



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**DESARROLLO DE UNA METODOLOGÍA
PARA REPRESENTAR CARTERAS DE
INVERSIÓN MEDIANTE CONJUNTOS DE
CLASES DE ACTIVOS DE DISTINTA
DIMENSIONALIDAD EN UN MERCADO DE
POCAS TRANSACCIONES**

PAULA DONOSO Y.

Tesis para optar al grado de
Magíster en Ciencias de la Ingeniería

Profesor Supervisor:
GONZALO CORTAZAR S.

Santiago de Chile, Agosto, 2007



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA
Departamento de Ingeniería Industrial y de Sistemas

**DESARROLLO DE UNA METODOLOGÍA
PARA REPRESENTAR CARTERAS DE
INVERSIÓN MEDIANTE CONJUNTOS DE
CLASES DE ACTIVOS DE DISTINTA
DIMENSIONALIDAD EN UN MERCADO CON
POCAS TRANSACCIONES**

PAULA DONOSO Y.

Tesis presentada a la Comisión integrada por los profesores:

GONZALO CORTAZAR S.

JAIME CASASSUS V.

AUGUSTO CASTILLO R.

LUIS FERNANDO ALARCÓN C.

Para completar las exigencias del grado de
Magíster en Ciencias de la Ingeniería

Santiago de Chile, Agosto, 2007

A mis padres, que siempre
me apoyaron.

AGRADECIMIENTOS

En primer lugar quiero agradecer al profesor Gonzalo Cortázar por el tiempo dedicado y el constante apoyo a lo largo de la realización de esta tesis. Sus consejos y orientación me han permitido comprender los fundamentos de un trabajo de investigación y han sido sumamente importantes en mi formación profesional.

También quiero agradecer a mis compañeros y amigos de magíster: Claudia Salas y Felipe Severino por su valiosa ayuda y por los gratos momentos vividos durante la realización de este trabajo. Un reconocimiento especial para el profesor Jaime Casassus y para Carolina Mery (AFP Habitat) por su gran disposición e importante colaboración en el desarrollo de esta tesis y para todos los integrantes del FinLabUC y RiskAmerica, encabezados por Carlos Milla, por su importante ayuda y por el gran apoyo prestado durante la realización de este trabajo.

Se agradece también el apoyo financiero proporcionado por Fondef (Proyecto D03I1039), y Fondecyt (Proyecto 1070771)

Finalmente, quiero agradecer a mi familia y a Aldo, por su gran apoyo y porque siempre han estado conmigo.

INDICE GENERAL

	Pág.
DEDICATORIA	II
AGRADECIMIENTOS	III
INDICE DE TABLAS	VII
INDICE DE FIGURAS	IX
RESUMEN	XI
ABSTRACT	XII
I. INTRODUCCIÓN	1
II. MÉTODOS DE SELECCIÓN DE SUB-SETS ÓPTIMOS.....	5
II.1 Método de selección utilizando variables compuestas	6
II.1.1 Regresión con Componentes Principales.....	9
II.1.2 Estrategias para seleccionar Componentes en Regresiones	11
II.2 Métodos de selección utilizando variables puras	13
II.2.1 Introducción	13
II.2.2 Métodos Estadísticos	14
II.2.3 Análisis Heurísticos	26
III. PROBLEMA Y METODOLOGÍA PROPUESTA.....	34
III.1 Explicación del Problema	34
III.2 Discusión de la Propuesta Metodológica	36

III.2.1 Precios para representar la Cartera de Inversión	37
III.2.2 Diferenciación entre tipos de Índices	40
III.2.3 Función objetivo o criterio de ajuste de la representación	41
III.2.4 Selección de variables “Compuestas” o variables “Puras”	42
III.2.5 Algoritmo de Optimización	43
III.3 Descripción de la Metodología Propuesta.....	46
IV. EJEMPLO DEL USO DE LA METODOLOGÍA EN UNA CARTERA DE RENTA FIJA NACIONAL.....	50
IV.1 Descripción de Datos	50
IV.1.1 Mercado de Renta Fija en Chile	50
IV.1.2 Cartera de Saldos Nominales de Instrumentos de Renta Fija en Depósito Central de Valores DCV	54
IV.1.3 Índices de Renta Fija RA.....	56
IV.2 Resultados.....	61
IV.2.1 Pruebas realizadas	61
IV.2.2 Discusión de los Resultados.....	80
V. CONCLUSIONES GENERALES	83
BIBLIOGRAFIA.....	85
ANEXOS.....	89
ANEXO A: CARACTERÍSTICAS DE LOS INSTRUMENTOS DEL MERCADO DE RENTA FIJA EN CHILE.....	90
A.1 Bonos de Gobierno.....	90

A.1.1 Pagarés descontables del Banco Central de Chile (PDBC).....	90
A.1.2 Pagarés reajustables del Banco Central de Chile con pago en cupones (PRC).....	91
A.1.3 Cupones de emisión reajustables opcionales (CERO) en UF.....	92
A.1.4 Bonos del Banco Central de Chile en pesos y en UF (BCP) (BCU).....	93
A.1.5 Bonos de la Tesorería General de la República en UF (BTU).....	94
A.2 Bono Corporativo	95
A.3 Depósito a plazo	97
A.4 Letras Hipotecarias.....	99
A.5 Bonos de Reconocimiento	101
ANEXO B: COMPOSICIÓN CARTRERA DCV Y ASIGNACIÓN ACTIVOS	103
ANEXO C: COMPOSICIÓN COMPONENTES PRINCIPALES.....	108
ANEXO D: REGRESIÓN CON COMPONENTES PRINCIPALES	109

INDICE DE TABLAS

	Pág.
Tabla IV- 1: Descripción Índices Renta Fija Nacional Seleccionados	59
Tabla IV- 2: % de la Varianza explicada por los componentes principales.....	63
Tabla IV- 3: 10 PCs con mayor estadístico t.....	64
Tabla IV- 4: Ajuste para k=3	69
Tabla IV- 5: Análisis estadístico Series In-sample	71
Tabla IV- 6: Análisis estadístico Series Out-of-sample	72
Tabla IV- 7: Ajuste para k=5	72
Tabla IV- 8: Análisis estadístico Series In sample.....	74
Tabla IV- 9: Análisis estadístico Series Out of sample.....	75
Tabla IV- 10: Ajuste para k=8	76
Tabla IV- 11: Análisis estadístico Series In-sample	77
Tabla IV-12: Análisis estadístico Series Out-of-sample.....	78
Tabla IV-13: Análisis estadístico Series Out-of-sample.....	79
Tabla B- 14: Información Cartera de Renta Fija del DCV y Asignación de activos	103
Tabla C- 15: Composición primeros 15 componentes Principales	108
Tabla D- 16: Coeficientes y estadístico t de la regresión entre la cartera Pre-Mapea y los PCs.....	109

Tabla D- 17: PCs ordenados de forma decreciente según valor absoluto del estadístico t.....	110
---	-----

INDICE DE FIGURAS

	Pág.
Figura 2- 1: Esquema selección de 3 variables de un universo de 6.....	6
Figura 2- 2: Métodos para selección del mejor subset.....	7
Figura 2- 3: Gráfico Cp vs p	25
Figura 4- 1: Composición Cartera Saldos Nominales RF DCV el 30/06/2006	55
Figura 4-2: Mapa Índices Renta Fija RA	58
Figura 4-3: Composición cartera con 3 PCs	65
Figura 4-4: Composición cartera con 5 PCs	66
Figura 4-5: Composición cartera con 8 PCs	67
Figura 4-6: N° de iteraciones vs N° de índices.....	68
Figura 4-7: Composición cartera óptima para k=3	70
Figura 4-8: Series In-sample	71
Figura 4-9 : Series Out-of-sample.....	71
Figura 4-10: Composición cartera óptima para k=5	73
Figura 4-11: Series In sample	74
Figura 4-12: Series Out of sample	75
Figura 4-13: Composición cartera óptima para k=8	76
Figura 4-14: Series In-sample	77

Figura 4-15: Series Out-of-sample..... 78

RESUMEN

La motivación de esta investigación es poder representar o *mapear* de la mejor forma posible una cartera de inversión. El proceso de *mapeo* de carteras es un proceso fundamental, que hasta hoy se realiza de manera intuitiva en la mayoría de las instituciones financieras del país y que parece no haber sido abordado formalmente en la literatura financiera.

