline{q}_p \\ &= \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1} (\mathbf{T}'_p \mathbf{T}_p)^{-1} \mathbf{T}'_p \underline{y} \\ &= \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1} \left( (\mathbf{X} \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1})' \mathbf{X} \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1} (\mathbf{X} \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1})' \right) \underline{y} \\ &= \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1} \left( (\mathbf{W}'_p \mathbf{P}_p)^{-1} (\mathbf{X} \mathbf{W}_p)' \mathbf{X} \mathbf{W}_p (\mathbf{P}'_p \mathbf{W}_p)^{-1} (\mathbf{W}'_p \mathbf{P}_p)^{-1} (\mathbf{X} \mathbf{W}_p)' \right) \underline{y} \\ &= \mathbf{W}_p \left( (\mathbf{X} \mathbf{W}_p)' \mathbf{X} \mathbf{W}_p \right)^{-1} (\mathbf{X} \mathbf{W}_p)' \underline{y} \\ &= \mathbf{W}_p (\mathbf{W}'_p \mathbf{X}' \mathbf{X} \mathbf{W}_p)^{-1} \mathbf{W}'_p \mathbf{X}' \underline{y} \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \underline{y} \\ &= \hat{\beta}_{MCO} \end{aligned}$$

Este resultado se verifica, pues el espacio generado por el número máximo de variables latentes que NIPALS MCP1 puede calcular es igual al espacio columna de la

## 4.2 Mínimos Cuadrados Parciales Univariado (MCP1)

matriz  $\mathbf{X}$  (Ver la propiedad 1). Por lo tanto, el estimador del vector de coeficientes de regresión por MCP1 es equivalente al estimador del vector de coeficientes de regresión de RLM, cuando esta última es aplicable. Ahora, en (4.15) se expusieron los valores ajustados  $\hat{\underline{y}}_{\text{MCP1}}^a = \mathbf{T}_a \underline{q}_a$ , por construcción sigue de manera inmediata que<sup>14</sup>:

$$\left\| \hat{\underline{y}}_{\text{MCP1}}^a \right\| \leq \left\| \hat{\underline{y}}_{\text{MCO}} \right\|$$

## 4.3 Mínimos Cuadrados Parciales Multivariado, MCP2

Es de interés científico desarrollar técnicas multivariadas para tener entendimiento de relaciones funcionales entre un conjunto de variables independientes y otro de dependientes; estamos situados en una generalización de MCP1 donde en  $n$  unidades observacionales, se tienen datos de  $p$  variables independientes y  $q > 1$  variables de respuesta. El modelo para el caso multivariado es:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{\Xi} \quad (4.31)$$

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1q} \\ y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nq} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nq} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1q} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nq} \end{pmatrix}$$

Donde:

$$\begin{aligned} \underline{y}_1 &= \mathbf{X}\underline{\beta}_1 + \varepsilon_1 \\ \underline{y}_2 &= \mathbf{X}\underline{\beta}_2 + \varepsilon_2 \\ &\vdots \\ \underline{y}_q &= \mathbf{X}\underline{\beta}_q + \varepsilon_q \end{aligned}$$

Como se presentó en el capítulo 1, originado de la dependencia lineal entre  $\mathbf{X}$  y  $\mathbf{Y}$ , cada  $y_{ij}$  sufre cierta cantidad de variación por cada unidad de variación de  $x_j$ , representada por  $\beta_{ij}$ . Para ajustar el modelo de regresión multivariado, se pide rango completo por columnas a  $\mathbf{Y}$  (independencia entre las variables de respuesta), de esta

<sup>14</sup> Denham (1991) pag. 74, Goutis, C. (1996)

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

manera, se puede usar MCP1 individualmente para cada columna  $\underline{y}_j$ , construyendo  $q > 1$  modelos predictivos óptimos por MCP; este razonamiento es sinónimo de encontrar  $q > 1$  conjuntos de variables latentes  $\left\{ \left\{ t_{i_1} \right\}_{i=1}^{a_1^*}, \left\{ t_{i_2} \right\}_{i=1}^{a_2^*}, \dots, \left\{ t_{i_q} \right\}_{i=1}^{a_q^*} \right\}$  diferentes entre sí, puesto que para cada  $\underline{y}_j$  ni los conjuntos de variables latentes son necesariamente los mismos ni el número óptimo  $a^*$  de variables latentes a retener para ajustar el modelo de regresión final, lo debe de ser. La contraparte a este procedimiento conjetura que, es ventajoso incluir información de las columnas restantes en la predicción de una  $\underline{y}_j$  en específico, por ser las variables independientes parte de la información; además en la práctica, aparentemente se gana precisión al reemplazar predicciones univariadas por un modelo de predicción multivariado<sup>15</sup>, aunado a esto, se pretende reducir los cálculos al no construir  $q > 1$  modelos de regresión por MCP1, explicando cada columna  $\underline{y}_j$ , con  $j=1,2,\dots,q$ , mediante un modelo general que describa a  $\mathbf{Y}$  como función de  $\mathbf{X}$ , para lograr todo ésto, el método MCP2 fue creado; las columnas de  $\mathbf{Y}$  son predichas simultáneamente y la información de los distintos conjuntos de variables latentes, es incluida dentro de uno solo, lo cual implica un algoritmo más complejo y de mayores pasos, comparado con MCP1.

El principio detrás del método de MCP2, puede resumirse de la siguiente manera: la descomposición bilineal de  $\mathbf{X}$  y de  $\mathbf{Y}$ , se efectuará por medio de dos conjuntos ortogonales de variables latentes; MCP2 construye un conjunto adicional de variables latentes para  $\mathbf{Y}$ , a saber:

$$\begin{aligned} \mathbf{X} &= \underline{t}_1 \underline{p}'_1 + \underline{t}_2 \underline{p}'_2 + \dots + \underline{t}_a \underline{p}'_a + \mathbf{E}_a \\ \mathbf{X} &= \mathbf{T}_a \mathbf{P}'_a + \mathbf{E}_a \end{aligned} \quad (4.32)$$

$$\begin{aligned} \mathbf{Y} &= \underline{u}_1 \underline{q}'_1 + \underline{u}_2 \underline{q}'_2 + \dots + \underline{u}_a \underline{q}'_a + \mathbf{F}_a \\ \mathbf{Y} &= \mathbf{U}_a \mathbf{Q}'_a + \mathbf{F}_a \end{aligned} \quad (4.33)$$

---

<sup>15</sup> Sundberg, R. (2002)

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

$$\mathbf{U}_a = \mathbf{T}_a \mathbf{C}_a + \mathbf{\Omega}_a \quad (4.34)$$

Donde  $\mathbf{\Omega}_a$  es una matriz de residuales; el modelo interno es un modelo estadístico que modela la relación lineal entre las variables latentes de  $\mathbf{X}$  y las variables latentes de  $\mathbf{Y}$ , y es el resultado de regresar a  $\underline{u}_i$  sobre  $\underline{t}_i$  de  $i=1,2,\dots,a$  de forma iterativa, por lo que los valores ajustados para  $\hat{\underline{u}}_1, \hat{\underline{u}}_2, \dots, \hat{\underline{u}}_a$  están representados como:

$$\hat{\mathbf{U}}_a = \mathbf{T}_a \hat{\mathbf{C}}_a \quad (4.35)$$

Con  $\hat{\mathbf{C}}_a = \text{diag}(\hat{c}_1, \hat{c}_2, \dots, \hat{c}_a)$ . Sustituyendo (4.35) en (4.33) se tiene:

$$\begin{aligned} \hat{\mathbf{Y}} &= \hat{\mathbf{U}}_a \mathbf{Q}_a' \\ \hat{\mathbf{Y}} &= \mathbf{T}_a \hat{\mathbf{C}}_a \mathbf{Q}_a' \end{aligned} \quad (4.36)$$

$$\hat{\mathbf{Y}} = \mathbf{T}_a \hat{\boldsymbol{\beta}}_{\text{MCP2}}^a \quad (4.37)$$

Con  $\hat{\boldsymbol{\beta}}_{\text{MCP2}} = \hat{\mathbf{C}}_a \mathbf{Q}_a'$ .

#### 4.3.1 Planteamiento matemático

Sean  $\underline{x} = (x_1 \ x_2 \ \dots \ x_p)'$  y  $\underline{y} = (y_1 \ y_2 \ \dots \ y_q)'$  vectores aleatorios y  $\underline{w}_1 = (w_{11} \ w_{21} \ \dots \ w_{p1})'$  y  $\underline{v}_1 = (v_{11} \ v_{21} \ \dots \ v_{q1})'$  vectores de pesos. Se define a la primera variable latente de  $\underline{y}$  como  $u_1 = \underline{v}_1' \underline{y}$ , tal que el criterio de maximización de la covarianza se cumple, es decir,  $\text{Cov}(t_1, u_1)$  es máxima. Para la primera iteración, se plantean las variables latentes  $t_1$  y  $u_1$  como sigue:

$$t_1 = \underline{w}_1' \underline{x} \quad \text{y} \quad u_1 = \underline{v}_1' \underline{y}$$

Para estos dos arreglos, MCP2 busca  $\underline{w}_1$  y  $\underline{v}_1$  tales que:

## Capítulo 4. Mínimos Cuadrados Parciales

$$\text{Max}(\text{Cov}(\underline{w}_1' \underline{x}, \underline{v}_1' \underline{y})) = \text{Max}(\underline{w}_1' \text{Cov}(\underline{x}, \underline{y}) \underline{v}_1)$$

$$\text{Sujeto a: } \|\underline{w}_1\| = 1 \text{ y } \|\underline{v}_1\| = 1$$

$$\text{Sea } \Sigma_{XY} = \text{Cov}(\underline{x}, \underline{y}) = \begin{pmatrix} \text{Cov}(x_1, y_1) & \text{Cov}(x_1, y_2) & \cdots & \text{Cov}(x_1, y_q) \\ \text{Cov}(x_2, y_1) & \text{Cov}(x_2, y_2) & \cdots & \text{Cov}(x_2, y_q) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(x_p, y_1) & \text{Cov}(x_p, y_2) & \cdots & \text{Cov}(x_p, y_q) \end{pmatrix} \quad (4.38)$$

La matriz de covarianzas entre  $\underline{x}$  y  $\underline{y}$ . Sea  $\phi$  la función a maximizar, que contiene las restricciones anteriores, utilizando los multiplicadores de Lagrange  $\delta$  y  $\gamma$ , tenemos:

$$\phi = \underline{w}_1' \Sigma_{XY} \underline{v}_1 - \delta (\underline{w}_1' \underline{w}_1 - 1) - \gamma (\underline{v}_1' \underline{v}_1 - 1)$$

Obteniendo las derivadas parciales con respecto a  $\underline{w}_1$  y  $\underline{v}_1$ :

$$\frac{\partial \phi}{\partial \underline{w}_1} = \Sigma_{XY} \underline{v}_1 - \delta \underline{w}_1 ; \quad \frac{\partial \phi}{\partial \underline{v}_1} = \Sigma_{XY}' \underline{w}_1 - \gamma \underline{v}_1$$

Igualando a cero, se forma el sistema de ecuaciones siguiente:

$$\begin{cases} \Sigma_{XY} \underline{v}_1 = \delta \underline{w}_1 & (1) \\ \Sigma_{XY}' \underline{w}_1 = \gamma \underline{v}_1 & (2) \end{cases}$$

En el sistema anterior, se tiene que  $\delta = \gamma$ . Despejando  $\underline{w}_1$  de (1) y sustituyendo este valor en (2), se obtiene:

$$\delta^{-1} \Sigma_{XY} \underline{v}_1 = \underline{w}_1 \Rightarrow \Sigma_{XY}' \Sigma_{XY} \underline{v}_1 = \lambda \underline{v}_1 \quad (4.39)$$

Despejando  $\underline{v}_1$  de (2) y sustituyendo este valor en (1)

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

$$\gamma^{-1} \Sigma'_{XY} \underline{w}_1 = \underline{v}_1 \Rightarrow \Sigma_{XY} \Sigma'_{XY} \underline{w}_1 = \lambda \underline{w}_1 \quad (4.40)$$

Con el escalar  $\lambda = \delta^2$ . De esta manera, los vectores de pesos  $\underline{w}_1$  y  $\underline{v}_1$  que maximizan la covarianza cuadrada entre  $t_1$  y  $u_1$ , son los vectores propios asociados a los valores propios más grandes de las matrices  $\Sigma_{XY} \Sigma'_{XY}$  y  $\Sigma'_{XY} \Sigma_{XY}$  respectivamente, para lograr maximizar la función  $\phi$ . Reemplazando a  $\Sigma_{XY}$  por su estimador insesgado, la matriz de covarianza muestral de las matrices de datos  $\mathbf{X}$  y  $\mathbf{Y}$ , dada por  $\hat{\Sigma}_{XY} = S_{XY} = \frac{1}{n-1} \mathbf{X}'\mathbf{Y}$ , en (4.39) y (4.40), se tiene:

$$\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y} \underline{v}_1 = \lambda_1 \underline{v}_1 \quad (4.41)$$

$$\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X} \underline{w}_1 = \lambda_1 \underline{w}_1 \quad (4.42)$$

Dados estos resultados, los primeros vectores de pesos  $\underline{w}_1$  y  $\underline{v}_1$  son el vector singular izquierdo y derecho respectivamente, asociados al primer valor singular de la matriz  $\mathbf{X}'\mathbf{Y}$ . Esto da un claro indicio de que MCP2, aplica la descomposición en valores singulares a la matriz  $\mathbf{S}_{XY}$ , esto es:

$$\mathbf{X}'\mathbf{Y} = \mathbf{U}\Delta\mathbf{V}' \quad (4.43)$$

Donde:

$\mathbf{U}$  := Matriz que contiene como columnas los vectores singulares izquierdos de  $\mathbf{X}'\mathbf{Y}$ , esto es, los vectores propios de la matriz  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$

$\mathbf{V}$  := Matriz que contiene como columnas los vectores singulares derechos de  $\mathbf{X}'\mathbf{Y}$ , esto es, los vectores propios de la matriz  $\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}$

$\Delta := \text{diag}\left((n-1)^2 \sqrt{\lambda_1}, (n-1)^2 \sqrt{\lambda_2}, \dots, (n-1)^2 \sqrt{\lambda_z}\right)$  tal que  $\lambda_1 > \lambda_2 > \dots > \lambda_z$

Con  $z = \min(\text{rango}(\mathbf{X}), \text{rango}(\mathbf{Y}))$



## Capítulo 4. Mínimos Cuadrados Parciales

Al igual que en MCP1, para simplificar los cálculos se normalizan las variables latentes de  $\mathbf{X}$  y de  $\mathbf{Y}$  como sigue:

$$\underline{t}_1 \leftarrow \frac{\underline{t}_1}{\|\underline{t}_1\|}, \quad \underline{u}_1 \leftarrow \frac{\underline{u}_1}{\|\underline{u}_1\|} \quad (4.44)$$

Una vez obtenidas las variables latentes  $\underline{t}_1$  y  $\underline{u}_1$ , las descomposiciones bilineales de  $\mathbf{X}$  y de  $\mathbf{Y}$ , se efectúan de manera simultánea<sup>16</sup>, es decir:

$$\mathbf{X} = \underline{t}_1 \underline{p}'_1 + \mathbf{E}_1 \quad (a)$$

$$\mathbf{Y} = \underline{u}_1 \underline{q}'_1 + \mathbf{F}_1 \quad (b)$$

Con  $\underline{p}'_1 = \underline{t}'_1 \mathbf{X}$  y  $\underline{q}'_1 = \underline{u}'_1 \mathbf{Y}$ . La relación lineal “interna” entre las variables latentes  $\underline{t}_1$  y  $\underline{u}_1$  de acuerdo a lo que establece (4.34) es:  $\underline{u}_1 = \underline{t}_1 \hat{c}_1 + \omega_1$  con el escalar  $\hat{c}_1 = \underline{t}'_1 \underline{u}_1$ , pues se busca que  $\|\underline{u}_1 - \hat{u}_1\|^2$  sea mínima, entonces:

$$\hat{u}_1 = \underline{t}_1 \hat{c}_1 \quad (c)$$

Sustituyendo lo anterior en (b), se tiene  $\hat{\mathbf{Y}}_{MCP2}^1 = \underline{t}_1 \hat{c}_1 \underline{q}'_1$ . El producto  $\hat{c}_1 \underline{q}'_1$  es interpretado como la parte de las variables de respuesta que ha sido predicha por la primera variable latente de MCP1.

El planteamiento iterativo para la construcción de las variables latentes de NIPALS MCP2, sigue la posible no ortonormalidad del conjunto  $\{\underline{t}_1, \dots, \underline{t}_a\}$ , si se eligiera a los vectores de pesos de  $\mathbf{X}$  como los vectores singulares izquierdos de la matriz  $\mathbf{X}'\mathbf{Y}$ , es decir,  $\underline{w}_i = \underline{v}_i$  asociados a  $\lambda_1 > \lambda_2 > \dots > \lambda_a$ , ya que multiplicando por la izquierda por  $\mathbf{X}$  a  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\underline{w}_i = \lambda_i^2 \underline{w}_i$ , se obtiene  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\underline{w}_i = \lambda_i \mathbf{X}\underline{w}_i$  entonces  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\underline{t}_i = \lambda_i \mathbf{X}\underline{t}_i$  para  $i = 1, 2, \dots, a$ ; las variables latentes  $\underline{t}_i$  no tienen por qué ser mutuamente ortogonales, a menos que  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'$  sea una matriz simétrica y por lo

<sup>16</sup> La ausencia del “sombrero” sobre las cargas  $\underline{p}$  y  $\underline{q}$ , se debe a que los modelos (4.32) y (4.33) no comprenden modelos estadísticos, por tanto no hay parámetros a estimar.

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

tanto, diagonalizable ortonormal, condición que no necesariamente se cumple; el mismo planteamiento sigue para las variables latentes de  $\mathbf{Y}$ .

Las segundas variables latentes, por ejemplo  $t_2$  y  $u_2$ , son construidas con la finalidad de recuperar variabilidad sistemática en  $\mathbf{X}$  y en  $\mathbf{Y}$ , contenida en las matrices de residuales  $\mathbf{E}_1$  y  $\mathbf{F}_1$ , no capturada por  $t_1$  y  $u_1$  en los modelos (a) y (b). Sintetizando, las variables latentes de  $\mathbf{X}$  y de  $\mathbf{Y}$  son construidas tales que la relación entre parejas sucesivas de ellas sea fuerte; entonces, su construcción seguirá el supuesto de maximización de covarianza entre las combinaciones lineales de las matrices de residuales de  $\mathbf{X}$  y de  $\mathbf{Y}$  de pasos anteriores, eligiendo a los vectores de pesos  $\underline{w}_i$  y  $\underline{v}_i$ , tal que el supuesto se cumpla.

En general,

$$\text{Max}\left(\text{cov}\left(\underline{w}_i' \mathbf{E}_{i-1}, \underline{v}_i' \mathbf{F}_{i-1}\right)\right) = \text{Max}\left(\underline{w}_i' \text{cov}\left(\mathbf{E}_{i-1}, \mathbf{F}_{i-1}\right) \underline{v}_i\right)$$

Sujeto a:  $\|\underline{w}_i\| = 1$  y  $\|\underline{v}_i\| = 1$

Realizando el mismo desarrollo matemático que para  $t_1$  y  $u_1$ , por (4.41) y (4.42) los vectores de pesos  $\underline{w}_i$  y  $\underline{v}_i$  resultan ser los vectores singulares izquierdo y derecho del producto cruzado de  $\mathbf{E}_{i-1}$  y  $\mathbf{F}_{i-1}$ , asociados al primer valor singular  $\tau_1^i$ , respectivamente. Desde un punto de vista práctico, el método MCP puede comprenderse como una técnica que rompe un problema de regresión lineal múltiple dentro de una serie de problemas de regresión lineal simple<sup>17</sup>.

Los pasos de MCP2 son resumidos en la siguiente tabla.

---

<sup>17</sup> Lakshminarayanan, S.; Shah, S. and Nandakumar, K. (1997)

**NIPALS MCP2**

Sean  $\mathbf{E}_0 = \mathbf{X}$  y  $\mathbf{F}_0 = \mathbf{Y}$  estandarizadas. Sea  $i = 0$ ,

1. – a) Asignar a  $\underline{w}_{i+1}$  el vector singular izquierdo y a  $\underline{v}_{i+1}$  el vector singular derecho del producto  $\mathbf{E}'_{i-1} \mathbf{F}_{i-1}$ , asociados al valor singular más grande  $\tau_1^{i+1}$ , respectivamente.

$$\text{b) Hacer } \underline{w}_{i+1} \leftarrow \frac{\underline{w}_{i+1}}{\|\underline{w}_{i+1}\|} \text{ y } \underline{v}_{i+1} \leftarrow \frac{\underline{v}_{i+1}}{\|\underline{v}_{i+1}\|}$$

$$2. - \text{a) } \underline{t}_{i+1} = \mathbf{E}_i \underline{w}_{i+1}, \underline{u}_{i+1} = \mathbf{F}_i \underline{v}_{i+1}$$

$$\text{b) Hacer } \underline{t}_{i+1} \leftarrow \frac{\underline{t}_{i+1}}{\|\underline{t}_{i+1}\|} \text{ y } \underline{u}_{i+1} \leftarrow \frac{\underline{u}_{i+1}}{\|\underline{u}_{i+1}\|}$$

$$3. - \text{a) } \underline{p}'_{i+1} = \underline{t}'_{i+1} \mathbf{E}_i$$

$$\text{b) } \underline{q}'_{i+1} = \underline{u}'_{i+1} \mathbf{F}_i$$

$$\text{c) } \underline{\hat{c}}_{i+1} = \underline{t}'_{i+1} \underline{u}_{i+1}$$

$$4. - \mathbf{E}_{i+1} = \mathbf{E}_i - \underline{t}_{i+1} \underline{p}'_{i+1} \text{ y } \mathbf{F}_{i+1} = \mathbf{F}_i - \underline{t}_{i+1} \underline{\hat{c}}_{i+1} \underline{q}'_{i+1}$$

5. – Parar si  $i = \min(\text{rango}(\mathbf{X}), \text{rango}(\mathbf{Y}))$ , en otro caso hacer  $i = i + 1$  e ir en paso 1

Figura 4.2<sup>18</sup>: Algoritmo NIPALS MCP2

El número máximo de iteraciones que el algoritmo NIPALS MCP2 puede realizar es:

<sup>18</sup> Wegelin, A. J. (2000)

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

$$a = \min(\text{rango}(\mathbf{X}), \text{rango}(\mathbf{Y}))$$

No se puede continuar más allá de  $a$ , por ser el número para el cual alguna de las matrices de residuales  $\mathbf{E}_a$  y  $\mathbf{F}_a$  sería nula. Típicamente, la sucesión  $\{\tau_1^1, \tau_1^2, \dots, \tau_1^a\}$  converge rápidamente a cero<sup>19</sup>; el Algoritmo MCP2 puede ser detenido en cuanto los valores de los  $\tau_1^i$ 's decrezcan significativamente; siendo ésto una clara indicación de que la dimensión del espacio generado por las primeras variables latentes será óptima para el ajuste a  $\mathbf{Y}$ , en un modelo de regresión, como se detallará en la sección 4.4. Finalmente, se almacenan las variables obtenidas en cada iteración en las siguientes matrices:

$\mathbf{T} = (\underline{t}_1, \underline{t}_2, \dots, \underline{t}_a) :=$  Matriz de variables latentes de  $\mathbf{X}$

$\mathbf{P} = (\underline{p}_1, \underline{p}_2, \dots, \underline{p}_a) :=$  Matriz de cargas de  $\mathbf{X}$

$\mathbf{U} = (\underline{u}_1, \underline{u}_2, \dots, \underline{u}_a) :=$  Matriz de variables latentes de  $\mathbf{Y}$

$\mathbf{Q} = (\underline{q}_1, \underline{q}_2, \dots, \underline{q}_a) :=$  Matriz de cargas de  $\mathbf{Y}$

$\mathbf{C} = \text{diag}(\hat{c}_1, \hat{c}_2, \dots, \hat{c}_a) :=$  Matriz diagonal de coeficientes de regresión del modelo interno correspondientes a la  $i$ -ésima iteración.

$\mathbf{W} = (\underline{w}_1, \underline{w}_2, \dots, \underline{w}_a) :=$  Matriz de vectores de pesos de  $\mathbf{X}$

$\mathbf{V} = (\underline{v}_1, \underline{v}_2, \dots, \underline{v}_a) :=$  Matriz de vectores de pesos de  $\mathbf{Y}$ .

La siguiente figura representa la forma esquemática del algoritmo MCP2, donde todas las matrices anteriores intervienen.

---

<sup>19</sup> Höskuldsson, A. (2004)

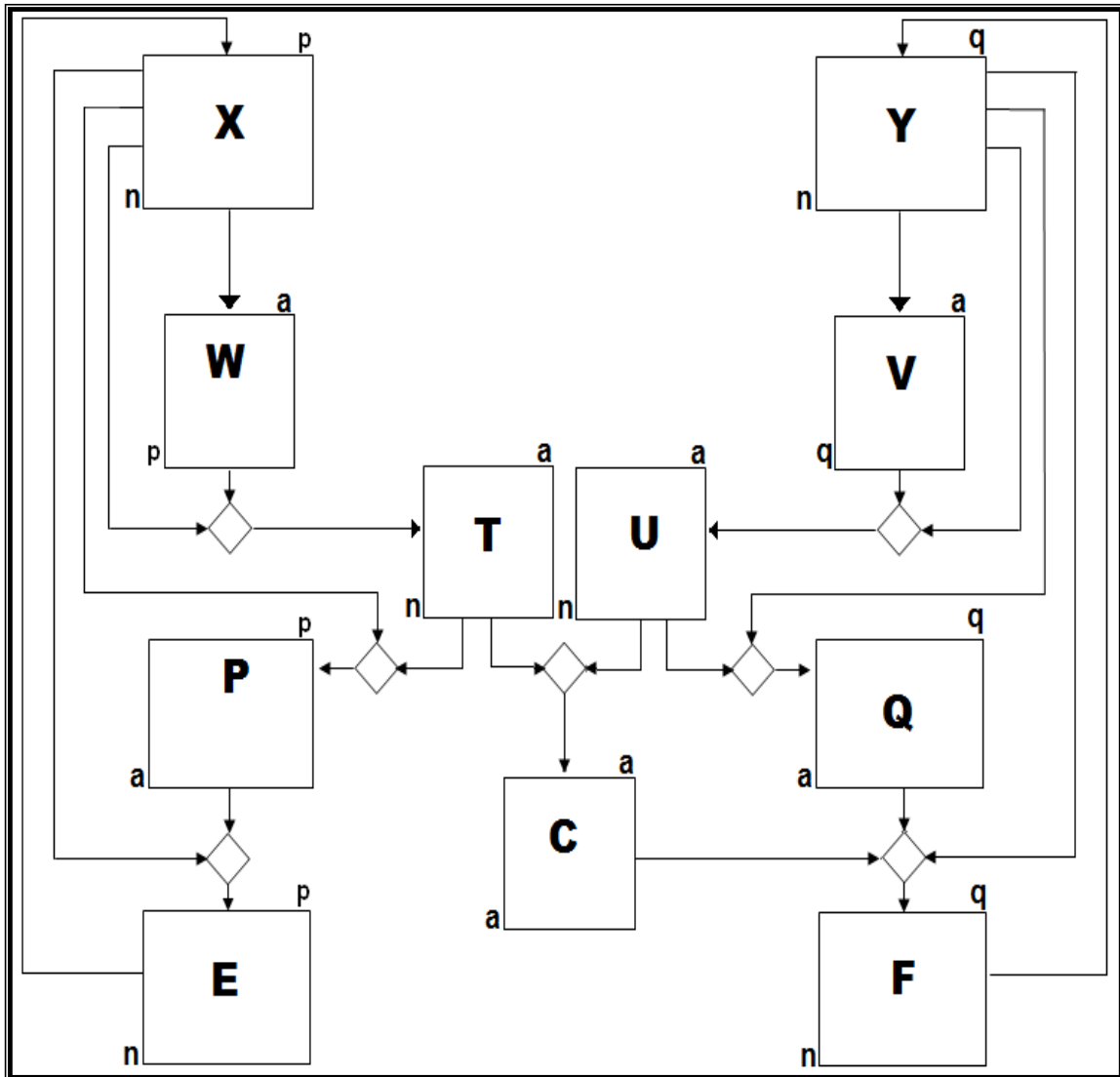


Figura 4.3<sup>20</sup>: En el diagrama de flujo del algoritmo MCP2, se representan las matrices de las variables latentes, cargas, vectores de pesos y matrices de residuales de  $X$  y de  $Y$ .

### 4.3.2 Propiedades básicas de MCP2

En esta sección se detallan las propiedades del algoritmo MCP2, suponiendo lo siguiente:  $X_{n \times p}$ ,  $Y_{n \times q}$  con  $p > q$ ,  $\text{rango}(X) = r$  y  $\text{rango}(Y) = h$ .

En (4.32) y (4.33), se plantearon los modelos externos que expresan a las matrices  $X$  y  $Y$ , como suma de aproximaciones de rango uno más un residual. Se denotan las aproximaciones de rango uno que se realizan en cada iteración como:

<sup>20</sup> Fuente: Elaboración propia.

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

$$\mathbf{X}^{(1)} = \underline{t}_1 \mathbf{p}'_1, \mathbf{X}^{(2)} = \underline{t}_2 \mathbf{p}'_2, \dots, \mathbf{X}^{(a)} = \underline{t}_a \mathbf{p}'_a$$

$$\mathbf{Y}^{(1)} = \underline{u}_1 \mathbf{q}'_1, \mathbf{Y}^{(2)} = \underline{u}_2 \mathbf{q}'_2, \dots, \mathbf{Y}^{(a)} = \underline{u}_a \mathbf{q}'_a$$

**Propiedad 1.**—a) Las variables latentes  $\{\underline{t}_1, \dots, \underline{t}_a\}$  son mutuamente ortogonales, y forman una base ortonormal para un subespacio  $a$ -dimensional del espacio columna de  $\mathbf{X}$ . b) Las variables latentes  $\{\underline{u}_1, \dots, \underline{u}_a\}$  son mutuamente ortogonales y forman una base ortonormal para un subespacio  $a$ -dimensional del espacio columna de  $\mathbf{Y}$ .