El problema consiste de 2 etapas, primero cómo representar una cartera de inversión compuesta de m activos mediante n clases de activos con $m \geq n$, para posteriormente representar estas n clases mediante un conjunto de k clases de activos.

Esta tesis investiga distintos métodos de selección de un subset óptimo de k variables, a partir de un conjunto inicial de n , donde $n > k$. En primer lugar, se presenta un análisis detallado de las distintas metodologías desarrolladas por diversos autores con el fin de seleccionar un subset óptimo de la manera más eficiente posible.

Dentro de los métodos señalados, se pueden mencionar aquellos que seleccionan variables generadas en base a los datos originales como el método de componentes principales, y los métodos que seleccionan variables dentro del conjunto inicial, como los métodos estadísticos. Los componentes principales en regresiones entregan de forma eficiente respuestas muy satisfactorias cuando se seleccionan por criterios de ajuste, pero presentan el inconveniente de entregar combinaciones de variables, en lugar de las mismas por separado, lo cual dificulta algunos de los análisis posteriores. Los métodos estadísticos se caracterizan por entregar respuestas de buena calidad y además son sencillos en su implementación y comprensión.

En segundo lugar, se presenta la metodología propuesta para la resolución final del problema planteado, así como los alcances de su implementación.

Finalmente, se ilustra el uso de la metodología aplicándola a una cartera del mercado de Renta Fija Nacional, representada por los saldos nominales en las custodias del Depósito Central de Valores (DCV) de Chile y se plantean las conclusiones principales.

ABSTRACT

The motivation of this investigation is to propose a methodology to represent or to map in the best possible way an investment portfolio. This process is fundamental, and until today it is made intuitively most financial institutions and seems not to have been analyzed in the literature.

The problem consists of 2 stages, first how to represent an investment portfolio composed of m assets through n asset classes with $m \geq n$, and second to represent these n asset classes through by a subset of k asset classes.

This thesis reviews different k -subset selection methods, from an initial n -variable set, where $n > k$. In the first place, a detailed analysis of the different existing methodologies to solve the problem is described.

These methods can be classified into those that select variables generated on the basis of the original data like the method of principal components, and the methods that select variables within the initial set, like the statistical methods. Regressions with principal components give very satisfactory answers when the components are chosen by the adjustment criteria, but have the disadvantage of providing a combination of variables, instead of a group of single variables, which difficult some analysis. The statistical methods tend to give answers of good quality and they are simple in their implementation and understanding.

Secondly, the proposed methodology for the final resolution is developed along with a discussion on how to implement it.

Finally, an empirical implementation of the proposed methodology is presented, applying it to the Fixed Income Market portfolio of the Depósito Central de Valores (DCV) of Chile.

I. INTRODUCCIÓN

Dentro de los mercados financieros resulta de vital importancia que sus actores constantemente revisen sus inversiones, tanto para analizar el comportamiento de sus propias carteras como el que siguen el resto de los participantes del mercado.

Muchos autores han señalado que herramientas como asset allocation (Markowitz 1959), análisis de riesgo (J.P. Morgan Bank, 1995; Albrecht, 2003; Biglova et al., 2004; Satchell y Hwang, 2001) y performance (Sharpe 1994; Gupta et al., 1999) son la base del análisis de inversión de cualquier institución financiera.

Con el fin de llevar a cabo estos análisis, es necesario representar de la mejor manera posible cualquier cartera de inversión. El problema de la representación de un portafolio de inversión mediante un subconjunto de clases de activos, es enfrentado diariamente dentro del mundo financiero, siendo informalmente denominado como proceso de “Mapeo de carteras”.

Sin embargo, debido al gran número de activos que componen los portafolios de inversión, es de gran importancia, ser capaz de representar los retornos de un portafolio, a través de un subconjunto de clases de activos, de dimensión inferior al universo total. Esta reducción resulta fundamental, ya que permite una disminución considerable en la complejidad de los modelos, permitiendo realizar análisis que serían impensables debido a la dificultad numérica si se consideraran todos los activos.

Ahora bien, la complejidad del problema anterior radica básicamente en 3 puntos.

Primero, la dimensionalidad de la representación puede ser variable de acuerdo al uso que se le va a dar. Si hablamos por ejemplo de asset allocation, podríamos necesitar representar la cartera global considerando sólo unas cuantas regiones geográficas por lo que, el número final de clases requeridas sería muy inferior al requerido en el caso de un análisis de riesgo, donde probablemente nos interesaría modelar la cartera de una forma más detallada, considerando características propias de los activos además de su distribución geográfica.

Segundo, la dificultad en la selección de los datos a utilizar para la representación. La tendencia natural, sería utilizar los precios históricos que han tenido tanto la cartera de inversión como los activos a lo largo del tiempo.

Sin embargo, el uso de los precios históricos presenta varios problemas.

- a) La cartera de inversión es activa (cambia sus posiciones en el tiempo)
- b) Los papeles que componen la cartera cambian sus características (particularidad de los instrumentos de renta fija)
- c) Falta de precios

Los dos primeros, son propios de cualquier mercado de inversión, independiente de la profundidad de éste, y el último es una característica de mercados emergentes, como el mercado nacional, en el que debido a la falta de transacciones e iliquidez de los papeles, no es posible obtener series de precios que representen en forma fidedigna las inversiones de cada institución.

Para solucionar estos problemas, es posible representar cada activo de la cartera mediante una serie de índices, los cuales se definen como un activo o portafolio de activos que poseen la característica de ser estables en su definición y completos en precios. A través de estos índices es posible generar una serie que represente el comportamiento de esa cartera fija a lo largo del tiempo.

Finalmente, debido a la gran cantidad de índices pasivos disponibles en el mercado, es necesaria la capacidad de selección de un subset óptimo del conjunto inicial de índices que represente de la mejor forma posible los retornos de un portafolio.

Aún cuando el problema parece no haberse abordado formalmente en la literatura académica, existen trabajos que tratan de resolver problemas similares al planteado. Sharpe (1992) desarrolla una metodología denominada “Return Based Style Analysis” en la cual mediante una regresión modificada es posible encontrar qué combinaciones de índices pasivos formados por activos financieros mejor explica el retorno de un fondo a lo largo del tiempo. Sin embargo, el principal objetivo de esta metodología es reconocer

los sectores de inversión del fondo, cuando no se posee ninguna información acerca de la cartera de inversión. Además, el método no permite seleccionar un número específico de índices pasivos con los que representar la cartera de inversión, no permitiendo solucionar el problema de la dimensionalidad de representación variable.