**Prueba a)** Debido a que las variables latentes son normalizadas en cada iteración, se tiene que  $\underline{t}'_n \underline{t}_n = 1$ , para  $n=1, \dots, a$ . Ahora, es necesario probar que el producto  $\underline{t}'_m \mathbf{E}_{n-1} = \underline{0}$  para  $m < n$ ; de manera recursiva se puede escribir el  $n-1$  residual como:

$$\mathbf{E}_{n-1} = \mathbf{E}_{n-2} - \underline{t}_{n-1} \mathbf{p}'_{n-1} = \mathbf{E}_{n-2} - \underline{t}_{n-1} \underline{t}'_{n-1} \mathbf{E}_{n-2}$$

Para  $n=1, \dots, a$ . Por otro lado, sustituyendo  $\underline{t}_{n-1} = \mathbf{E}_{n-2} \underline{w}_{n-1}$  se cumple:

$$\mathbf{E}_{n-1} = \mathbf{E}_{n-2} - (\mathbf{E}_{n-2} \underline{w}_{n-1}) \underline{t}'_{n-1} \mathbf{E}_{n-2} = \mathbf{E}_{n-2} \left( \mathbf{I} - \underline{w}_{n-1} \underline{t}'_{n-1} \mathbf{E}_{n-2} \right)$$

Sea  $\Theta_n = \mathbf{I} - \underline{w}_{n-1} \underline{t}'_{n-1} \mathbf{E}_{n-2}$ , con  $\mathbf{I}$  la matriz identidad de orden  $p$ ; entonces:

$$\begin{aligned} \mathbf{E}_{n-1} &= \mathbf{E}_{n-2} \Theta_{n-1} \\ &= \mathbf{E}_{n-3} \Theta_{n-2} \Theta_{n-1} \\ &= \mathbf{E}_m \Theta_{m+1} \dots \Theta_{n-1} \\ &= \mathbf{E}_m \mathbf{Z} \end{aligned}$$

Para  $\mathbf{Z} = \Theta_{m+1} \dots \Theta_n$ . Multiplicando esta igualdad por la izquierda por  $\underline{t}'_m$

$$\underline{t}'_m \mathbf{E}_{n-1} = \underline{t}'_m \mathbf{E}_m \mathbf{Z}$$

Por otro lado, se tiene:

## Capítulo 4. Mínimos Cuadrados Parciales

$$\mathbf{E}_m = \mathbf{E}_{m-1} - \underline{t}_m \underline{p}_m' = \mathbf{E}_{m-1} - \underline{t}_m \underline{t}_m' \mathbf{E}_{m-1} = (\mathbf{I} - \underline{t}_m \underline{t}_m') \mathbf{E}_{m-1} = \mathbf{A}_m \mathbf{E}_{m-1}$$

Con  $\mathbf{I}$  la matriz identidad de orden  $n$ . Entonces:

$$\begin{aligned} \underline{t}_m' \mathbf{E}_{n-1} &= \underline{t}_m' \mathbf{E}_m \mathbf{Z} \\ &= \underline{t}_m' \mathbf{A}_m \mathbf{E}_{m-1} \mathbf{Z} \\ &= \underline{t}_m' (\mathbf{I} - \underline{t}_m \underline{t}_m') \mathbf{E}_{m-1} \mathbf{Z} \\ &= [\underline{t}_m' - \underline{t}_m'] \mathbf{E}_{m-1} \mathbf{Z} \\ &= 0 \end{aligned} \tag{1.a}$$

Se concluye que:

$$\underline{t}_m' \underline{t}_n = \underline{t}_m' \mathbf{E}_{n-1} \underline{w}_n = 0 \text{ para } m < n$$

Por construcción, cada variable latente  $\underline{t}_m$  está en el espacio columna de  $\mathbf{X}$ ; así  $\{\underline{t}_1, \dots, \underline{t}_a\}$  forma una base ortonormal para un subespacio  $a$ -dimensional del espacio columna de  $\mathbf{X}$ .

**Prueba b)** El mismo planteamiento se considera para el modelo externo para  $\mathbf{Y}$ , esto es:

$$\mathbf{F}_n = \mathbf{F}_{n-1} - \underline{u}_n \underline{p}_n' = \mathbf{F}_{n-1} - \underline{u}_n \underline{u}_n' \mathbf{F}_{n-1} = \mathbf{F}_{n-1} - \mathbf{F}_{n-1}' \underline{v}_n \underline{u}_n' \mathbf{F}_{n-1} = \mathbf{F}_{n-1} \mathbf{O}_n$$

Con  $\mathbf{O}_n = (\mathbf{I} - \underline{v}_n \underline{u}_n' \mathbf{F}_{n-1})$ , siguiendo el mismo desarrollo se concluye.

Por construcción, cada variable latente  $\underline{u}_m$  está en el espacio columna de  $\mathbf{Y}$ , así  $\{\underline{u}_1, \dots, \underline{u}_a\}$  forma una base ortonormal para un subespacio  $a$ -dimensional del espacio columna de  $\mathbf{Y}$ .

A consecuencia de esta propiedad, de derivan las siguientes afirmaciones:

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

- El conjunto  $\{\underline{\hat{u}}_1, \dots, \underline{\hat{u}}_a\}$  es mutuamente ortogonal, puesto que para  $m < n$  se cumple:

$$\underline{\hat{u}}'_m \underline{\hat{u}}_n = \left( \underline{t}_m \hat{\underline{c}}_m \underline{q}'_m \right)' \left( \underline{t}_n \hat{\underline{c}}_n \underline{q}'_n \right) = \underline{q}_m \hat{\underline{c}}_m \left( \underline{t}'_m \underline{t}_n \right) \hat{\underline{c}}_n \underline{q}'_n = \underline{0}$$

- Las cargas de  $\mathbf{X}$  se pueden expresar como:

$$\begin{aligned} \underline{p}'_m &= \underline{t}'_m \mathbf{E}_{m-1} \\ &= \underline{t}'_m \mathbf{E}_{m-1} \\ &= \underline{t}'_m \left( \mathbf{X} - \underline{t}_1 \underline{p}'_1 - \dots - \underline{t}_{m-1} \underline{p}'_{m-1} \right) \\ &= \underline{t}'_m \mathbf{X} - \left( \underline{t}'_m \underline{t}_1 \right) \underline{p}'_1 - \dots - \left( \underline{t}'_m \underline{t}_{m-1} \right) \underline{p}'_{m-1} \\ &= \underline{t}'_m \mathbf{X} \end{aligned}$$

- Las cargas de  $\mathbf{Y}$  se pueden expresar como:

$$\begin{aligned} \underline{q}'_m &= \underline{u}'_m \mathbf{F}_{m-1} \\ &= \underline{u}'_m \left( \mathbf{Y} - \underline{u}_1 \underline{q}'_1 - \dots - \underline{u}_{m-1} \underline{q}'_{m-1} \right) \\ &= \underline{u}'_m \mathbf{Y} - \left( \underline{u}'_m \underline{u}_1 \right) \underline{q}'_1 - \dots - \left( \underline{u}'_m \underline{u}_{m-1} \right) \underline{q}'_{m-1} \\ &= \underline{u}'_m \mathbf{Y} \end{aligned}$$

**Propiedad 2.** – Las aproximaciones de rango uno para  $\mathbf{X}$  y para  $\mathbf{Y}$ , son mutuamente ortogonales, i.e:

$$m \neq n \Rightarrow \begin{cases} \left( \mathbf{X}^{(m)} \right)' \left( \mathbf{X}^{(n)} \right) = \underline{p}_m \underline{t}'_m \underline{t}_n \underline{p}'_n = \underline{0} \\ \left( \mathbf{Y}^{(m)} \right)' \left( \mathbf{Y}^{(n)} \right) = \underline{q}_m \underline{u}'_m \underline{u}_n \underline{q}'_n = \underline{0} \end{cases}$$

**Prueba:** Se sigue de la ortogonalidad de los conjuntos de variables latentes de  $\mathbf{X}$  y de  $\mathbf{Y}$ , de la propiedad 1.



## Capítulo 4. Mínimos Cuadrados Parciales

**Propiedad 3.** –Las aproximaciones de rango uno de  $\mathbf{X}$  y de  $\mathbf{Y}$ , son ortogonales a las matrices de residuales  $\mathbf{E}_a$  y  $\mathbf{F}_a$ , respectivamente.

$$m \leq n \Rightarrow \begin{cases} (\mathbf{X}^{(m)})' \mathbf{E}_n = \underline{0} \\ (\mathbf{Y}^{(m)})' \mathbf{F}_n = \underline{0} \end{cases}$$

**Prueba:**

$$\begin{aligned} (\mathbf{X}^{(m)})' \mathbf{E}_n &= (\mathbf{t}_m \mathbf{p}_m')' \mathbf{E}_n \\ &= \mathbf{p}_m' \mathbf{t}_m \mathbf{E}_n \quad \text{por (1.a)} \\ &= \underline{0} \end{aligned}$$

Para  $\mathbf{Y}$ , se tiene que:

$$\begin{aligned} (\mathbf{Y}^{(m)})' \mathbf{F}_n &= (\mathbf{u}_m \mathbf{q}_m')' \mathbf{F}_n \\ &= \mathbf{q}_m' \mathbf{u}_m \mathbf{F}_n \\ &= \underline{0} \end{aligned}$$

**Propiedad 4.** –Debido a la actualización de las matrices de residuales, se cumple que:

$$\text{rango}(\mathbf{E}_{m+1}) \leq \text{rango}(\mathbf{E}_m) - 1$$

$$\text{rango}(\mathbf{F}_{m+1}) \leq \text{rango}(\mathbf{F}_m) - 1$$

NIPALS MCP2, no puede continuar más allá de  $\min(r, k)$ , número para el cual al menos una de las matrices  $\mathbf{F}_a$  y  $\mathbf{E}_a$  sería nula, haciendo el producto cruzado entre éstas nulo; el número máximo de variables latentes a calcular es aquel para el cual  $\mathbf{E}_{a+1}$  no contiene más información sobre  $\mathbf{F}_{a+1}$ , en otras palabras,  $\mathbf{E}_{a+1}$  y  $\mathbf{F}_{a+1}$  están no correlacionadas; ésto no significa que las matrices de residuales deben ser cero para que el algoritmo pare, por ejemplo, cuando no existe dependencia lineal entre  $\mathbf{X}$  y  $\mathbf{Y}$ ,

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

el producto  $(n-1)^{-1} \mathbf{X}'\mathbf{Y} = \mathbf{0}$  y, NIPALS MCP2 no puede realizar ningún cálculo, entonces las matrices de residuales  $\mathbf{E}$  y  $\mathbf{F}$  son iguales a  $\mathbf{X}$  y  $\mathbf{Y}$ , respectivamente.

Por construcción de los conjuntos de variables latentes de  $\mathbf{X}$  y  $\mathbf{Y}$ , los vectores de pesos  $\underline{w}_{m+1}$  y  $\underline{v}_{m+1}$  son elegidos como:

$$\underline{w}_{m+1} \text{ es un vector propio de } \mathbf{E}'_m \mathbf{F}_m \mathbf{F}'_m \mathbf{E}_m$$

$$\underline{v}_{m+1} \text{ es un vector propio de } \mathbf{F}'_m \mathbf{E}_m \mathbf{E}'_m \mathbf{F}_m$$

Nota<sup>21</sup>: Siguiendo NIPALS MCP2, los vectores  $\underline{w}_{m+1}$  y  $\underline{v}_{m+1}$  no necesariamente son el vector singular izquierdo y derecho del producto  $\mathbf{X}'\mathbf{Y}$ , respectivamente.

**Propiedad 5.** – a) La variable latente  $\underline{t}_{m+1}$  de  $\mathbf{X}$ , es vector propio de la matriz  $\mathbf{E}_m \mathbf{E}'_m \mathbf{F}_m \mathbf{F}'_m$ . b) La variable latente  $\underline{u}_{m+1}$  de  $\mathbf{Y}$ , es vector propio de la matriz  $\mathbf{F}_m \mathbf{F}'_m \mathbf{E}_m \mathbf{E}'_m$ .

**Prueba a)** Por construcción, se tiene que:

$$\mathbf{E}'_m \mathbf{F}_m \mathbf{F}'_m \mathbf{E}_m \underline{w}_{m+1} = \tau_1^{m+1} \underline{w}_{m+1}$$

Multiplicando por la izquierda por  $\mathbf{E}_m$ , resulta:

$$\mathbf{E}_m \mathbf{E}'_m \mathbf{F}_m \mathbf{F}'_m \mathbf{E}_m \underline{w}_{m+1} = \tau_1^{m+1} \mathbf{E}_m \underline{w}_{m+1}$$

Pero  $\underline{t}_{m+1} = \mathbf{E}_m \underline{w}_{m+1}$ , entonces:

$$\mathbf{E}_m \mathbf{E}'_m \mathbf{F}_m \mathbf{F}'_m \underline{t}_{m+1} = \tau_1^{m+1} \underline{t}_{m+1}$$

**Prueba b)**

---

<sup>21</sup> Wegelin, A. J. (2000)

## Capítulo 4. Mínimos Cuadrados Parciales

$$\mathbf{F}'_m \mathbf{E}_m \mathbf{E}'_m \mathbf{F}_m \underline{v}_{m+1} = \tau_1^{m+1} \underline{v}_{m+1}.$$

Multiplicando por la izquierda por  $\mathbf{F}_m$ , se obtiene:

$$\mathbf{F}_m \mathbf{F}'_m \mathbf{E}_m \mathbf{E}'_m \underline{u}_{m+1} = \tau_1^{m+1} \underline{u}_{m+1}.$$

**Propiedad 6.** – a) Los vectores de pesos  $\underline{w}_1, \dots, \underline{w}_a$  de  $\mathbf{X}$ , son mutuamente ortogonales. b) Los vectores de pesos  $\underline{v}_1, \dots, \underline{v}_a$  de  $\mathbf{Y}$ , son mutuamente ortogonales.

**Prueba a)** De nueva cuenta, se expresa a  $\mathbf{E}_{m+1}$  de manera recursiva como:

$$\begin{aligned} \mathbf{E}_{m+1} &= \mathbf{E}_m - \underline{t}_{m+1} \underline{p}'_{m+1} \\ &= \mathbf{E}_m - \underline{t}_{m+1} \underline{t}'_{m+1} \mathbf{E}_m \\ &= \mathbf{A}_{m+1} \mathbf{E}_m \end{aligned} \quad (6.a)$$

Donde:  $\mathbf{A}_{m+1} = \mathbf{I} - \underline{t}_{m+1} \underline{t}'_{m+1}$ . De forma general para  $1 \leq m < n$

$$\begin{aligned} \mathbf{E}_n &= \mathbf{A}_n \mathbf{E}_{n-1} \\ &= \mathbf{A}_n \mathbf{A}_{n-1} \mathbf{E}_{n-2} \\ &= \mathbf{A}_n \cdots \mathbf{A}_m \mathbf{A}_{m+1} \mathbf{E}_m \\ &= \mathbf{H} \mathbf{A}_{m+1} \mathbf{E}_m \end{aligned} \quad (6.b)$$

Para  $\mathbf{H} = \mathbf{A}_n, \dots, \mathbf{A}_m$ . Por otro lado, se tiene que:

$$\begin{aligned} \mathbf{A}_{m+1} \mathbf{E}_m \underline{w}_{m+1} &= \mathbf{A}_{m+1} \underline{t}_{m+1} \\ &= \left( \mathbf{I} - \underline{t}_{m+1} \underline{t}'_{m+1} \right) \underline{t}_{m+1} \\ &= \underline{t}_{m+1} - \underline{t}_{m+1} \\ &= \underline{0} \end{aligned} \quad (6.c)$$

De (6.b) y (6.c), se obtiene:

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

$$\begin{aligned}
 \mathbf{E}_n \underline{w}_{m+1} &= \mathbf{H} \mathbf{A}_{m+1} \mathbf{E}_m \underline{w}_{m+1} \\
 &= \mathbf{H} \left( \mathbf{I} - \underline{t}_{m+1} \underline{t}'_{m+1} \right) \underline{t}_{m+1} \\
 &= \mathbf{H} \left( \underline{t}_{m+1} - \underline{t}_{m+1} \right) \\
 &= \mathbf{0}
 \end{aligned} \tag{6.d}$$

Ahora, por construcción

$$\mathbf{E}'_n \mathbf{F}_n \mathbf{F}'_n \mathbf{E}_n \underline{w}_{n+1} = \tau_1^{m+1} \underline{w}_{n+1}$$

Aplicando la transposición se obtiene:

$$\underline{w}'_{n+1} \mathbf{E}'_n \mathbf{F}_n \mathbf{F}'_n \mathbf{E}_n = \tau_1^{m+1} \underline{w}'_{n+1}$$

Multiplicando lo anterior por la derecha por  $\underline{w}_{m+1}$  se obtiene:

$$\underline{w}'_{n+1} \mathbf{E}'_n \mathbf{F}_n \mathbf{F}'_n \mathbf{E}_n \underline{w}_{m+1} = \tau_1^{m+1} \underline{w}'_{n+1} \underline{w}_{m+1}$$

Pero por (6.d) se cumple que  $\mathbf{E}_n \underline{w}_{m+1} = \mathbf{0}$ , entonces  $\underline{w}'_{n+1} \underline{w}_{m+1} = 0$ . Así, la matriz  $\mathbf{W}$  es ortonormal, pues se pide que  $\|\underline{w}_i\| = 1$  para  $i = 1, 2, \dots, a$ . Por simetría, las variables latentes de  $\mathbf{Y}$  son mutuamente ortogonales, esto es  $\underline{v}'_n \underline{v}_m = 0$ , para  $n > m$ ; haciendo a  $\mathbf{V}$  una matriz ortonormal siguiendo el mismo argumento que para  $\mathbf{W}$ .

**Propiedad 9.** – a) Los vectores de pesos  $\underline{w}_m$  son ortogonales a los vectores de cargas  $\underline{p}_n$ . b) Los vectores de pesos  $\underline{v}_m$  son ortogonales a vectores de cargas  $\underline{q}_n$ , esto es:

$$m < n \Rightarrow \begin{cases} \underline{w}_m \perp \underline{p}_n \\ \underline{v}_m \perp \underline{q}_n \end{cases}$$

**Prueba a)** Según lo definido en NIPALS MCP2, la  $n$ -ésima carga es:  $\underline{p}'_n = \underline{t}'_n \mathbf{E}_{n-1}$  operando la transposición se tiene que:

## Capítulo 4. Mínimos Cuadrados Parciales

$$\begin{aligned} \underline{p}_n &= E'_{n-1} t_n \Rightarrow w'_m \underline{p}_n = w'_m E'_{n-1} t_n \\ &= (E_{n-1} w_m)' t_n = 0 \text{ por (6.d)} \\ \therefore w'_m \underline{p}_n &= 0 \text{ para } m < n \end{aligned}$$

**Prueba b)** El razonamiento es el mismo para  $Y$ .

Esta propiedad garantiza la inversión del producto  $P'_a W_a$ , por ser una matriz triangular superior; la cual permite relacionar los vectores de pesos de  $X$  con las regresoras originales; para esto, considere el número  $a$  tal que la matriz de residuales  $E_a$  tenga entradas casi nulas (despreciables), entonces  $T = X M_a$ , ya que multiplicando por la izquierda a (4.32) por  $W_a$  se tiene:

$$XW_a = T_a P'_a W_a \Rightarrow T_a = XW_a (P'_a W_a)^{-1} = X M_a \text{ donde } M_a = W_a (P'_a W_a)^{-1}$$

La matriz  $P'_a W_a = \Psi_a$  es triangular superior con la estructura siguiente:

$$\begin{aligned} P'_a W_a &= \begin{pmatrix} \underline{p}_1 \\ \underline{p}_2 \\ \vdots \\ \underline{p}_a \end{pmatrix} (\underline{w}_1 \quad \underline{w}_2 \quad \cdots \quad \underline{w}_a) \\ &= \begin{pmatrix} 1 & \underline{p}_1 \underline{w}_2 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \underline{p}_2 \underline{w}_3 & 0 & & & 0 \\ \vdots & & 1 & \ddots & & & \vdots \\ 0 & & \ddots & \ddots & \underline{p}_{(i-1)} \underline{w}_i & \ddots & 0 \\ 0 & & & \ddots & 1 & \ddots & 0 \\ \vdots & & & & \ddots & \ddots & \underline{p}_{(a-1)} \underline{w}_a \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} = \Psi \end{aligned}$$

Pues, los elementos de la diagonal corresponden a:

$$\begin{aligned} \underline{p}_m \underline{w}_m &= t'_m E_{m-1} \underline{w}_m \\ &= t'_m t_m \\ &= 1 \end{aligned}$$

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

**Propiedad 10.** – Para  $m < n$ , se cumple:

$$\begin{cases} \mathbf{E}_n \mathbf{w}_m = 0 \\ \mathbf{F}_n \mathbf{v}_m = 0 \end{cases}$$

**Prueba:** Ver la propiedad 6, inciso (6.d). Por simetría, sigue para  $\mathbf{Y}$ .

**Propiedad 11.** – a) El producto  $\mathbf{P}_m' \mathbf{M}_m$  es igual a la matriz identidad de orden  $m \times m$ .

b) El producto  $\mathbf{M}_m' \mathbf{P}_m$  es igual a la matriz identidad de orden  $m \times m$

**Prueba a)**

$$\begin{aligned} \mathbf{P}_m' \mathbf{M}_m &= (\mathbf{T}_m' \mathbf{T}_m)^{-1} \mathbf{T}_m' \mathbf{X} \mathbf{M}_m \\ &= (\mathbf{T}_m' \mathbf{T}_m)^{-1} \mathbf{T}_m' \mathbf{T}_m \\ &= \mathbf{I}_m \end{aligned}$$

**Prueba b)**

$$\begin{aligned} \mathbf{M}_m' \mathbf{P}_m &= \mathbf{M}_m' \mathbf{X}' \mathbf{T}_m (\mathbf{T}_m' \mathbf{T}_m)^{-1} \\ &= \mathbf{T}_m' \mathbf{T}_m (\mathbf{T}_m' \mathbf{T}_m)^{-1} \\ &= \mathbf{I}_m \end{aligned}$$

#### 4.3.3 Regresión por Mínimos Cuadrados Parciales Multivariado (RMCP2)

Una vez construidos los conjuntos  $\{\underline{t}_1, \underline{t}_2, \dots, \underline{t}_a\}$  y  $\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_a\}$ , el modelo de regresión constituido por las primeras  $a$  variables latentes retenidas será utilizado para la predicción de observaciones futuras. Partiendo de los modelos externos (4.32) y (4.33) y del modelo interno dado en (4.34):

$$\mathbf{X} = \mathbf{T}_a \mathbf{P}_a' + \mathbf{E}_a, \quad \mathbf{Y} = \mathbf{U}_a \mathbf{Q}_a' + \mathbf{F}_a \quad \text{y} \quad \mathbf{U}_a = \mathbf{T}_a \mathbf{C}_a + \mathbf{\Omega}_a$$

Los valores ajustados para las variables de respuesta por MCP2, quedan expresados como:

## Capítulo 4. Mínimos Cuadrados Parciales

$$\widehat{Y}_{MCP2}^a = T_a \widehat{C}_a Q_a' = T_a \widehat{\beta}_{MCP2}^a \quad (4.45)$$

Al igual que MCP1, la matriz de variables latentes  $T_a$  está relacionada con las variables independientes originales de la siguiente manera:

$$T_a = X_a W_a (P_a' W_a)^{-1} = X M_a \quad (4.46)$$

Con  $M_a = W_a (P_a' W_a)^{-1}$ . El modelo de regresión para  $Y$  dadas las variables regresoras originales explícitamente, está dado por:

$$\widehat{Y}_{MCP2}^a = X W_a (P_a' W_a)^{-1} \widehat{C}_a Q_a'$$

Con  $\widehat{\beta}_{MCP2}^a = W_a (P_a' W_a)^{-1} \widehat{C}_a Q_a'$ .

### 4.3.4 Predicción

Para los propósitos de predicción, contando con  $\underline{x}_0 = (x_{01}, x_{02}, \dots, x_{0p})'$  una nueva observación de las variables de  $X$ , entonces:

$$Y_{MCP2}^a = \underline{\mu}_y^* + \left( (\underline{x}_0 - \underline{\mu}_x^*)' \Psi \right) \widehat{\beta}_{MCP2}^a \Omega \quad (4.47)$$

Donde:  $\widehat{\beta}_{MCP2}^a = W_a (P_a' W_a)^{-1} \widehat{C}_a Q_a'$ ,  $\underline{\mu}_y^* = \frac{1}{n} Y' \underline{1}_n$ ,  $\Omega = \text{diag}(\sigma_{y_1^*}, \sigma_{y_2^*}, \dots, \sigma_{y_q^*})$  con

$\sigma_{y_j^*} = \sqrt{\frac{\sum_{i=1}^n (y_{ij} - \mu_j)^2}{n-1}}$  para  $i = 1, \dots, q$ , corresponden al vector de medias muestral y a las desviaciones estándar de las variables de respuesta originales (no estandarizadas). Asimismo, los valores de  $\underline{\mu}_x^* = \frac{1}{n} X' \underline{1}_n$ ,  $\Psi = \text{diag}(\sigma_{x_1^*}^{-1}, \sigma_{x_2^*}^{-1}, \dots, \sigma_{x_p^*}^{-1})$

con  $\sigma_{x_j^*} = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \mu_j)^2}{n-1}}$  para  $i = 1, \dots, p$ , corresponden a las variables regresoras

### 4.3 Mínimos Cuadrados Parciales Multivariado (MCP2)

originales (no estandarizadas). Es decir, el proceso de predicción se inicia con la nueva unidad estandarizada. Los valores de (4.47) están en las unidades de medida de las variables de respuesta originales.

#### 4.3.5 Intervalos de confianza para $\hat{\beta}_{MCP}$ y $\hat{y}_{MCP}$

Debido a que la distribución de MCP es desconocida, no existen pruebas de significancia convencionales; éstas pueden ser desarrolladas a través de métodos bootstrap, tales como jackknife el cual es un método de re-muestreo<sup>22</sup>.

### 4.4 Selección del modelo

En regresión, la meta principal es la predicción de la(s) variable(s) de respuesta, por tanto, el dilema es elegir el número óptimo de variables latentes que proporcione gran calidad predictiva al modelo de RMCP. Ya que la calidad de un modelo de regresión es equivalente a la calidad de predicción que éste posea, se introducen varias medidas de “ajuste del modelo a los datos” en las cuales se intenta estimar la desviación promedio del modelo a los datos. Se debe tener en cuenta que, la capacidad de interpretación decrece al aumentar el número de variables latentes.

#### 4.4.1 Por validación cruzada

En esta tesis, se elige la técnica de Leave One Out Cross-Validation (LOOCV) como método de evaluación para la capacidad predictiva de un modelo de RMCP. Esta es un caso particular del método de validación cruzada (cross-validation), el cual deja una unidad de observación fuera de tal forma que, con los datos restantes se construye el modelo. El proceso se repite tantas veces como el número de observaciones. Este procedimiento se lleva a cabo para cada variable latente agregada al modelo.

- **Leave one out cross-validation (LOOCV)**

Para un modelo de RMCP, con  $a$  variables latentes:

---

<sup>22</sup> Pirouz, D. (2006), Abdi, H. (2010)



## Capítulo 4. Mínimos Cuadrados Parciales

1. Construir el modelo de regresión, excluyendo la  $i$ -ésima observación, con  $i = 1, \dots, n$
2. Calcular la predicción para la unidad de observación que no fue incluida,  $\hat{y}_i$  de  $i = 1, \dots, n$ , Para un valor  $i$  fijo.
3. Calcular el residual correspondiente  $\varepsilon_i = y_i - \hat{y}_i$  de  $i = 1, \dots, n$
4. Calcular la Suma de Cuadrados de los Residuales de Validación Cruzada, SCRVC como:

$$SCRVC = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

La SCRVC para un modelo MCP2, es calculado como:

$$SCRVC = \left\| \mathbf{Y} - \hat{\mathbf{Y}}_{MCP2}^a \right\|^2$$

$$\text{Donde } \left\| \mathbf{Y} - \hat{\mathbf{Y}}_{MCP2}^a \right\|^2 = \text{tr} \left( \left( \mathbf{Y} - \hat{\mathbf{Y}}_{MCP2}^a \right) \left( \mathbf{Y} - \hat{\mathbf{Y}}_{MCP2}^a \right)' \right)$$

5. Calcular la SCRVC promedio, SCRPMC , como:

$$SCRPMC = \frac{SCRVC}{n}$$

O la raíz de la SCRPMC , dada por:

$$RSCRMCP = \sqrt{SCRPMC}$$

Realizar los cálculos anteriores, utilizando como regresora la primera variable latente, después, la primera y segunda variable latente, e ir incrementando el número de variables latentes, es decir, ir ajustando el modelo sobre las primeras  $a$  variables latentes con  $a = 1, 2, \dots, r$ .

El número de variables latentes que será utilizado es obtenido siguiendo:

## 4.4 Selección del Modelo

$$a^* = \min \left\{ a > 1 : \text{SCRVC}_{a+1} - \text{SCRVC}_a > 0 \right\} \quad (4.48)$$

Donde  $a^*$  es el número óptimo.

En otras palabras, el modelo óptimo de RMCP es aquel que contenga el primer mínimo local del valor de la SCRVC, y será equivalente al primer mínimo local de la RSCRVC. Otro juicio sencillo para parar el incremento de variables latentes al modelo, es tan pronto como la SCRVC decrezca ligeramente.

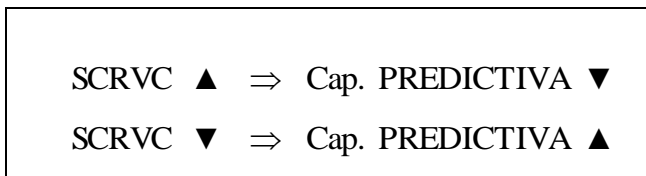


Figura: Relación existente entre SCRVC y la capacidad predictiva.

La calidad de predicción, aumenta conforme el número de variables latentes lo hace, debido a la variabilidad aportada por cada variable latente, sin embargo este aumento posee un límite, un

decremento en la calidad predictiva sucede si se añaden más variables latentes de las necesarias, este punto de decremento indica un “sobre ajuste” del modelo a los datos; en tal caso, el modelo de regresión propuesto no es el adecuado para predecir observaciones futuras, en virtud de que usualmente la SCRVC incrementa ligeramente cuando una variable innecesaria es agregada, señalando sobre ajuste, pero incrementa bastante si alguna variable importante es removida.

### 4.4.2 Por variabilidad explicada por las primeras $j$ – variables latentes

La varianza de  $\mathbf{X}$  explicada por las primeras  $\{t_1, \dots, t_j\}$  variables latentes está dada por:

$$\frac{\|\mathbf{X}\|^2 - \|\mathbf{E}_j\|^2}{\|\mathbf{X}\|^2} (100) \quad (4.49)$$

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

Donde  $\|\mathbf{X}\|^2 = \text{tr}(\mathbf{X}'\mathbf{X})$  llamada norma de Frobenius. Este resultado sigue la descomposición bilineal de  $\mathbf{X}$  que efectúa el método MCP, ya que la variabilidad de  $\mathbf{X}$  no explicada por las  $j$ -primeras variables latentes es capturada en la matriz de residuales  $\mathbf{E}_j$ . Si la matriz de residuales  $\|\mathbf{E}_j\|^2$  es nula, el conjunto de variables latentes explica completamente la variabilidad total de  $\mathbf{X}$ . En el caso multivariado, el porcentaje de varianza de  $\mathbf{Y}$  explicada por las primeras  $\{u_1, \dots, u_j\}$  variables latentes está dado por:

$$\frac{\|\mathbf{Y}\|^2 - \|\mathbf{F}_j\|^2}{\|\mathbf{Y}\|^2} \quad (100) \quad (4.50)$$

Para la RMCP1, la varianza de la variable de respuesta explicada por el modelo con las primeras  $j$ - variables latentes de  $\mathbf{X}$  como nuevas regresoras es:

$$R_j^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i^j)^2}{\sum_{i=1}^n (y_i^j - \bar{y})^2}$$

Para RMCP2, la varianza de las variables de respuesta explicada por el modelo con las primeras  $j$ - variables latentes de  $\mathbf{X}$  como nuevas regresoras es:

$$R_j^2 = 1 - \frac{\text{tr}((\mathbf{Y} - \hat{\mathbf{Y}}_{\text{MCP2}}^j)'(\mathbf{Y} - \hat{\mathbf{Y}}_{\text{MCP2}}^j))}{\text{tr}(\mathbf{Y}'\mathbf{Y})} \quad (4.51)$$

La elección del número  $a^*$  de variables latentes a retener, se realiza de tal manera que la varianza explicada de la(s) variable(s) dependiente(s) y de las variables independientes sea grande para ambas.