Por otra parte, existen trabajos en el área de Index Tracking, nombre que describe el proceso de seguir (o reproducir) el comportamiento de un índice. En esta área de investigación, existe una similitud entre aquellos métodos que seleccionan un subconjunto de activos que repliquen de mejor forma el retorno del índice, de un universo total, con un número determinado de activos y el problema aquí planteado. Sin embargo, en muchos casos el objetivo es minimizar los costos de transacción, más que buscar exactitud en la representación

Dentro de la línea en que es posible generar analogías entre ambos problemas, destacan los trabajos de optimización por pasos, como el presentado por Jansen y Van Dijk (2002), que consiste en una optimización secuencial basada en la mejora de un método básico denominado *sequential quadratic*, consistente en la optimización por pasos eliminando en cada iteración los activos con menor ponderación.

Como no es posible encontrar literatura que aborde el mismo problema, el enfoque se centra en aquellas áreas que resuelven el problema de cómo determinar el mejor subconjunto de k variables de un set que contiene n variables. Existen en la literatura varias formas potenciales de resolver este problema.

Una rama de la literatura está compuesta por los métodos estadísticos. A grandes rasgos, el foco de los estudios de búsqueda de subconjunto óptimos, han sido las metodologías de selección de subconjuntos, el criterio de selección, o una combinación de ambas.

Textos estándar como Draper y Smith (1981) y Montgomery y Peck (1991) proveen una buena descripción de estas metodologías y criterios.

Una segunda rama está compuesta por los llamados análisis heurísticos, concepto generado por el matemático George Polya, (1957). Dentro de los análisis heurísticos destaca procedimientos de búsqueda tales como el Simulated Annealing (S.

Kirkpatrick, C. D. Gelatt y M. P. Vecchi en 1983, y por V. Cerný en 1985) y el Tabú Search (Glover, 1966, 1969; Glover y Laguna, 1977)

El presente documento es uno de los primeros en abordar el problema del Mapeo de carteras de manera formal, desarrollando una metodología de selección simple que puede ser utilizada en un mercado con pocas transacciones (donde no hay precios diarios para todos los activos) basada en métodos estadísticos que permiten obtener resultados empíricos que significan un importante avance en el desarrollo de un proceso fundamental, que hasta hoy se realiza de manera informal e intuitiva en la mayoría de las instituciones financieras del país.

Además ilustra el uso de la metodología implementándola para representar la cartera de Renta Fija Nacional del Depósito Central de Valores (DCV) de Chile, utilizando para este fin índices de Renta Fija generados por RiskAmérica.

La Tesis está estructurada de la siguiente forma, el Capítulo 2 explica los distintos métodos presentes en la literatura para seleccionar un subset óptimo de variables. A continuación, el Capítulo 3 describe formalmente el problema que se desea abordar, para describir, a continuación, la metodología propuesta, explicando primero los distintos problemas que es necesario resolver para ser capaces de generar un algoritmo, para posteriormente formalizar su descripción. El Capítulo 4 presenta una ilustración del uso de la metodología, buscando, como se señaló anteriormente, representar la cartera Renta Fija Nacional del Depósito Central de Valores (DCV) de Chile, para finalmente extraer las principales conclusiones de la investigación en el Capítulo 5.

II. MÉTODOS DE SELECCIÓN DE SUB-SETS ÓPTIMOS

Existen en la literatura diversos métodos y heurísticas para determinar formas de encontrar un subset óptimo. La forma de seleccionar el subset varía según las necesidades y objetivos que el usuario persigue con la representación.

El objetivo de este Capítulo es presentar las diversas formas de selección presentes en la literatura, buscando enfocarse principalmente en aquellos métodos que satisfagan los requerimientos básicos preestablecidos en esta investigación, es decir, que cumplan con las características necesarias para permitir tanto flexibilidad, en términos de capacidad de satisfacer distintos requerimientos del usuario, como también, permitan generar una metodología sencilla, de fácil comprensión e implementación.

Como señala Draper y Smith (1981), el problema de selección del subset de variables, tiene como principal objetivo seleccionar un subconjunto mínimo de variables sin sacrificar el poder explicativo de estas.

En términos del criterio utilizado para reducir la dimensionalidad de la representación, es posible diferenciar entre aquellos métodos que reducen dimensionalidad mediante la generación de un nuevo conjunto de variables, de aquellos que reducen el espacio mediante la selección de variables dentro del universo inicial. Con el objeto de diferenciar entre el uso de variables originales y el uso de variables generadas, se designará a cada grupo de métodos con un nombre específico.

Al primer conjunto de métodos, aquellos que buscan reducir la dimensionalidad mediante la generación de un nuevo conjunto de variables, se les denominará “Métodos de selección utilizando variables compuestas”, ya que ellos utilizan alguna transformación de las variables originales. Al segundo conjunto de métodos, aquellos que buscan reducir mediante la selección de variables originales, se les denominará “Métodos de selección utilizando variables puras”, ya que seleccionan variables dentro del universo inicial.

A modo de ejemplo, la Figura 2-1 representa de forma esquemática un caso en que para un set de datos cualquiera compuesto de 6 variables (X_i), se desea seleccionar 3, donde para el caso de la metodología seleccionando variables puras, las variables representativas son: X_2 , X_3 y X_6 y para el caso de la metodología seleccionando variables compuestas, se generan las nuevas variables: Y_1 , Y_2 y Y_3

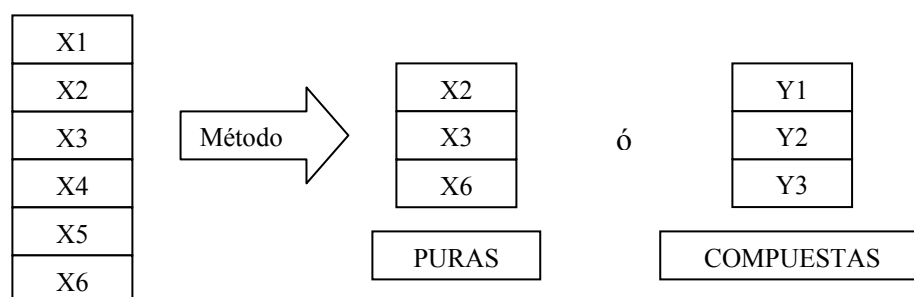


Figura 2- 1: Esquema selección de 3 variables de un universo de 6

En las siguientes secciones se describen cada conjunto de métodos, buscando centrarse en la descripción de aquellos que parezcan ser trascendentes en la generación de la metodología que motiva esta investigación.

II.1 Método de selección utilizando variables compuestas

El objetivo de esta sección es describir el método más representativo de esta clasificación, el Análisis de Componentes Principales.

Como señala Jolliffe (2002), probablemente uno de los métodos de análisis multivariado más antiguo y conocido es el Análisis de Componentes Principales (Principal Component Analysis, "PCA"). Es importante señalar que una variación de este método es conocida con el nombre de Análisis de Factores (Factor Analysis) y muchos textos refieren erróneamente el término PCA como un caso especial de Análisis de Factores.

La técnica 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).