### 4.4.3 Interpretación de los variables obtenidas por MCP

Para la interpretación de las variables de MCP, el analista puede ayudarse examinando gráficos semejantes a los empleados en el ACP; por ejemplo, diagramas

## 4.4 Selección del Modelo

de dispersión (nube de puntos-variable, nube de puntos-observaciones) para las variables latentes  $\underline{t}_i$  y  $\underline{u}_i$ , así como de los vectores de pesos  $\underline{w}_i$  y  $\underline{v}_i$ ; también es posible construir círculos de correlación, gráficos de barras, etc., con el propósito de detectar alguna anomalía en la información, tal como no-linealidad de los datos, clusters o datos atípicos.

Por construcción, las entradas de los vectores de pesos, indican cuanto contribuye cada una de las variables regresoras en la descripción de las variables de respuesta; éstos son usados para graficar las posiciones de las variables regresoras originales en el espacio generado por las variables latentes. En el caso multivariado, los vectores de pesos de  $\mathbf{X}$  muestran la correlación de las variables regresoras con las variables latentes de  $\mathbf{Y}$ ; los signos de las entradas indican la dirección de la correlación (positiva, negativa o correlación cero).

Los siguientes gráficos pueden ser construidos:

- Scatterplot de cada  $\underline{t}_i$  vs  $\underline{t}_j$ , para  $i=1,2,\dots,a$  y  $j=1,2,\dots,a$
- Scatterplot de cada  $\underline{p}_i$  vs  $\underline{p}_j$ , para  $i=1,2,\dots,a$  y  $j=1,2,\dots,a$
- Gráfica de valores observados y valores ajustados,  $y_i$  vs  $\hat{y}_i$ , para  $i=1,2,\dots,a$

Entre otros.

## 4.5 Otras variantes de MCP

Existen múltiples algoritmos que responden al nombre de Mínimos Cuadrados Parciales, que siguen un proceso iterativo para extraer los conjuntos de vectores de pesos, de variables latentes y de cargas de  $\mathbf{X}$  y en el caso multivariado, el de cargas de  $\mathbf{Y}$ , sin embargo se diferencian por las imposiciones aplicadas para su construcción.

- **“MCP con cargas ortogonales”**

Herarld Martens (1985) con la cooperación de Wold y en paralelo a su trabajo<sup>23</sup>, diseñó otro algoritmo matemáticamente equivalente a NIPALS MCP, en el cual los

---

<sup>23</sup> Ergon, R. (2009).

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

requerimientos de ortogonalidad recaen sobre el conjunto de cargas  $\{p_1, \dots, p_a\}$  más no en el conjunto  $\{t_1, \dots, t_a\}$ . Helland (1988) da una prueba formal de la equivalencia de ambos algoritmos, al desglosar las estructuras algebraicas que los relacionan, demostrando también que dan la misma ecuación de predicción. Desde un punto de vista práctico, el algoritmo de Wold es preferido sobre el de Martens por varias razones: NIPALS MCP1 comprende cálculos más sencillos y a la vez, los conjuntos de variables latentes y de cargas son más fáciles de entender e interpretar, que el algoritmo de Martens<sup>24</sup>.

### • Método PLS-SVD

La técnica PLS-SVD (Partial Least Squares-Singular Value Decomposition), es una variante de MCP que sigue el mismo criterio de maximización de covarianza, pero solo tiene una iteración; PLS-SVD realiza la descomposición en valores singulares del producto cruzado de  $X$  y  $Y$  solo una vez, a saber:

$$X'Y = W\Delta V'$$

Entonces se calculan las matrices de variables latentes para las variables originales como:  $T = XW$  y  $U = YV$ . En este caso, las matrices de variables latentes  $T$  y  $U$  no necesariamente son ortogonales; su producto resulta ser una matriz diagonal, pues:

$$\begin{aligned} T'U &= (XW)'YV \\ &= W'X'YV \\ \therefore T'U &= \Delta \end{aligned}$$

En PLS-SVD, los papeles de  $X$  y  $Y$  son simétricos, por tanto, si se desea predecir  $X$  de  $Y$ , los conjuntos de variables latentes serán idénticos, lo que clasifica a PLS-SVD como bi-direccional, a diferencia de MCP2, que toma los roles de  $X$  y  $Y$  de forma asimétrica, considerándolo uni-direccional<sup>25</sup>.

---

<sup>24</sup> Helland, I. (1988).

<sup>25</sup> Abdi, H. (2010)

- **Método SIMPLS**

Lyan Trigg modificó NIPALS MCP para dar origen a ésta versión, la cual evita los pasos de actualización de la matriz  $\mathbf{X}$  y del vector o matriz de respuestas  $\mathbf{Y}$ , al actualizar directamente el producto  $\mathbf{X}'\mathbf{y}$  y los vectores de pesos  $\underline{w}_i$  ponderan directamente a las columnas de las variables originales. Bajo esta versión, las variables latentes son mutuamente ortogonales no siendo así para el conjunto de las cargas. Se ha mostrado que la ejecución de SIMPLS es equivalente a la de MCP1, pero difiere en el caso multivariado, MCP2<sup>26</sup>.

- **Normalización en diferentes puntos del algoritmo**

Versiones modificadas de NIPALS MCP (univariado/multivariado), ejecutan los pasos de normalización de las variables en otros puntos del algoritmo, esto dificulta la comparación de los valores de las variables obtenidas de diferentes versiones; sin embargo el resultado final de predicción no se ve alterado<sup>27</sup>.

- **Respuestas Categóricas**

Aunque MCP fue desarrollado para variables continuas (regresoras y respuestas), éste puede ser exitosamente aplicado sobre respuestas categóricas<sup>28</sup> (binarias o multicategóricas).

## 4.6 Consideraciones Finales

A pesar de que los orígenes de MCP estuvieron basados en el ACP, en esta sección se discuten las principales diferencias entre ellos, pues sus resultados difieren significativamente, sobre todo si los datos presentan alta dimensionalidad de regresoras y/o multicolinealidad perfecta.

La literatura sobre el método de MCP, generalmente argumenta la preferencia de MCP sobre el ACP; de hecho algunos autores consideran al ACP como un método de reducción de dimensionalidad que permite visualizar, clasificar y diagnosticar la

---

<sup>26</sup> Boulesteix, A. L. and Strimmer, K. (2006).

<sup>27</sup> Kleinknecht, D. (Junio, 2002)

<sup>28</sup> Vega, J.C. (2004)

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

información de las regresoras más no como un método de regresión, pues nada garantiza que las componentes principales sean relevantes para predecir o modelar la información de la(s) variable(s) de respuesta, ya que en la construcción de las componentes, la información de la respuesta no es tomada en consideración. Para ilustrar esto, considere el caso donde el ACP excluya una o varias variables regresoras por tener varianza mínima, pero que tengan alta correlación con  $\underline{y}$ , esta acción resta ajuste al modelo de regresión y afecta las predicciones; en cambio, por la naturaleza de las variables latentes de MCP, toda la variabilidad de  $\underline{X}$  que influencia a la(s) variable(s) de respuesta, es utilizada en el modelo de regresión y el primer mínimo local de la SCRVC del modelo de RMCP, se consigue con un menor número de componentes que con la RCP<sup>29</sup>, lo cual, además de minimizar los cálculos, puede ayudar a la interpretación de los modelos obtenidos por MCP1 y MCP2, en algunos casos.

El método de MCP es contemplado como método de predicción más que de interpretación<sup>30</sup>, lo que se podría considerar como desventaja, pues muchos analistas lo distinguen por su buen funcionamiento, sin embargo la comprensión de sus propiedades e interpretación de las variables obtenidas mediante ambos algoritmos (univariado y multivariado), es compleja. Otra desventaja, engloba el hecho de que bajo MCP, la distribución del estimador de coeficientes de regresión es desconocida, al igual que la distribución de la respuesta estimada, no existiendo pruebas de significancia para los parámetros.

Asimismo, el modelo de MCP es correlativo y no causal, por consiguiente, no ofrece información fundamental acerca del fenómeno estudiado. Una crítica más profunda, desde el punto de vista de la teoría estadística, es que MCP no puede ser considerado un método óptimo<sup>31</sup>.

El siguiente cuadro resume algunas de las principales diferencias en las variables obtenidas por medio de ambos métodos.

---

<sup>29</sup> Frank and Friedman. (2003)

<sup>30</sup> Pirouz, D. (2006), Helland, I. (1988)

<sup>31</sup> Helland (2001)

## 4.6 Consideraciones Finales

ACP vs MCP	
Método ACP	Método MCP
El ACP está basado en la descomposición en valores singulares de la matriz de varianzas y covarianzas de $X$ .	El método MCP está basado en descomposiciones en valores singulares sucesivas de la matriz de covarianzas de los residuales de $X$ y $y$ .
Los vectores de pesos son ponderadores de las variables regresoras, tal que explican la covarianza entre las variables regresoras. La matriz $W$ es ortogonal.  No hay vectores de pesos para la(s) variable(s) de respuesta.	Los vectores de pesos de $X$ son ponderadores de las variables regresoras tales que explican la covarianza entre las variables regresoras y las variables respuesta.  En caso multivariado, los vectores de pesos de $Y$ ponderan a las variables de respuesta tal que las variables latentes de $X$ tengan correlación máxima con las variables latentes de $Y$ .  La matriz $W$ es ortonormal. La matriz $V$ es ortonormal.
Las componentes principales explican a las variables regresoras más no a la(s) variable(s) de respuesta	Las variables latentes explican la covarianza entre las variables regresoras y las variables de respuesta.
El conjunto de componentes principales es ortogonal; a su vez el conjunto de cargas también es ortogonal.	El conjunto de variables latentes de $X$ es ortogonal no siendo así el conjunto de cargas de $X$ .  En caso multivariado, el conjunto de variables latentes de $Y$ es ortogonal, no siendo así el conjunto de cargas de $Y$ .

Tabla<sup>32</sup>: Principales diferencias entre el ACP y MCP.

<sup>32</sup> Elaboración propia.



## Capítulo 4. Mínimos Cuadrados Parciales

Sin embargo, cuando la RLM es aplicable a los datos se tienen las siguientes igualdades:

$$\hat{\beta}_{RLM} = \hat{\beta}_{PCA} = \hat{\beta}_{MCP}$$

### 4.7 Ejemplo de MCP1 en R

En la tabla 1, se presenta un típico proceso de quimiometría<sup>33</sup>, área donde la multicolinealidad es un problema frecuente. Se trata de datos concernientes al porcentaje de conversión de *n*-heptano en acetileno, dados los incrementos de tres variables regresoras; la primera variable representa la temperatura medida en grados Celsius (°C), la segunda variable, la proporción de H<sub>2</sub> para el *n*-heptano medido en moles; finalmente, la tercera variable representa el tiempo de contacto, medido en segundos.

El objetivo es ajustar un modelo de regresión, para poder obtener predicciones sobre la conversión de *n*-heptano en acetileno para muestras futuras.

En la práctica, se acostumbra dividir el número de observaciones (muestras) en dos partes, la primera comúnmente llamada conjunto de entrenamiento<sup>34</sup>, y la segunda denominada conjunto de prueba<sup>35</sup>, que tiene la finalidad de validar el modelo; al implementar el modelo obtenido del conjunto de entrenamiento sobre el conjunto de prueba. Por tanto, dividiremos los datos en dos conjuntos, el primero estará conformado por las primeras diez observaciones, el segundo por las restantes.

Donde:

$\underline{y}$  := Conversión de *n*-heptano en acetileno

$\underline{x}_1$  := Temperatura

$\underline{x}_2$  := Razón Molar de H<sub>2</sub>

$\underline{x}_3$  := Tiempo de contacto

---

<sup>33</sup> Montgomery, D. C. and Peck, A. E. (1992)

<sup>34</sup> Un conjunto de entrenamiento es aquel para el cual se conoce la propiedad a determinar, es representativo del sistema en estudio y debe permitir realizar predicciones sobre un conjunto de muestras futuras.

<sup>35</sup> Para la etapa de validación del modelo, el conjunto de prueba es aquel para el cual se conoce la propiedad a determinar y no ha sido utilizado para la construcción del modelo.

## 4.7 Ejemplo de MCP1 en R

No. Observación	$\underline{y}$	$\underline{X}_1$	$\underline{X}_2$	$\underline{X}_3$
1	49.0	1,300	7.5	0.0120
2	50.2	1,300	9.0	0.0120
3	50.5	1,300	11.0	0.0115
4	48.5	1,300	13.5	0.0130
5	47.5	1,300	17.0	0.0135
6	44.5	1,300	23.0	0.0120
7	28.0	1,200	5.3	0.4000
8	31.5	1,200	7.5	0.0380
9	34.5	1,200	11.0	0.0320
10	35.0	1,200	13.5	0.0260
11	38.0	1,200	17.0	0.0340
12	38.5	1,200	23.0	0.0410
13	15.0	1,100	5.3	0.0840
14	17.0	1,100	7.5	0.0980
15	20.5	1,100	11.0	0.0920
16	29.5	1,100	17.0	0.0860

- **Pre-procesamiento de los datos**

Las tres variables regresoras del ejemplo, están en unidades de medida diferentes, entonces se comienza el análisis estandarizando las variables originales (ver sección 2.2).

- **Análisis de multicolinealidad**

Para la detección de multicolinealidad en los datos, primero se analiza la matriz de correlaciones de las variables regresoras, dada por:

---

<sup>36</sup> Ryan, P. T. (1997)

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

$$\mathbf{R}_{\mathbf{X}\mathbf{X}} = \begin{pmatrix} 1 & 0.2236 & -0.9582 \\ 0.2236 & 1 & -0.2402 \\ -0.9582 & -0.2402 & 1 \end{pmatrix}$$

Note que la primera regresora está negativamente correlacionada con la tercera, lo cual indica que a mayor temperatura menor tiempo de contacto y viceversa, entonces se puede decir que existe multicolinealidad en los datos. Para reafirmar este hecho, obtengamos el Número Condición; los valores propios de  $\mathbf{X}'\mathbf{X}$  son:

$$\lambda_1 = 23'622,500, \lambda_2 = 459.6583, \lambda_3 = 0.01732$$

Las diferencias numéricas entre ellos, nos llevan a sospechar algún grado de multicolinealidad para los datos; el Número Condición es:

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{23'622,500}{0.01732} = 1364'261,405$$

Ya que  $\kappa > 1000$ , entonces las regresoras presentan multicolinealidad perfecta. Se considera que es implementar MCP1 es ideal, a pesar de no tener alta dimensionalidad de variables regresoras.

- **Mínimos cuadrados parciales univariado, MCP1**

Las tablas 3–6, muestran las matrices obtenidas al ejecutar NIPALS MCP1 en el paquete estadístico R<sup>37</sup>.

Debido a la estandarización de las variables originales, el vector de pesos  $\underline{w}_1$  refleja la razón de cambio del porcentaje de n – heptano de acetileno, respecto a los valores de las tres variables regresoras.