La idea central del PCA es reducir la dimensión de un conjunto de datos, consistente en un gran número de variables interrelacionadas, pero manteniendo al máximo posible la variación presente en el conjunto de datos. Esto se logra transformando las variables iniciales en un nuevo conjunto de variables que se denominan Componentes Principales (PCs), las cuales no están correlacionadas (son ortogonales), y se ordenan de manera tal que las primeras retienen la mayor parte de la variación presente en todas las variables originales.

Técnicamente hablando, el método PCA es una transformación lineal que representa los datos iniciales en un nuevo sistema coordinado, tal que la mayor varianza de cualquier proyección de los datos quede representada en la primera coordenada (denominada primer componente principal), la segunda mayor varianza en la segunda coordenada y así sucesivamente, cumpliendo siempre la condición que la siguiente componente principal debe ser ortogonal a las anteriores¹.

Los PCs pueden ser calculados en función de la matriz de varianza-covarianza o en función de la matriz de correlaciones. Para esta investigación, se explica el método en función de la matriz de varianza-covarianza de los datos Σ de dimensión $p \times p$.

Se define z como el vector cuyo késimo elemento es z_k , correspondiente al késimo PC, $k=1,2,\dots,p$. Entonces

$$z = A'X \quad (2.1)$$

donde X es la matriz cuya késima columna es la serie de valores de la variable k , y A es la matriz ortogonal cuya késima columna, a_k , es el késimo vector propio de Σ . Así, los PCs son definidos como la transformación ortonormal de X , de esta forma se cumple la ecuación

¹ Para ver la derivación matemática del problema, ver Jolliffe (2002). Capítulo 1. Pág. 2-6

$$\Sigma A = A\Lambda \quad (2.2)$$

donde Λ es la matriz diagonal cuyo k ésimo elemento de la diagonal es λ_k , el k ésimo valor propio de la matriz Σ , y $\lambda_k = \text{var}(\alpha_k'X) = \text{var}(z_k)$.

Este método puede ser utilizado en reducciones de dimensión en conjuntos de datos, cuando se desea retener aquellas características de los datos que contribuyen más a su varianza, manteniendo los componentes principales de orden menor e ignorando aquellos de orden mayor.

El caso de análisis que motiva esta investigación, busca representar una serie (por ejemplo la serie de retornos de una cartera) en función de diferentes variables (por ejemplo un subconjunto de índices), por este motivo se cree de vital importancia analizar el uso de componentes principales como estimadores de una serie, es decir, el uso de componentes principales en una regresión, analizando además, las estrategias en la selección de los componentes principales que son relevantes en la regresión.

Las siguientes secciones analizan ambos aspectos basándose principalmente en los textos: Jackson (1991) "A User's Guide to Principal Components" y Jolliffe (2002) "Principal Components Analysis".

II.1.1 Regresión con Componentes Principales

Esta sección tiene como objetivo ilustrar el uso de componentes principales en el análisis de regresiones lineales múltiples. Para poder explicar los alcances de la aplicación de PCs en modelos de regresión lineal, se hará un breve resumen de los principios básicos de la regresión de mínimos cuadrados.

Sea y el vector de $N \times 1$ medidas respuesta y X la matriz de $N \times p$ observaciones de las p variables regresoras. El modelo tradicional de regresión es

$$y = Xb \quad (2.3)$$

donde b es el vector de $p \times 1$ coeficientes de regresión. La solución de mínimos cuadrados para b es

$$b = [X'X]^{-1} X'y \quad (2.4)$$

Como se señala en Gunst (1983) una de las mayores dificultades con los estimadores de mínimos cuadrados en las regresiones con múltiples variables, es el problema de la multicolinealidad, el cual ocurre cuando una o más variables son una combinación lineal de las otras, problema que puede ser detectado en algunos conjuntos de datos debido a la presencia de alta correlación entre las variables.

Si existe multicolinealidad, las varianzas de algunos de los coeficientes estimados en la regresión son muy altas, y por tanto, pueden llevar a errores en la estimación. En la literatura estadística existen variadas formas de lidiar con el problema de la selección de variables para evitar multicolinealidad, y una de ellas es el uso de componentes principales como variables regresoras.

Como los componentes principales son no correlacionados, no existe multicolinealidad entre ellos, además los cálculos para obtener los regresores se simplifican.

Técnicamente hablando, para realizar una regresión utilizando componentes principales, primero se debe transformar las variables a PCs. Para esta sección se utiliza la relación

$$z = XU \quad (2.5)$$

donde $U'U = I$. Esta descomposición escala todos los vectores de coeficientes a ± 1 .

Luego de obtener los PCs, se deben regresionar contra la variable respuesta obteniendo

$$y = zb_z \quad (2.6)$$

donde b_z es el vector que contiene a los coeficientes de regresión obtenidos utilizando los PCs.

Como se mencionó anteriormente, el uso de componentes principales en regresiones busca principalmente solucionar el problema de la multicolinealidad entre las variables regresoras y sus efectos sobre la inversa de $X'X$, los errores estándar de los coeficientes de regresión y la correlación entre estos coeficientes.

El método de PCs abarca todos estos aspectos. La matriz $Z'Z$ es diagonal, por lo tanto los elementos de la inversa son los recíprocos de los elementos de la diagonal. Los coeficientes de la regresión relativos a los PCs tendrán errores estándar mínimos ya que los PCs no están correlacionados y, por la misma razón, los coeficientes de regresión serán no correlacionados.

Si todos los componentes principales son incluidos en la regresión, el modelo obtenido es equivalente al que se obtiene con el método de mínimos cuadrados. Por lo tanto, el real beneficio del uso de componentes principales se observa cuando se desea seleccionar un número menor al total de variables.

La siguiente sección, explica las implicancias de la selección de componentes principales con el fin de reducir el espacio de variables.

II.1.2 Estrategias para seleccionar Componentes en Regresiones

Una de las ventajas del PCA es su potencial para representar un set de variables en un espacio de menor dimensionalidad. Cuando se escoge un subconjunto de k componentes en la ecuación de regresión, existen 2 objetivos diferentes que son considerados como conflictivos. Si el objetivo es eliminar varianzas grandes de los coeficientes de la regresión producidas por el problema de multicolinealidad, se deben borrar aquellas variables cuyo aporte a explicar la varianza total de los datos es muy pequeña, pero, al mismo tiempo, se desea mantener aquellos componentes que tengan alta correlación con la variable dependiente y .

Una estrategia para escoger k , es borrar aquellas variables cuya varianza sea menor que un nivel de corte específico, es decir, retener aquellos componentes que en conjunto permitan explicar un $m\%$ de la varianza total de los datos. Borrar variables basándose solamente en las varianzas, es una estrategia sumamente simple, y por lo tanto, resulta atractiva para los fines que se persiguen con esta investigación.

Sin embargo, una baja varianza para un componente no implica necesariamente que el respectivo componente no tenga importancia en el modelo de regresión. Por ejemplo, Kung and Sharif (1980) proporciona un ejemplo del área de la meteorología donde, en una regresión, los PCs más importantes para fines de la predicción son en orden decreciente, el octavo, segundo y el décimo. El décimo componente aporta menos del 1% de la varianza total de los datos, pero es un predictor importante de la variable dependiente, y el PC más importante en la regresión aporta el 24% de la varianza en y pero sólo el 3% de la varianza total de los datos. Más ejemplos de este mismo tipo son presentados en Jolliffe (1982).

De esta forma, es posible ver que, los dos objetivos de borrar aquellos componentes con varianzas pequeñas y mantener aquellos componentes que son buenos predictores de la variable dependiente, pueden no ser alcanzables de manera simultánea.

En contraste con la selección basada solamente en el tamaño de la varianza, el extremo opuesto es basar la selección sólo en los valores de la medida de los estadísticos t , midiendo la contribución independiente de cada PC a la ecuación de regresión. Hill et al. (1977) proporciona un análisis de varios criterios para decidir cuál PCs borrar de la ecuación de regresión. Sus criterios son de 2 tipos, dependiendo de si el objetivo principal es obtener un \tilde{b} cercano a b , u obtener $\tilde{b}X$, la estimación de y , cercana a y o $E(y)$.

Es posible concluir que la selección de los componentes principales relevantes dependerá exclusivamente del objetivo que se persigue con la representación. En el Capítulo 3 se analiza en detalle cuál de los objetivos será considerado como prioritario en la implementación de la metodología que se busca desarrollar.

Otra forma de utilizar componentes principales para seleccionar variables regresoras es utilizar los datos proporcionados por la composición de cada componente principal y en base a éstos, determinar las variables del universo original que son relevantes para el análisis.

Para fines de esta investigación, el uso de componentes principales en la regresión será enfocado sólo como método en que se utiliza como variable regresora a los PCs relevantes en la regresión, y no como método de selección de variables puras.²

Si se desea seleccionar variables puras, la siguiente sección analiza diversos métodos presentes en la literatura estadística, y, siguiendo la línea investigativa, describe en detalle sólo aquellos considerados como atractivos por su potencialidad como candidatos a utilizar en la metodología que motiva la investigación.

² Si se desea profundizar en el método, vea Jackson (1991) Capítulo 14.4 “Variate Selection”

II.2 Métodos de selección utilizando variables puras

II.2.1 Introducción

Esta sección busca mostrar las diferentes formas que existen en la literatura para seleccionar un subconjunto de variables que permitan ser consideradas como óptimas dentro del universo total de éstas. Tetko et al (1996) plantea que existen varias formas potenciales de resolver este problema, que se diferencian principalmente en el número de iteraciones necesarias para cada solución (tiempo respuesta). Algunos de los algoritmos presentados y sus respectivos tiempos de solución se presentan en la Figura 2-2.

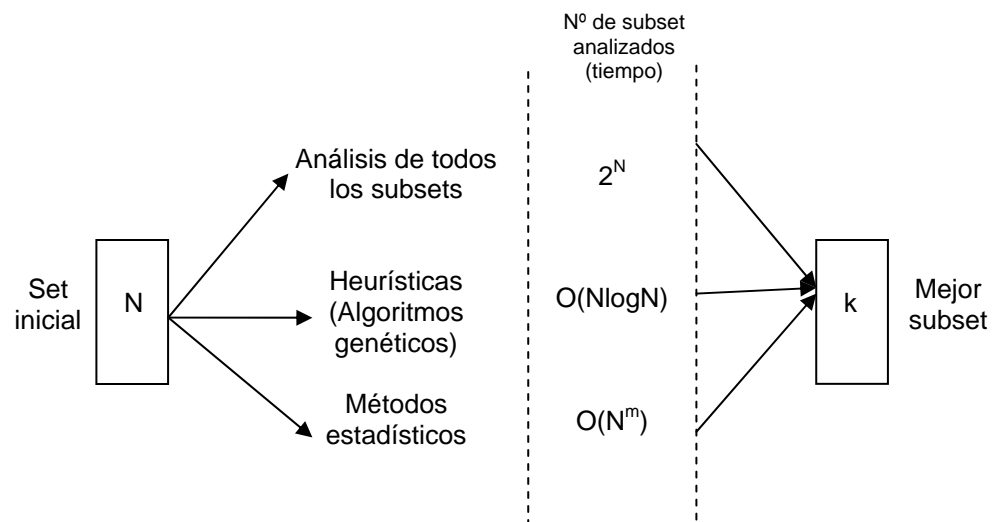


Figura 2- 2: Métodos para selección del mejor subset

Con el fin de analizar cada uno de los métodos presentados y su aplicabilidad al problema, las siguientes secciones presentan una descripción de cada uno de ellos.

II.2.2 Métodos Estadísticos

Como todo problema de selección de subset óptimo, el uso de métodos estadísticos tiene como principal objetivo seleccionar un subset mínimo de variables predictivas sin sacrificar el poder explicativo de éstas.

La principal motivación de la selección de subset parece surgir de inquietudes como: “si 3 regresores pueden “explicar” o “ajustar satisfactoriamente” una respuesta Y , ¿por qué usar 4?, como lo muestran las notas de Mandel (1989). Algunas de las razones para utilizar sólo un subset de las variables predictivas disponibles son dadas por Miller (1984):

- Para estimar o predecir a un costo menor reduciendo el número de variables sobre las cuales se deben recolectar los datos
- Para predecir más adecuadamente eliminando variables que no aportan información
- Para describir un set de datos multivariados.
- Para estimar coeficientes de regresión con menores errores estándar (particularmente cuando algunos de los predictores están altamente correlacionados)

Textos estándar como Draper y Smith (1981) y Montgomery y Peck (1991) señalan que, a grandes rasgos, es posible reconocer 2 grandes focos de estudio:

- a) Metodologías de selección de subsets
- b) Criterios de selección de subsets.

Estos focos de estudio han dado origen a una gran cantidad de artículos. Es por este motivo, que algunos autores se han enfocado en estudiar el estado del arte en la investigación. En relación a la selección de subset óptimo, el trabajo de Hocking (1976)

provee una descripción detallada del campo hasta mediados del 70. Al mismo tiempo, Berk (1977) reporta una comparación computacional de varios procedimientos de selección, y Thompson (1978) da una revisión y una evaluación de procedimientos y criterios de selección. Posteriormente, Miller (1984) y Grechanovsky (1987) ofrecen un examen comprensivo de métodos y criterios de selección y discuten las potenciales fallas que un analista puede enfrentar utilizando selección de subsets. Sparks y Zucchini (1985) examinan los mismos temas, pero para el caso cuando hay múltiples variables Y . Hoerl et al. (1986) muestra un estudio computacional que implica *ridge regression*, y selección de subset secuencial. Finalmente, destacan los trabajos de Kashid, y Kulkarni (2002, 2003) quienes buscan una aproximación para abarcar el problema cuando la distribución no presenta comportamiento normal y particularmente cuando presenta colas gordas (fat tails).

Tanto para la descripción de las metodologías de selección, como para los criterios, el análisis se centrará en explicar aquellos con énfasis en los más tradicionales y sencillos de entender por el usuario. Con este fin, se utiliza la descripción que provee el texto de Draper y Smith (1981).

En relación a las metodologías de selección, el análisis se centra en:

- i) Análisis de todos los subset
- ii) Selección *forward*
- iii) Eliminación *backward*
- iv) Regresión *stepwise*

Con respecto al criterio de selección, las medidas a analizar son:

- i) R^2
- ii) Varianza residual (MSE) s^2
- iii) Cp de Mallows

Es importante mencionar que existen muchas otras metodologías que no serán analizadas en este trabajo, tales como el *stepwise directed search* de Broersen (1986) o *ridge regression* de Hoerl (1986) o el *nonnegative garrote* de Breiman (1995).