---

<sup>37</sup> Paquete Estadístico R versión 2.11.0

## 4.7 Ejemplo de MCP1 en R

No. Observación	$\underline{y}$	$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$
1	0.8153	0.7746	-0.8272	-0.7638
2	0.9535	0.7746	-0.5406	-0.7638
3	0.9880	0.7746	-0.1586	-0.8063
4	0.7577	0.7746	0.3190	-0.6790
5	0.6426	0.7746	0.9877	-0.6365
6	0.2971	0.7746	2.1339	-0.7638
7	-1.6030	-1.1619	-1.2475	1.6125
8	-1.1999	-1.1619	-0.8272	1.4428
9	-0.8545	-1.1619	-0.1586	0.9336
10	-0.7969	-1.1619	0.3190	0.4243

Variable	$\underline{w}_1$	$\underline{w}_2$	$\underline{w}_3$
Temperatura (°C)	0.682	0.171	-0.712
Razón Molar de H <sub>2</sub>	0.240	-0.971	0
Tiempo	-0.691	-0.169	-0.703

Una vez obtenidos los pesos, las tres primeras variables latentes resultan una suma ponderada de las variables regresoras, esto es:

$$\underline{t}_1 = 0.682 \underline{x}_1 + 0.240 \underline{x}_2 - 0.691 \underline{x}_3$$

$$\underline{t}_2 = 0.171 \underline{x}_1 - 0.971 \underline{x}_2 - 0.169 \underline{x}_3$$

$$\underline{t}_3 = -0.712 \underline{x}_1 - 0.703 \underline{x}_3$$

<sup>38</sup> Las tablas 2-13 y las gráficas 1-5, se elaboraron en el paquete R versión 2.11.0. Las instrucciones para ejecutar RMCP pueden verificarse en el Anexo B.

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

Tabla 4 : Matriz T			
Observación	$t_1$	$t_2$	$t_3$
1	0.8574	1.2012	0.0889
2	0.9262	0.9340	0.0656
3	1.0472	0.5896	0.0653
4	1.0739	0.1087	-0.0660
5	1.2050	-0.5266	-0.1512
6	1.5681	-1.5598	-0.1521
7	-2.2060	0.3879	-0.2698
8	-1.9878	0.0434	-0.1807
9	-1.4753	-0.4377	0.1346
10	-1.0087	-0.7408	0.4654

Para  $t_1$ , el menor peso lo posee la variable razón molar de  $H_2$  ( $x_2$ ), mientras que la temperatura ( $x_1$ ) y el tiempo ( $x_3$ ), ya que éstas presentan una fuerte correlación positiva y negativa con el porcentaje de conversión de n-heptano en acetileno ( $y$ ), respectivamente. La variable latente  $t_2$  recoge la variabilidad no explicada por  $t_1$ , entonces la ponderación es mayor sobre la variable que mide la razón molar de  $H_2$  ( $x_2$ ) pues en el primer residual  $f_1$ , existe información de esta variable.

Tabla 5 : Matriz P			
Variable	$p_1$	$p_2$	$p_3$
°C	0.654	0.231	-0.712
Razón Molar de $H_2$	0.395	-0.971	0
Tiempo	-0.664	-0.11	-0.703

Tabla 6 : Vector $\underline{q}$		
$q_1$	$q_2$	$q_3$
0.640	0.322	0.419

- **Regresión por Mínimos Cuadrados Parciales Univariado**

A partir de las matrices anteriores, se calculan los coeficientes de regresión por MCP1 para distintos números de variables latentes, como se muestra en la siguiente tabla.

Tabla 7 : Coeficientes de Regresión para $\underline{y}$ por MCP1			
	Con una variable latente	Con 2 variables latentes	Con 3 variables latentes
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
	%	%	%
°C	0.4364	0.5265	0.2383
Razón Molar de H <sub>2</sub>	0.1537	-0.1469	-0.1810
Tiempo	-0.4427	-0.5328	-0.8369

Las entradas de los  $\hat{\beta}_i$  son muy parecidas, esto se debe a que la primera variable latente aporta la mayoría de variabilidad global del sistema (ver tablas 12 y 13), pues explica casi perfectamente a la respuesta y las variables latentes restantes aportan porcentajes de variabilidad muy pequeños, lo que hace muy parecidas las entradas de  $\hat{\beta}_2$  y  $\hat{\beta}_3$ . Este hecho, trae consigo que las predicciones para  $\underline{y}$  sean bastantes similares, utilizando a  $\underline{t}_1$  y  $\underline{t}_2$  como regresoras que con las tres variables latentes (ver tabla 8), lo que significa que prescindir de  $\underline{t}_3$  para el modelo de RMCP1 no altera las

## Capítulo 4. Mínimos Cuadrados Parciales (MCP)

predicciones, prueba de esto se presenta en la tabla 10. Asimismo, aplicando la regla del codo para la SCRVC de la gráfica 6, se concluye que el primer mínimo local se da en la segunda componente; cabe mencionar que la primera variable latente es por sí sola una excelente regresora para la variable de respuesta al explicar el 90 % su variabilidad, pues compacta la información de los datos originales con la menor pérdida de información posible. La tabla 9 muestra los residuales de  $\underline{y}$ , los cuales son muy pequeños para todos los casos, se puede decir que tanto  $\underline{f}_2$  como  $\underline{f}_3$  contienen residuales puramente aleatorios.

Tabla 8: Valores Ajustados de $\underline{y}$			
Observación	Con una variable latente $\underline{\hat{y}}_{MCP1}^1$	Con 2 variables latentes $\underline{\hat{y}}_{MCP1}^2$	Con 3 variables latentes $\underline{\hat{y}}_{MCP1}^3$
1	46.6879	50.0504	50.3737
2	47.0703	49.6848	49.9233
3	47.7433	49.3938	49.6312
4	47.8912	48.1956	47.9555
5	48.6203	47.1463	46.5963
6	50.6393	46.2732	45.7200
7	29.6533	30.7390	29.7579
8	30.8666	30.9882	30.3309
9	33.7164	32.4911	32.9806
10	36.3113	34.2377	35.9305

#### 4.7 Ejemplo de MCP1 en R

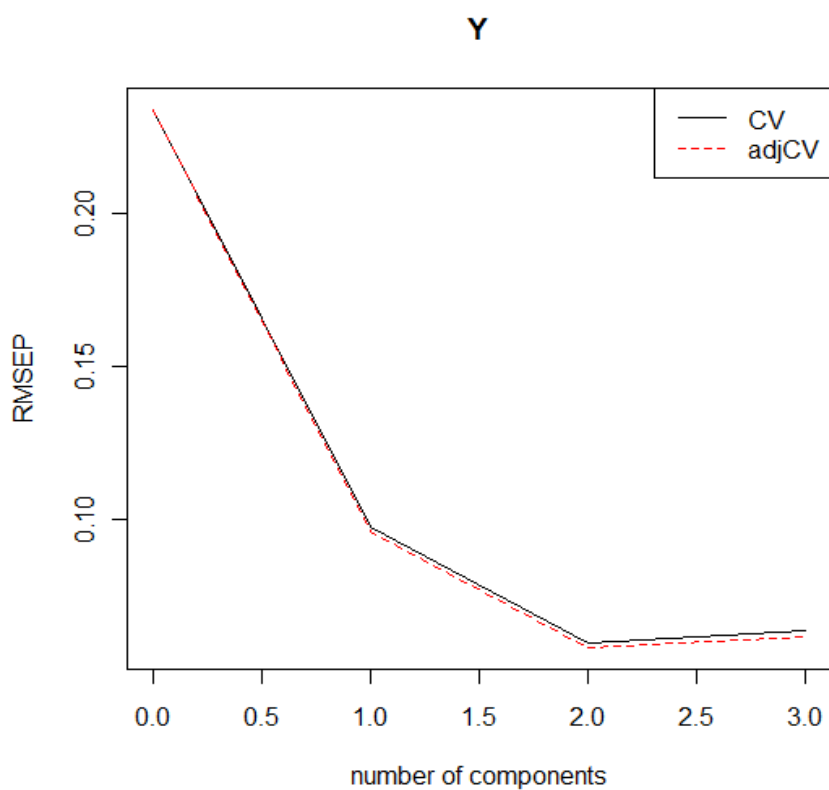
Tabla 9: Residuales de $\underline{y}$			
Observación	Con una variable latente $\underline{f}_1$	Con 2 variables latentes $\underline{f}_2$	Con 3 variables latentes $\underline{f}_3$
1	0.2663	-0.1210	-0.1582
2	0.3604	0.0593	0.0319
3	0.3175	0.1274	0.1000
4	0.0701	0.0351	0.0627
5	-0.1290	0.0407	0.1041
6	-0.7070	-0.2042	-0.1405
7	-0.1904	-0.3154	-0.2024
8	0.0729	0.0589	0.1346
9	0.0902	0.2313	0.1750
10	-0.1510	0.0878	-0.1072

Tabla 10: Predicciones			
Observación	Con una variable latente $\underline{y}_{MCP1}^1$	Con 2 variables latentes $\underline{y}_{MCP1}^2$	Con 3 variables latentes $\underline{y}_{MCP1}^3$
11	42.07	40.56	41.79
12	40.82	37.64	37.42
13	40.27	34.57	32.96
14	21.38	18.93	14.56
15	18.45	14.53	7.77
16	20.53	15.63	9.71



Capítulo 4. Mínimos Cuadrados Parciales (MCP)

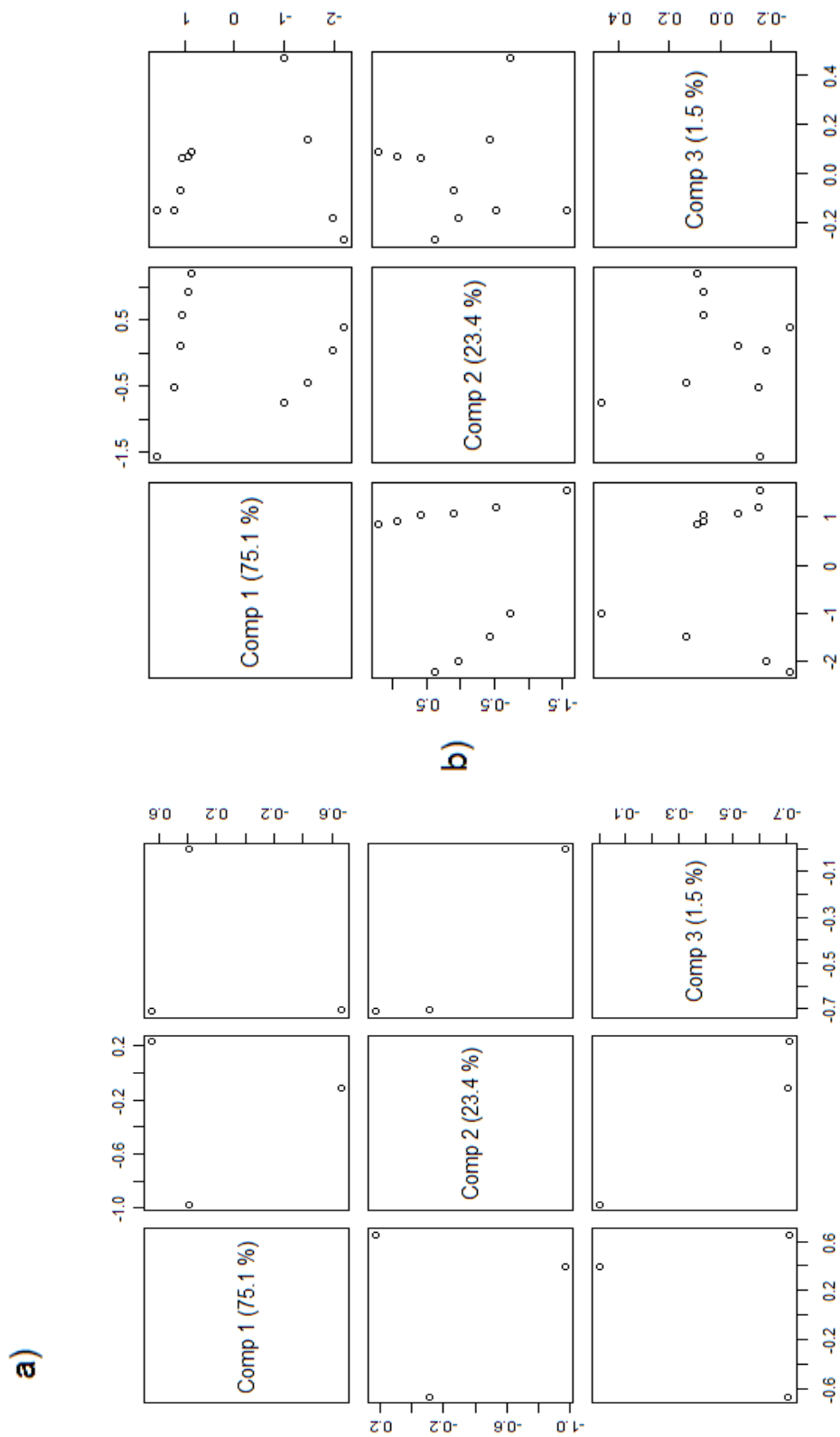
Tabla 11: MSEP , RMSEP y R <sup>2</sup> usando LOOCV (10 segmentos)				
Variable	Intersección	t <sub>1</sub>	t <sub>1</sub> , t <sub>2</sub>	t <sub>1</sub> , t <sub>2</sub> , t <sub>3</sub>
SCPVC	0.5918	0.1029	0.0382	0.0436
RSCPVC	0.7693	0.3208	0.1955	0.2089
R <sup>2</sup>		0.7853	0.9202	0.9090



Gráfica 1: Este tipo de graficas sirve para determinar el número óptimo de variables latentes a retener por conseguir el mínimo del RSCPVC , al aplicar la regla “del codo”.

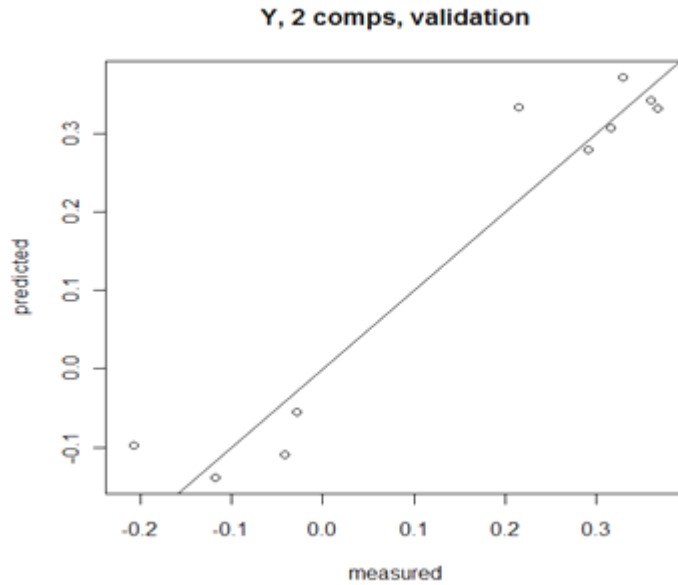
Evidentemente, las predicciones se mejoran considerando las dos primeras variables latentes, sin embargo corroboramos que la primera variable latente t<sub>1</sub>, rinde buena predicción de la respuesta.

## 4.7 Ejemplo de MCP1 en R



Gráfica 2: a) Proyección de las variables regresoras sobre el espacio definido por las variables latentes, por parejas (Matriz  $\mathbf{P}$ ) y, b) Proyección de las 10 observaciones sobre el espacio definido por las variables latentes por parejas (Matriz  $\mathbf{T}$ )

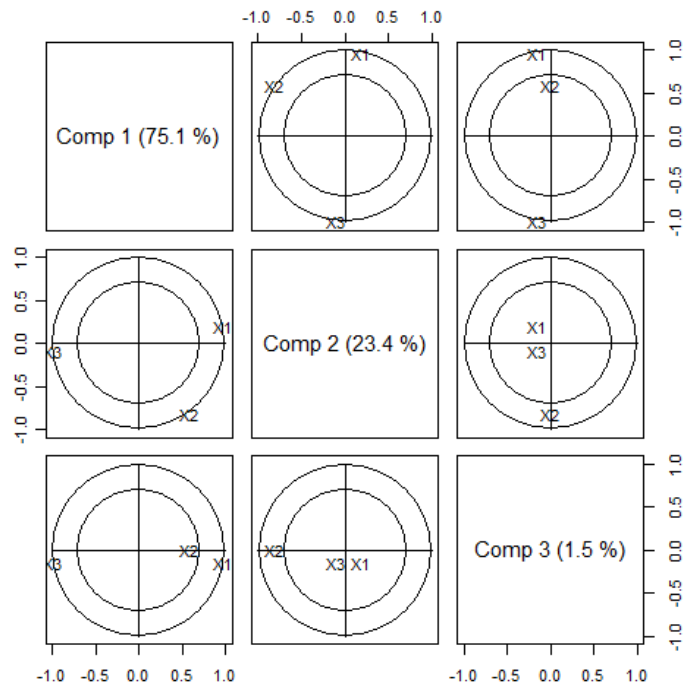
## Capítulo 4. Mínimos Cuadrados Parciales (MCP)



Gráfica 3 : Esta grafica muestra la similitud de las predicciones con las primeras dos variables latentes y los valores observados del conjunto de prueba.