Existe además un enfoque paralelo al problema de selección de subset que utiliza una perspectiva Bayesiana y que se presenta en el trabajo de Mitchell y Beauchamp (1988).

A continuación se describe inicialmente los métodos de selección para luego abordar los principales criterios de selección.

a) Metodologías de selección

En relación a las metodologías de selección, Draper y Smith (1981) diferencia entre aquellas consideradas como puramente enumerativas, como es el caso del análisis de todos los subset, de aquellas denominadas como secuenciales, que van paso a paso eliminando o agregando variables según algún criterio. Este es el caso de la selección forward, eliminación backward y la regresión stepwise.

i. Análisis de todos los subsets:

Este primer procedimiento requiere el ajuste de cualquier regresión posible que involucra a cualquier número de variables X_1, \dots, X_n . Como cada X_i puede estar, o no estar en la ecuación (dos posibilidades) y es verdadero para cada X_i , $i=1,2,\dots,n$ existen 2^n ecuaciones posibles. Si $n=10$, un número no despreciable, $2^n = 1024$ ecuaciones deben ser examinadas. Posteriormente se debe analizar la óptimalidad de cada regresión, determinada de acuerdo a un criterio de selección preespecificado.

En general, el análisis de todas las regresiones es un método cuya conveniencia dependerá del tiempo computacional que está dispuesto a utilizar el usuario. Por una parte, significa que el usuario ha analizado todas las posibilidades, pero a su vez también implica que ha examinado un gran número de ecuaciones de regresión, que

muchas veces, analizando cuidadosamente el problema, serían rechazadas de plano. El tiempo computacional usado cuando el número de variables examinadas es no despreciable, es enorme. Algunos tipos de procedimientos de selección que acortan el tiempo de ejecución analizando un número menor de regresiones son preferibles, en aquellos casos en que el usuario está dispuesto a sacrificar la precisión de la respuesta, en la búsqueda de una disminución en el tiempo computacional utilizado.

ii. Selección forward:

Este procedimiento parte del modelo de regresión sin ninguna variable y en cada etapa se introduce la más significativa hasta una cierta regla que determina cuando ya no es necesario seguir ingresando variables. Los pasos básicos de este procedimiento son los siguientes:

1. Se selecciona el X más correlacionado con Y (supongamos que es X_i) y se encuentra la ecuación de regresión lineal de primer orden $\hat{Y} = f(X_i)$
2. Se chequea si esta variable es significativa. Si no lo es, se quita y se adopta el modelo $Y = \bar{Y}$ como el mejor, si no, se busca la segunda variable predictiva a entrar en la regresión.
3. Se calcula el valor del test F parcial³ para todas las variable predictivas que no están en la regresión.
4. El valor del test F parcial mayor, llamado, F_H , se compara con un nivel de significancia preseleccionado, llamado F_0 .
5. Si $F_H > F_0$, se debe agregar la variable X_H , la que originó el F_H , posteriormente se debe recalcular la ecuación de regresión en las variables restantes, y reingresar a la etapa 3

³ El valor F-parcial es asociado con un test de $H_0: \beta=0$ versus $H_1: \beta \neq 0$ para cualquier coeficiente particular de regresión, dependiendo del texto el término F parcial se puede referir también como el F-estadístico para una variable predictiva particular. El valor F-parcial se calcula como

6. Si $F_H < F_0$, se debe seleccionar la ecuación de regresión como la calculada.

Este procedimiento de introducción progresiva tiene la ventaja de necesitar poco cálculo, pero presenta un grave inconveniente, y este es que pueden aparecer errores porque las variables introducidas permanecen en el modelo aunque el algoritmo en pasos sucesivos introduzca nuevas variables que aportan la misma información de las primeras.

iii. Eliminación Backward:

El método de eliminación backward (*backward elimination*) es también un procedimiento que necesita de pocos cálculos. Trata de examinar sólo la “mejor” regresión que contiene un cierto número de variables. Los pasos básicos de este procedimiento son los siguientes:

1. Parte del modelo de regresión que incluye a todas las variables.
2. Se calcula el valor del test F parcial para cada variable predictiva
3. El valor del test F parcial menor, llamado, F_L , se compara con un nivel de significancia preseleccionado, llamado F_0 .
4. Si $F_L < F_0$, se debe quitar la variable X_L , la que originó el F_L , posteriormente se debe recalcular la ecuación de regresión en las variables restantes, y reingresar a la etapa 2
5. Si $F_L > F_0$, se debe seleccionar la ecuación de regresión como la calculada.

Este procedimiento presenta la ventaja de permitir al usuario ver todas las variables en la ecuación una vez, con el objeto de poder observar el comportamiento del conjunto completo. Es económico en términos de tiempo computacional. Sin embargo, si los datos ingresados generan una matriz $X'X$ que es cercana a singular,

entonces la ecuación puede contener errores⁴. Se debe considerar además, que una vez que una variable ha sido eliminada en este proceso se irá para siempre. Así modelos alternativos que utilicen la variable eliminada no están disponibles para un posible análisis.

iv. Regresión Stepwise:

Este procedimiento puede ser visto como una metodología que mezcla los dos procedimientos anteriormente descritos.

El método Stepwise comienza eligiendo una ecuación que contiene la variable X que presenta el mayor ajuste (variable X que presenta el coeficiente de correlación más alto con la serie Y) y luego apunta a construir la ecuación con adiciones secuenciales de X , uno a la vez, mientras estas adiciones sean útiles. El orden de las adiciones es determinado usando los valores del test F parcial para seleccionar cual variable debe entrar la próxima vez. El mayor valor de test F parcial es comparado a un nivel de significancia predeterminado que determina el valor F que se denomina como F_{in} , y que será la cota impuesta por el usuario para que una variable ingrese a la regresión. Después de que una variable ha sido agregada, la ecuación es analizada para ver si cualquier variable debe ser borrada con un criterio de significancia F_{out} .

Los pasos básicos de este procedimiento son los siguientes:

1. Se selecciona el X más correlacionado con Y (supongamos que es X_1) y se encuentra la ecuación de regresión lineal de primer orden $\hat{Y} = f(X_1)$
2. Se chequea si esta variable es significativa. Si no lo es, se quita y se adopta el modelo $Y = \bar{Y}$ como el mejor.

⁴ Estos errores están asociados a la dificultad de invertir matrices singulares. Actualmente, el problema de la inversión de la matrices singulares no es considerado un problema, ya que se soluciona con algoritmos de inversión de matrices más sofisticados.

3. Si es significativa, se examinan los valores de test F parciales de todas las variables predictivas que no están en la regresión. El X_j con el valor más alto (supongamos X_2) es seleccionado y se ajusta una segunda ecuación de regresión $\hat{Y} = f(X_1, X_2)$
4. La regresión es chequeada por significancia, se nota una mejora en el criterio de ajuste, y los valores de F parciales para ambas variables ahora en la ecuación (no sólo la recién ingresada) deben ser estimados.
5. El menor de estos Fs parciales es comparado a un F con un nivel de significancia predeterminado que se denomina como F_{out} y que será la cota impuesta por el usuario para que una variable salga de la regresión. La variable predictiva correspondiente es retenida o rechazada en la ecuación de acuerdo a si con el test es significativa o no.
6. Este testeo del “predictor útil” en la ecuación debe ser realizado en cada etapa del procedimiento stepwise.
7. El procedimiento termina cuando ninguna variable en la ecuación actual puede ser removida y el mejor candidato próximo no puede mantener su lugar en la ecuación (a menos que los niveles de α de entrada y salida sean escogidos equivocadamente para proveer un efecto cíclico⁵).