De la gráfica 4 , note que las predicciones y los valores observados se encuentran relativamente cerca de la recta identidad, lo cual confirma que el ajuste a la respuesta es óptimo.

### Circulo de correlación de la nube de puntos-variables



Gráfica 4 : Circulo de correlaciones que muestran la correlación entre las variables regresoras originales y las variables latentes.

## 4.7 Ejemplo de MCP1 en R

Tabla 12: Varianza Explicada de $X$		
$t_1$	$t_2$	$t_3$
75.1	23.4	1.5
Acumulada	98.5	100

Tabla 13: Varianza Explicada de $y$		
$t_1$	$t_2$	$t_3$
90.1	7.3	0.8
Acumulada	97.3	98.1

El objetivo principal de aplicar MCP sobre un conjunto de datos, es obtener predicciones confiables con nuevas variables regresoras, que en número sean mucho menores que las originales. En este ejemplo, se observó que las predicciones para el porcentaje de conversión de  $n$ -heptano en acetileno, pueden realizarse utilizando como regresora la primera variable latente, con resultados confiables, esto se debe a que existe dependencia lineal entre la temperatura, la proporción molar de  $H_2$  y la temperatura con respecto a la conversión de  $n$ -heptano en acetileno; sin embargo, la mejor decisión corresponde a utilizar las dos primeras variables latentes como regresoras, obteniendo el óptimo ajuste.

## 4.8 Software

La técnica de MCP, necesita de cálculos sofisticados sobre todo si la base de datos a analizar es grande, implicando el uso de software que permita ejecutar el análisis con la mayor precisión posible.

Existen diversos paquetes comerciales que traen como opción la RMCP, por ejemplo, SAS/STAT (SAS 2001), SPSS (SPSS 16 en adelante, <spss.com>), STATISTICA (StatSoft 2001), Statgraphics (<www.statgraphics.com>), MatLab (Anderson and Bro 2000) o bien, como un agregado de Microsoft Excel spreadsheets (XL-Stat, <www.xlstat.com>). El paquete SIMCA 8.0 fue específicamente diseñado para el análisis de MCP; la paquetería libre que la ejecuta es R y SmartPLS.

# COMENTARIOS FINALES

El método de MCP fue desarrollado por Herman Wold como un método de optimización empírico, enfocado a las relaciones lineales que se presentan dentro de datos con ciertas características que hacen inapropiada la aplicación del método estándar de regresión, pues los supuestos con los cuales ésta trabaja, se violan. Con el paso del tiempo la popularidad de MCP se ha incrementado, pues los avances tecnológicos han abierto la brecha hacia nuevas áreas de investigación, con la necesidad imperiosa de modelar fenómenos que requieren técnicas menos restrictivas y más flexibles que la RLM; por ejemplo, en quimiometría, MCP es la herramienta estándar por excelencia para la predicción. Este buen funcionamiento, ha traído consigo el interés de mostrar a MCP como un método matemático riguroso más que un simple algoritmo; encontrando la equivalencia matemática de MCP con el método de Bidiagonalización de Lanczos. Por otra parte, la literatura que expone a MCP es bastante diversa, debido a la existencia de un gran número de variantes del algoritmo, lo cual dificulta el entendimiento de los principios en los cuales está basado; sin embargo los resultados de predicción que ofrecen, son los mismos. La falta de teoría para intervalos de confianza de los parámetros estimados por MCP, así como para las predicciones, caracterizan a MCP como un método estadístico no óptimo; aunque las estimaciones puntuales son consideradas óptimas, catalogando a MCP como una técnica predictiva más que interpretativa. La interpretación de las variables obtenidas por MCP es compleja, de hecho, la información al respecto es limitada; confirmando que el objetivo es la predicción. En la actualidad, existen numerosos paquetes de cómputo que permiten ejecutar MCP de manera eficaz, otorgándole mayor popularidad entre la comunidad científica.

Por todo esto, MCP puede ser implementado como herramienta estándar para realizar predicciones con la seguridad de que otorga resultados confiables, a pesar de ser de difícil entendimiento e interpretación, comparado con los modelos del ACP y de la RLM.

# ANEXO A

## I. Notación

**Definición:** Una matriz  $\mathbf{X}$  es un arreglo de números o variables  $x_{ij} \in \mathfrak{R}$ , construida mediante columnas y renglones, está dada por:

$$\mathbf{X} = (x_{ij}) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

Con  $i = 1, \dots, n$  y  $j = 1, \dots, p$ ; si  $n = p$  se dice que  $\mathbf{X}$  es una matriz cuadrada, de lo contrario se le conoce como una matriz rectangular.

Un vector columna es un arreglo dado por:

$$\underline{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{pmatrix}$$

La notación  $\underline{x}_i' = (x_{1i} \ x_{2i} \ \cdots \ x_{pi})$  indica el transpuesto de  $\underline{x}_i$  (vector renglón); en caso de una matriz  $\mathbf{X}$ , se tiene que  $\mathbf{X}' = (x_{ji})$ .

## II. Definiciones básicas

**Definición:** La diagonal de una matriz  $\mathbf{X}_{p \times p}$  comprende los elementos de su diagonal, dados por:

$$\text{diag}(\mathbf{X}) = \text{diag}(x_{11} \ x_{22} \ \cdots \ x_{pp}) = (x_{ii})$$

**Definición:** La traza de una matriz  $\mathbf{X}_{p \times p}$ , está definida como la suma de los elementos de su diagonal, esto es:

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^p a_{ii}$$

**Definición:** Una matriz es simétrica si se cumple que  $\mathbf{X} = \mathbf{X}'$ .

**Definición:** Una matriz simétrica es llamada definida positiva si  $\underline{v}'\underline{X}\underline{v} > 0$  para todos los posibles vectores  $\underline{v} \neq 0$ . Si se cumple que  $\underline{v}'\underline{X}\underline{v} \geq 0$ , entonces a  $\underline{X}$  es una matriz semidefinida positiva ( $\underline{v} \neq 0$ ). Los elementos de la diagonal de una matriz definida positiva son positivos; similarmente los elementos de la diagonal de una matriz semidefinida positiva son no negativos.

### III. Independencia lineal

Se dice que un conjunto de vectores  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  es linealmente dependiente si para las cualesquiera constantes  $\{c_1, c_2, \dots, c_n\} \in \mathfrak{R}$ , se tiene que:

$$c_1\underline{x}_1 + c_2\underline{x}_2 + \dots + c_n\underline{x}_n = \underline{0}$$

Para alguna  $c_i \neq 0$ . Si lo anterior no se cumple, se dice que el conjunto de vectores es linealmente independiente.

### IV. Ortogonalidad

**Definición:** Se dice que dos vectores  $\underline{a}$  y  $\underline{b}$  son ortogonales sí:

$$\underline{a}'\underline{b} = a_1b_1 + a_2b_2 + \dots + a_pb_p = 0$$

**Definición:** Un conjunto de vectores  $\{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n\}$  se dice conjunto ortogonal si se cumple que:

$$\underline{x}_i'\underline{x}_j = 0 \quad \forall i \neq j \text{ e } i, j = 1, \dots, n$$

Al conjunto  $\{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n\}$  se le llama conjunto ortonormal, si sus elementos son ortogonales y tienen norma unitaria.

**Definición:** Una matriz  $\underline{X}$  se dice matriz ortogonal si es cuadrada y las columnas de  $\underline{X}$  forman un conjunto ortogonal. Si  $\underline{X}$  es ortogonal entonces  $\underline{X}' = \underline{X}^{-1}$ .

Por tanto se tiene que:  $\underline{X}'\underline{X} = \underline{X}^{-1}\underline{X} = \underline{I}$

**Definición:** Una matriz  $\underline{X}$  se dice matriz ortonormal si es cuadrada y sus columnas forman un conjunto ortonormal.

### V. Rango

**Definición:** El rango de una matriz  $\underline{X}$  está definido como:

Rango ( $\mathbf{X}$ ) = Número de renglones linealmente independientes de  $\mathbf{X}$

Se puede probar que el número de vectores columna linealmente independientes es igual al número de renglones linealmente independientes de  $\mathbf{X}$ .

## VI. Inversa

**Definición:** Si una matriz  $\mathbf{X}$  es cuadrada y de rango completo, entonces se dice que  $\mathbf{X}$  es no singular, y que tiene una inversa única, denotada por  $\mathbf{X}^{-1}$  tal que:

$$\mathbf{X}\mathbf{X}^{-1} = \mathbf{I} \text{ y } \mathbf{X}^{-1}\mathbf{X} = \mathbf{I}$$

Además, por la definición de inversa se tiene que, para una matriz ortogonal  $\mathbf{X}'$ :

$$\mathbf{X}' = \mathbf{X}^{-1}$$

## VII. Valores y vectores propios

**Definición:** Dada una matriz  $\mathbf{X}_{p \times p}$ , se denomina a  $\underline{x} \in \mathfrak{R}^p$  con  $x \neq 0$  un vector propio de  $\mathbf{X}$ , si se cumple  $\mathbf{X}\underline{v} = \lambda\underline{v}$  con  $\lambda \in \mathfrak{R}$  un escalar, llamado valor propio.

**Teorema:** Si  $\mathbf{X}_{p \times p}$  tiene  $\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_k$  valores propios asociados a los vectores propios  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_k$ , entonces  $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_k\}$  forma un conjunto ortogonal.

**Definición:** Una matriz  $\mathbf{X}_{p \times p}$ , es equivalente con la matriz  $\mathbf{B}_{p \times p}$  si existe  $\mathbf{E}^{-1}$  tal que

$$\mathbf{A} = \mathbf{E}^{-1}\mathbf{B}\mathbf{E}.$$

Nota: Una de las características de las matrices equivalentes es que tienen los mismos vectores propios.

**Definición:** Una matriz  $\mathbf{X}_{p \times p}$  es diagonalizable si es equivalente con una matriz diagonal  $\mathbf{B}_{p \times p}$ , es decir  $\mathbf{X} = \mathbf{E}^{-1}\mathbf{B}\mathbf{E}$ .

**Definición:** Al conjunto de vectores aleatorios  $\{x_1, x_2, \dots, x_p\}$  se le denomina muestra aleatoria de tamaño  $p$ , si y solo si, éstos son independientes e idénticamente distribuidos.

**Teorema:** Sea una matriz  $\mathbf{A}_{m \times p}$  de rango  $r$ , entonces existen dos matrices ortogonales  $\mathbf{U}$  y  $\mathbf{V}$  tal que:

$$\mathbf{U}'\mathbf{A}\mathbf{V} = \mathbf{D}. \quad (\text{i})$$

A la expresión (i), se le conoce como la descomposición en valores singulares de  $\mathbf{A}$ .



Donde:  $\mathbf{D}$  es una matriz diagonal de orden  $r \times r$ , cuyas entradas  $d_i$  son las raíces cuadradas de los  $r$  valores propios de la matriz  $\mathbf{A}'\mathbf{A}$ , a los  $d_i$  se les denomina los valores singulares de la matriz  $\mathbf{A}$  y estos están en orden decreciente.  $\mathbf{U}$  es una matriz ortogonal de orden  $m \times r$ , cuyas columnas contienen a los vectores propios de la matriz  $\mathbf{A}\mathbf{A}'$ , llamados vectores singulares izquierdos de  $\mathbf{A}$ . Así mismo,  $\mathbf{V}$  es una matriz ortogonal de orden  $r \times p$  y sus columnas contienen los vectores propios de la matriz  $\mathbf{A}'\mathbf{A}$ , llamados los vectores singulares derechos de  $\mathbf{A}$ .

Multiplicando por la izquierda por  $\mathbf{U}$  y por la derecha por  $\mathbf{V}'$  a (i), obtenemos que:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}' \quad (\text{ii})$$

Mejor conocida como descomposición en sus valores singulares de  $\mathbf{A}$ .

### VIII. Vectores aleatorios

Sea  $(\Omega, \mathcal{F}, \mathbf{P})$  un espacio de probabilidad.

**Definición:** Un vector aleatorio es una función

$$\mathbf{X} : (x_1, x_2, \dots, x_p) : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (\mathcal{R}^p, \mathbf{B}(\mathcal{R}^p))$$

Tal que  $\forall \mathbf{A} \in \mathbf{B}(\mathcal{R}^p)$ ,  $\mathbf{X}^{-1}(\mathbf{A}) \in \mathcal{F}$  con  $\mathbf{X}^{-1}$  la imagen inversa de la función.

**Definición:** A la matriz  $\mathbf{X}_{n \times p}$ , conformada por los vectores aleatorios  $\{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p\}$  como renglones, se le denomina matriz de datos.

## ANEXO B

En esta tesis, se utilizó el paquete estadístico R versión 2.11.0. Para ejecutar un análisis por MCP1, primero es necesario cargar el paquete `pls` en R, como sigue:

```
library(pls)
```

Una vez cargado el paquete `pls`, se debe ajustar la función `plsrf`, para iniciar un análisis por MCP; esta función puede ejecutar diferentes métodos para implementar diversas versiones de MCP; en esta tesis se utilizó el método `oscorespls`, el cual da un análisis por MCP con el conjunto de variables latentes ortogonal. Se debe tener en cuenta que, si desean estandarizar las variables originales, la función `plsrf` solo centra las variables más no las estandariza, entonces la opción es introducir los datos estandarizados o bien, escalados a varianza unitaria. Si la estandarización es redundante, se introducen los datos con los valores originales.

Los datos de nuestro ejemplo se introducen en R, de la siguiente manera:

```
X<-matrix(c( 1300, 1300, 1300, 1300, 1300, 1300, 1200, 1200,
1200, 1200, 1200, 1200, 1100, 1100, 1100, 1100, 7.5, 9, 11,
13.5, 17, 23, 5.3, 7.5, 11, 13.5, 17, 23, 5.3, 7.5, 11, 17,
.0120, .0120, .0115, .0130, .0135, .0120, .0400, .0380, .0320,
.0260, 0.0340, .0410, .0840, .0980, .0920, .0860), ncol=3,
nrow=16)
```

```
Y<-matrix(c(49, 50.2, 50.5, 48.5, 47.5, 44.5, 28, 31.5, 34.5,
35.0, 38, 38.5, 15, 17, 20.5, 29.5), ncol=1, nrow=16)
Y
```

Las variables regresoras están en unidades de medida diferentes, por lo cual requieren ser estandarizadas previamente al análisis de MCP1, como sigue:

```
x1<-X[,1]-mean(X[,1])
x2<-X[,2]-mean(X[,2])
```

## ANEXO B

```
x3<-X[,3]-mean(X[,3])
z<-matrix(c(0,0,0,0,0,0,0,0,0,0),ncol=3, nrow=3)
diag(z)<-diag(var(X))
z<-sqrt(z)
X<-matrix(c(X1,X2,X3), ncol=3,nrow=16)
X<-X%%solve(z)
X

y<-Y[,1]-mean(Y[,1])
b<-(1/15)*t(Y)%%Y
Y<-a*(1/sqrt(b))
Y
```

Las funciones del paquete `pls`, trabajan con arreglos `dataframes`. Utilizando la función `data.frame` es posible crear estos arreglos, en nuestro ejemplo se hizo como sigue:

```
misdatos<-data.frame(I(Y), I(X))
```

Se dividieron los datos en dos partes: la de entrenamiento y la de prueba, mediante los siguientes `dataframes`:

```
datosEntrenamiento<-misdatos[1:10,]
datosEntrenamiento
datosPrueba<-misdatos[10:16,]
datosPrueba
```

El ajuste de la función `pls` necesita de los siguientes argumentos para ejecutar la técnica de MCP1: la *fórmula* especifica quienes son las variables dependientes (lado izquierdo de la fórmula) y quienes las independientes (lado derecho de la fórmula), separándolas por una tilde (~); `ncomp` especifica el número de variables latentes a calcular, `data` se refiere al `dataframe` que contiene los datos, `validation` se refiere al tipo de validación cruzada que será implementada para la validación del modelo por MCP; `scale` se refiere a la opción de centrar las variables o no, entre otros.

En nuestro ejemplo, el ajuste de la función `pls` estuvo dado por:

```
plsresultados<-pls(Y~X, ncomp=3, data=datosEntrenamiento,
validation = "LOO", method="oscorespls", scale=TRUE)
```

## ANEXO B

Los resultados del ajuste de la función `plsr`, pueden ser visualizados mediante las siguientes instrucciones:

```
summary(plsresultados)
plsresultados$scores
plsresultados$loadings
plsresultados$residuals
explvar(plsresultados)
```

La manera explícita de obtener el RSCR<sub>P</sub> de validación cruzada (RMSEP<sup>46</sup>, por sus siglas en inglés), SCR<sub>P</sub> (MSEP<sup>47</sup>, por sus siglas en inglés), y el coeficiente de determinación  $R^2$ , derivado de un ajuste `plsr` es:

```
RMSEP(plsresultados)
MSEP(plsresultados)
R2(plsresultados)
```

El modelo de RMCP es utilizado para predecir respuestas futuras; se utilizó el conjunto de prueba para la predicción de la respuesta (observación 11 a la 16), mediante la función:

```
predict(plsresultados, ncomp=1:3, newdata=datosPrueba)
```

los resultados de la predicción deben ser regresados a la escala original, esto es, sumar las medias muestrales de las variables del conjunto de entrenamiento, y multiplicar por las desviaciones estándar de las mismas, con la finalidad de dar contexto a los resultados.

Las gráficas expuestas en la sección del ejemplo, estuvieron dadas por las siguientes instrucciones:

```
Gráfica 1: plot(plsresultados, plottype="coef", ncomp=1:3,
asp=1, legendpos="topright")
```

```
Gráfica 2: plot(RMSEP(plsresultados), legend="topright")
```

---

<sup>46</sup> Root Mean Squares Error Prediction, RMSEP

<sup>47</sup> Mean Squares Error Prediction, MSEP

## ANEXO B

Gráfica 3: `plot(plsresultados, plottype="scores",  
comps=1:3)`

Gráfica 4: `plot(plsresultados, ncomp=2, asp=1, line=TRUE)`

Gráfica 5: `plot(plsresultados, plottype="correlation",  
comps=1:3)`

Gráfica 6: `plot(RMSEP(plsresultados), main="Predicciones",  
newdata=datosPrueba)`

## REFERENCIAS BIBLIOGRÁFICAS

1. Abdi, H. (2010), "*Partial Least Squares Regression and Projection on Latent Structure Regression (PLS Regression)*". John Wiley & Sons, Inc
2. Abdi, H. & Williams, J. L. (2010), "*Principal Component Analysis*". Wiley Interdisciplinary Reviews: Computational Statistics, 2.
3. Abdi, H. (2003), "*Least-Squares*". Encyclopedia for Research Methods for the Social Sciences. Thousand Oaks (CA): 559-561.
4. Alciaturi, C. E.; Escobar, M. E.; De La Cruz, C. & Rincón, C. (2003), "*Partial Least Squares (PLS) Regression and its Application to Coal Analysis*". Revista Técnica de la Facultad de Ingeniería, Universidad de Zulia. Vol. 26 (3):197-204.
5. Benzi, M. & Tuma, M. (2003), "*A Robust Incomplete Factorization Preconditioner for Positive Definite Matrices*". Numerical Linear Algebra with Applications, Vol. 10: 385-400.
6. Björkström, A. & Sundberg, R. (2007), "*A Two-Parametric Class of Predictors in Multivariate Regression*". Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
7. Boulesteix, A. L. & Strimmer, K. (2006). "*Partial Least Squares: A Versatile Tool for Analysis of High-Dimensional Genomic Data*". Briefings in Bioinformatics, Vol. 8 (1)1: 32- 44.
8. Butler, N. A. & Denham, M. C. (2000), "*The Peculiar Shrinkage Properties of Partial Least Squares Regression*". Journal of the Royal Statistical Society Series B. Vol. 62 (2):585-593.
9. Carrascal, M. L, Galván, I. & Gordo, O. (2009), "*Partial Least Squares Regression as an Alternative to Current Regression Methods Used in Ecology*". Oikos 118: 681-690, Madrid, España.

10. Chun, H. & Keles, S. (2010), "*Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection*". Journal of the Royal Statistical Society B, 72(1):3-25.
11. Denham, M.C. (1991), "*Calibration in infrared spectroscopy*". Ph.D. dissertation, Univ. Liverpool.
12. Eldén, L. (2004), "*Partial Least-Squares vs. Lanczos Bidiagonalization-I: Analysis of a Projection Method for Multiple Regression*". Computational Statistics & Data Analysis, Vol. 46: 11-31.
13. Ergon, R. (2009), "*Re-interpretation de NIPALS Results Solves PLSR Inconsistency Problem*". Journal of Chemometrics; Vol. 23 (1): 72-75.
14. Ergon, R. (2002), "*Noise Handling Capabilities of Multivariate Calibration Methods*". Modeling, Identification and Control, Vol. 23 (4): 259-273.
15. Faber, K. (2005), "*Multivariate Prediction Uncertainty*". Homepage of Chemometrics.
16. Faber, K. & Kowalski, B.R. (1997), "*Propagation of Measurement Errors for the Validation of Predictions Obtained by Principal Component Regression and Partial Least Squares*". Journal of Chemometrics, Vol. 11: 181-238.
17. Ferré, B. J. (1997). "*Experimental Design Applied to the Selection of Samples and Sensors in Multivariate Calibration*". Tesis Doctoral, Universitat Rovira I Virgili.
18. Frank, L. & Friedman, J. (1993), "*A Statistical View of Some Chemometrics Regression Tools*", Technometrics, Vol. 35/2: 109-135.
19. Garthwaite, P. H. (1994), "*An Interpretation of Partial Least Squares*". Journal of the American Statistical Association, Vol. 89 (425): 122-127.
20. Geladi, P. & Kowalski B. R. (1986), "*Partial Least-Squares Regression: A Tutorial*". Analytica Chimica Acta 185: 1-17.
21. Goutis, C. (1996), "*Partial Least Squares Algorithm Yields Shrinkage Estimators*". The Annals of Statistics, Vol. 24 (2): 816-824.

22. Graber, S.; Czellar, S. & Jean-Emile, D. (2002), "*Using Partial Least Squares Regression in Marketing Reserch*". University of Geneva.
23. Grupo de Quimiometría & Cualimetría de Tarragona, "*Quimiometría, Una Disciplina Útil para el Análisis Químico*". Universitat Rovira i Virgili. Tarragona, España.
24. Helland, I. S. (2001), "*Some theoretical aspects of partial least square regression*". Chemometrics. Intell. Lab. Syst: Vol. 58 No. 2: 97-107.
25. Helland, I.S. (2000), "*Model Reduction for Prediction in Regression Models*". Scandinavian Journal of Statistics, Vol. 27: 1-20.
26. Helland, I.S. (1988), "*On the Structure of Partial Least Squares Regression*". Communications in Statistics, Simulation and Computation 17 (2): 581-607.
27. Hestenes, M. R. & Stiefel, E. (1952), "*Methods of Conjugate Gradients for Solving Linear Systems*". Journal of Research of the National Bureau of Standards, Vol. 49 (6): 409-438.
28. Hoegaerts, L.; Suykens J.A.K.; Vandewalle, J. & De Moor, B. (2003), "*Kernel PLS Variants for Regression*". KU Leuven, Dept. of Electrical Engineering ESAT-SCD-SISTA, Kasteelpark Arenberg 10, B-3001.
29. Höskuldsson, A. (September, 2004), "*Prediction in Linear Models*". Homepage of Chemometrics.
30. Höskuldsson, A. (June, 2004), "*PLS regression and the Covariance*". Homepage of Chemometrics.
31. Höskuldsson, A. (March, 2004), "*Centring and Scaling of Data*". Homepage of Chemometrics.
32. Höskuldsson, A. (November, 2003), "*Selection of Subsets of Variables in Linear Regression*". Homepage of Chemometrics.
33. Jørgensen, B. & Goegebeur, Y. (2007), "*Module 7: Partial Least Squares Regression I*". Department of Statistics, ST02: Multivariate Data Analysis and Chemometrics.



34. Kleinknecht, D. (June, 2002), "*A Statistician's View of the Single-Y Problem*". Homepage of Chemometrics.
35. Krämer, N. (2007), "*An Overview on the Shrinkage Properties of Partial Least Squares Regression*". Computational Statistics, Vol. 22 (2): 249-273.
36. Krämer, N.; Sugiyama, M. & Braun, M. (2009), "*Lanczos Approximations for the Speedup of Kernel Partial Least Squares Regression*". Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS): 272-279.
37. Lakshminarayanan, S.; Sirish, L. S. & Nandakumar. (1997), "*Modelling and Control of Multivariable Processes: The Dynamic Projection to Latent Structures Approach*". Department of Chemical Engineering, University of Alberta; Edmonton, Alberta, Canada T6G 2G6, AIChE Journal, Vol 43:2307-2323.
38. Lingjaerde, O. C. & Christophersen, N. (2000), "*Shrinkage Structure of Partial Least Squares*". Scandinavian Journal of Statistics, Vol. 27 (3): 459-473.
39. Maitra, S. & Yan, J. (2008), "*Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression*". Casualty Actuarial Society.
40. Merola, M. G & Abraham, B. (2001), "*Dimensionality Reduction Approach to Multivariate Prediction*". The Canadian Journal of Statistics, Vol. 29 (2): 191-200.
41. Mevik, B. & Wehrens, R. (2007), "*The pls Package: Principal Component and Partial Least Squares Regression in R*". Journal of Statistical Software: Vol 18 (2).
42. Miret, E. & Linares, G. "*El Teorema de Descomposición Singualr comoramienta Unificadora de Algunos Métodos de Analisis de Datos*". Revista Ciencias Matemáticas, Vol. 19(1). (2001)
43. Montgomery, D. C. & Peck, A. E. (1992), "*Introduction to Linear Regression Analysis*". Segunda edición, John Wiley & Sons, INC.
44. Mood, M. A. & Graybill, A. F. (1978), "*Introducción a la Teoría de la Estadística*". Adaptación a la 2da edición norteamericana por Rafael Pro

- Bermejo, profesor de la Escuela de Estadística de la Universidad de Madrid. Edición española, 4ta edición, España.
45. Nguyen, D. V & Roche, D. M. (2004), "*On Partial Least Squares Dimension Reduction for Microarray-Based Classification: A Simulation Study*". Computational Statistics & Data Analysis 46:407-425.
  46. Pell, R. J.; Ramos, L. S. & Manne, R. (2002), "*The Model Space in Partial Least Squares Regression*". Journal of Chemometrics. Vol. 21: 165-172.
  47. Porcel, Marta. (2001), "Aplicación de Técnicas Quimiométricas para el Desarrollo de Nuevos Métodos Cinético-Espectrofotométricos de Análisis". Tesis, Universitat Autònoma de Barcelona.
  48. Phatak, A. & De Hoog. (2002), "*Exploiting the Connection between PLS, Lanczos, and Conjugate Gradients: Alternative Proofs of Some Properties of PLS*". Journal of Chemometrics, Vol. 16: 361-367.
  49. Phatak, A. & Jong, S. D. (1997), "*The Geometry of Partial Least Squares*". Journal of Chemometrics 11: 311-338.
  50. Pirouz, D. M. (2006) "*An Overview of Partial Least Squares*". School of Business, University of California, Irvine.
  51. Ramos, P. (2006) "*Raman and X-Ray Fluorescence Spectroscopy Data Fusion for Identification of Pigments in Works of Art*". tesis Doctoral, Universitat Rovira I Virgili.
  52. Rao, C. R. & Mitra, S. K. (1971), "*Generalized Inverses of Matrices and Its Applications*". Wiley, New York.
  53. Rencher, A. C. (1995), "*Methods of Multivariate Analysis*". John Wiley & Sons, Inc. New York. Second edition.
  54. Risvik, H. (2007) "*Principal Component Analysis (PCA) & NIPALS Algorithm*".
  55. Roa, C. A. (2004), "*Métodos de Proyección Multivariada de Imágenes, Aplicados al Control Estadístico de Procesos*". Tesis de maestría, Universidad Simón Bolívar.

56. Rosipal, R. (2003), "*Kernel Partial Least Squares for Nonlinear Regression and Discrimination*". *Neural Network World*, 13 (3): 291-300.
57. Rosipal, R. & Krämer, N. (2006), "*Overview and Recent Advances in Partial Least Squares*". In C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, editors. "Subspace, Latent Structure and Feature Selection Techniques", *Lecture Notes in Computer Science*, 34-51.
58. Rosipal, R. & Trejo, J. L. (2001), "*Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space*". *Journal of Machine Learning Research* 2: 97-123.
59. Ryan, P. T. (1997), "*Modern Regression Methods*". A Wiley Interscience Publication". USA.
60. Seber, G. A. F. (1977), "*Linear Regression Analysis*". John Wiley & Sons.
61. Srivastava, M. S. & Carter, E. M. (1983), "*An Introduction to Applied Multivariate Statistics*". North Holland, New York.
62. Srivastava, M. S. & Khatri, C. G. (1979), "*An Introduction to Multivariate Statistics*". North Holland, New York.
63. Stoica, P. & Söderström, T. (1998), "*Partial Least Squares: A First-order Analysis*". *Scandinavian Journal of Statistics*, Vol. 25: 17-24.
64. Sundberg, R. (2002), "*Shrinkage Regression*". *Encyclopedia of Environmetrics*. Vol. 4:1994-1998.
65. Tenenhaus, M. (1998), "*La Régression PLS regression*". París: Technip.
66. Timm, N. H. (2002) "*Applied Multivariate Analysis*". Springer, New York.
67. Tobias, R.D. (1995), "*An Introduction to Partial Least Squares Regression*". SAS Institute Inc., Cary, NC.
68. Vega, J.C. (2004), "*Generalizaciones de Mínimos Cuadrados Parciales con Aplicación en Clasificación Supervisada*". Tesis para obtener el grado de Doctor, Universidad de Puerto Rico.

69. Wegelin, A. J. (2000), "*A Survey of Partial Least Squares (PLS) Methods, with Emphasis on the Two-Block Case*". TECHNICAL REPORT No. 371. Department of Statistics, University of Washington, Seattle, Washington, 98195 USA.
70. Wold, H. (1966), "*Estimation of Principal Components and Related Models by Iterative Least Squares*". In *Multivariate Analysis* (Ed., P.R. Krishnaiah), Academic Press, NY: 391-420.
71. Wold, S. (1995), "*Chemometrics; What Do We Mean With It, and What Do We Want From It?*". *Chemometrics and Intelligent Laboratory Systems* 30:109-115.
72. Yeniay, Ö & Göktas, A. (2002), "*A Comparison of Partial Least Squares Regression with Other Prediction Methods*". *Journal of Mathematics and Statistics*, Vol. 31: 99-111.