La razón del análisis del “predictor útil” en cada etapa del procedimiento radica en el hecho de que una variable que puede haber sido el mejor candidato entrante en una etapa anterior, puede en una etapa posterior, ser superfluo por las relaciones entre él y las otras variables que ahora están en la regresión.

Para chequear este efecto, se evalúa el criterio de F parcial para cada variable en la regresión en cualquier etapa de cálculo, y el menor de estos valores de F parciales (el

⁵ Es recomendado elegir los mismos niveles de significancia para ambos el test de entrada y el de salida. Si un α menor es escogido para el test de salida que para el de entrada, un patrón cíclico puede ocurrir. Usar un α mayor para el test de salida puede causar que variables cuya contribución se ha debilitado se mantengan.

que puede estar asociado con el ingresado más reciente o con uno ingresado anteriormente) se compara con un punto de porcentaje preseleccionado de la distribución F apropiada o a un valor F de falla correspondiente.

Una discusión de las ventajas y desventajas de este método se encuentra en Ralston y Wilf (1962).

b) Criterios de selección

Para decidir entre dos o más subconjuntos de variables regresoras en el estudio de un modelo de regresión múltiple, se requiere de medidas del ajuste del modelo construido.

Si se supone que el número máximo de variables explicativas que puede haber en el modelo es n , el número de observaciones es N y, si se ajusta un modelo de regresión lineal con k variables, el número de parámetros del modelo es $k + 1$. Entonces se definen las siguientes medidas o criterios que permiten medir el ajuste:

i. *Uso del Coeficiente R^2 :*

El coeficiente R^2 permite medir el porcentaje de la varianza de la serie Y que es explicado por la regresión \hat{Y} . Existen varias formas análogas de definir el coeficiente. Con fines de esta investigación se define R^2 como:

$$R^2 = \frac{\sigma_{\hat{Y}}^2}{\sigma_Y^2} = 1 - \frac{\sigma_e^2}{\sigma_Y^2} = 1 - \frac{E(e^2) - E(e)^2}{E(Y^2) - E(Y)^2} = 1 - \frac{\sum_{i=1}^N e_i^2 / N - 1}{\sum_{i=1}^N y_i^2 / N - 1 - \bar{Y}^2} \quad (2.7)$$

La principal dificultad que surge, es que este criterio aumenta al ir introduciendo nuevas variables en el modelo. De esta forma, si se define R_j^2 , $j = 1, \dots, n$, como el

máximo valor posible de R^2 cuando en el modelo hay j variables explicativas, es posible verificar que siempre $R_{j-1}^2 \leq R_j^2$.

Sin embargo, es posible establecer análisis complementarios que solucionen este problema. Un criterio sencillo es considerar un número pequeño Δ y elegir el modelo con j más pequeño tal que $R_k^2 - R_j^2 \leq \Delta$ (R_k^2 es el coeficiente del modelo con k variables regresoras).

Otra forma de solucionar el inconveniente anterior, es utilizar la medida de R^2 ajustado. Se define el coeficiente R^2 ajustado como:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{N - 1}{N - (k + 1)} \quad (2.8)$$

El coeficiente \bar{R}^2 tiene en cuenta el número de variables regresoras y no necesariamente crece al introducir un mayor número de variables. Por lo tanto, un buen criterio a utilizar es buscar aquella regresión que tenga el máximo \bar{R}^2 .

ii. *Uso de la varianza residual (Mean square error) s^2 :*

Se define s^2 como:

$$s_k^2 = \frac{1}{N - (k + 1)} \sum_{i=1}^N e_i^2 \quad (2.9)$$

Un buen criterio de selección del subconjunto de variables, es elegir el subconjunto de j variables que minimiza el valor s_j^2 , siendo ésta la varianza residual obtenida con el modelo de j variables.

Teniendo en cuenta que

$$\bar{R}^2 = 1 - \frac{s_j^2}{\sigma_Y^2} \quad (2.10)$$

se deduce que

$$\bar{R}_k^2 > \bar{R}_j^2 \Leftrightarrow s_k^2 < s_j^2 \quad (2.11)$$

por tanto, el criterio de minimizar la varianza residual es equivalente al criterio de maximizar el coeficiente R^2 ajustado.

Otra forma de utilizar este criterio es a través del análisis del gráfico s_j^2 vs j . Si se realizan todas las regresiones en un problema con un gran número de variables, el valor de s_j^2 , es decir, de la varianza residual mientras el número de variables aumenta, indica el mejor punto de corte de variables en la regresión.

Para realizar el gráfico se deben obtener, las varianzas residuales de todos los set de k variables, donde k es el número de parámetros en el modelo incluyendo β_0 .

El ajuste de ecuaciones de regresión que involucra más variables regresoras que las que son necesarias para obtener un ajuste satisfactorio es llamado *overfitting*. Mientras más variables regresoras son incluidas a una ecuación que ya exhibe *overfitting*, la varianza residual tenderá a estabilizarse y se aproxima al verdadero valor de σ^2 mientras el número de variables aumenta, asegurando que todas las variables importantes hayan sido incluidas.

Este procedimiento da una estimación asintótica de σ^2 con la cual se puede elegir un modelo o modelos cuya estimación de la varianza residual sea dada por la asíntota y que contengan el menor número de variables para alcanzarlo.

iii. *Uso del estadístico de Mallows Cp:*

Los criterios anteriores se basan en la media de los errores al cuadrado, pero también es interesante tener en cuenta el sesgo en la selección del modelo, ya que si se omite una variable regresora importante los estimadores de los coeficientes de regresión son sesgados y los criterios anteriores pueden elegir un modelo que tenga sesgo grande aunque su media de los errores al cuadrado sea pequeña.

Un criterio que tenga en cuenta el sesgo ayudará a elegir el modelo adecuadamente. Con este objetivo surge el estadístico C_p de Mallows definido como

$$C_p = p + (N - p) \left(\frac{s_p^2 - s^2}{s^2} \right) \quad (2.12)$$

donde s_p^2 es la varianza residual de un modelo que contiene p parámetros, p es el número de parámetros del modelo incluyendo β_0 , y s^2 es la varianza residual de la ecuación más grande postulada, que contiene todos los X_s .

Es importante notar que hay varios C_p para cada p , excepto para el valor de C_p cuando todas las variables son incluidas, C_{n+1} con:

$$C_{n+1} = n + 1 + (N - (n + 1)) \frac{s_{n+1}^2 - s^2}{s^2} = n + 1 + (N - n - 1) \frac{s^2 - s^2}{s^2} = n + 1 \quad (2.13)$$

Esencialmente se comparan los modelos con p parámetros al modelo completo con $n+1$ parámetros.

Si una ecuación con p parámetros es adecuada, $E(s_p^2) = \sigma^2$, además también se asume que $E(s^2) = \sigma^2$, y que por lo tanto, aproximadamente $E(C_p) = p$ para un modelo adecuado. Por lo tanto, se desprende que el gráfico C_p versus p mostrará

los “modelos adecuados” como puntos cercanos a la línea $C_p=p$. Además puede probarse que en los modelos sin sesgo⁶ $C_p = p$.

Ecuaciones con problemas de ajuste, es decir, ecuaciones sesgadas, se presentan como puntos sobre la línea $C_p=p$. La altura actual de cada C_p , de cada punto graficado, es también de importancia porque es una estimación de la suma total de cuadrados de discrepancia (errores de varianza más error de sesgo) del modelo ajustado con respecto al modelo verdadero.

Así luego de analizar el gráfico de C_p vs p , se busca una regresión con un bajo valor de C_p sobre la recta C_p igual a p . La Figura 2-3 muestra un ejemplo del Gráfico C_p vs p .

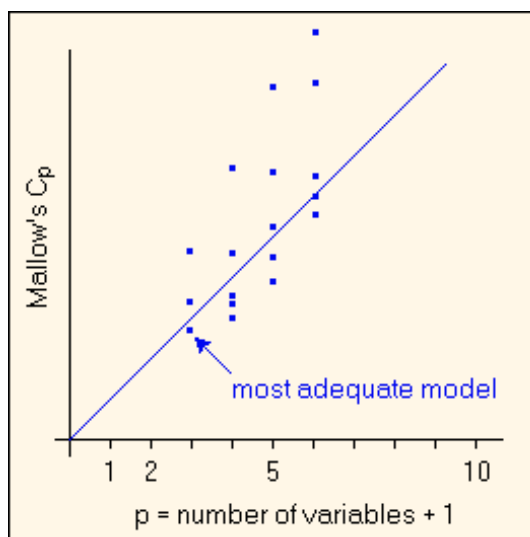


Figura 2- 3: Gráfico Cp vs p

⁶ Para ver más detalles del análisis del estadístico Cp de Mallows ver Mallows, C. (1973). Some comments on Cp.

II.2.3 Análisis Heurísticos

Como se señala a lo largo del texto, existe un gran número de métodos que permiten resolver el problema de la selección óptima de un subconjunto de variables.

Sólo como una forma de ilustrar la revisión bibliográfica que se hizo a lo largo de la investigación, esta sección explica algunos de los métodos que se analizaron como posibles alternativas, pero que finalmente no fueron utilizados en la metodología propuesta.

La rama de investigación que se describe en esta sección está compuesta por los llamados análisis heurísticos.

El término heurística es utilizado para referirse a aquellos métodos o algoritmos exploratorios para la resolución de problemas, en los que las soluciones se descubren por la evaluación del progreso logrado en la búsqueda de un resultado final. Se trata de métodos en los que, aunque la exploración se realiza de manera algorítmica, el progreso se logra por la evaluación puramente empírica del resultado. Con estos métodos se gana eficacia, sobre todo en términos de eficiencia computacional, a costa de la precisión.

Las técnicas heurísticas son usadas por ejemplo en problemas en los que la complejidad de la solución algorítmica disponible es función exponencial de algún parámetro; cuando el valor de éste crece, el problema se vuelve rápidamente inabordable. Estas técnicas no aseguran soluciones óptimas sino solamente soluciones factibles, aproximadas, y frecuentemente no es posible justificar la validez del resultado. La popularización del concepto se debe al matemático George Polya, con su libro *Cómo resolverlo* (How to solve it (1957)).

Por otra parte, existen formas de relacionar estas heurísticas de manera inteligente en la búsqueda de soluciones óptimas, procedimiento que se conoce con el término de meta-heurística (Glover, 1986), y es extensamente aplicado en la literatura (Laporte and Osman, 1995; Gendreau, Laporte and Potvin, 1995; Al-Mahmeed, 1996; Charon and Hudry, 1996; Osam and Nelly, 1996).

Una meta-heurística se refiere a una estrategia que guía y modifica otras heurísticas para producir soluciones mejores que aquellas que normalmente son generadas en búsquedas de optimalidad local. Las heurísticas guiadas por tales meta-heurísticas pueden ser procedimientos muy complejos o pueden implicar nada más una descripción de movimientos disponibles para transformar una solución en otra, unidas con una regla asociativa de evaluación.

Por muchos años, la concepción básica de un procedimiento heurístico era proveer de una regla inteligente o una regla interactiva que termine cuando ninguna solución inmediatamente accesible pueda mejorar la última encontrada. Estas heurísticas interactivas son generalmente referidas a métodos descendentes, métodos ascendentes o métodos de búsqueda local. Consecuentemente, el surgimiento de métodos que parten de este diseño clásico, y que lo hacen por medio de un sistema principal organizado, constituyen un importante avance. Un extenso conocimiento de este avance comenzó a aparecer solo durante la última década, aunque su origen se remonta a varios años atrás.

La evolución de las meta-heurísticas durante la pasada media docena de años ha experimentado una mejora explosiva. Las meta-heurísticas en su forma moderna se basan en una variedad de interpretaciones sobre qué constituye una búsqueda “inteligente”.

A continuación se presentan meta-heurísticas cuya aplicabilidad parece razonable en la búsqueda de la solución del problema planteado: Algoritmos Genéticos, Tabú Search y Simulated Annealing.

a) Algoritmos Genéticos (GA)

Es una técnica de búsqueda utilizada en ciencias de la computación para encontrar soluciones aproximadas a problemas de optimización y búsqueda. Específicamente esta técnica cae en la categoría de técnicas de búsqueda local y es por tanto generalmente una búsqueda incompleta. Los algoritmos genéticos son una clase particular de algoritmos evolutivos que usan técnicas inspiradas por la biología evolutiva como herencia, mutación, selección y cruza (también llamado recombinación).

Los algoritmos genéticos típicamente se implementan como simulaciones computacionales en las cuales una población de representaciones abstractas (llamados cromosomas) de soluciones candidatas (llamadas individuos) a un problema de optimización, evoluciona hacia mejores soluciones. Tradicionalmente, las soluciones se representan en binario como un vector de ceros y unos, pero es posible también presentar diferentes configuraciones.

La evolución comienza de una población de individuos completamente aleatorios y sucede en generaciones. En cada generación, las aptitudes de toda la población son evaluadas, individuos múltiples son seleccionados estocásticamente de la población actual (basado en sus aptitudes), y modificados (mutados o recombinados) para formar una nueva población. La nueva población es usada en la siguiente iteración del algoritmo.

Existen varias ventajas y desventajas en la utilización de este método. A continuación, se presentan algunas observaciones generales acerca de la generación de soluciones vía algoritmos genéticos:

- En algunos problemas con suficiente complejidad, GA puede tener la tendencia de converger a soluciones óptimas locales en lugar de hacerlo al óptimo global del problema. La probabilidad de que esto ocurra depende de la forma del entorno, ciertos problemas pueden proveer un fácil acceso al óptimo global, y otros pueden hacer más fácil para la función el encontrar el óptimo local.

- Operar en conjuntos dinámicos de datos es difícil, dado que los genomas comienzan a converger tempranamente en soluciones que pueden no ser válidas por mucho tiempo para datos posteriores. Algunos métodos han sido propuestos para remediar este problema aumentando la diversidad genética de alguna forma y previniendo convergencia temprana, también aumentando la probabilidad de mutación cuando la calidad de la solución baja o introduciendo ocasionalmente elementos completamente nuevos generados aleatoriamente dentro del pool genético.
- La selección es claramente un operador genético importante, pero la opinión está dividida sobre la importancia de la cruce versus la mutación. Algunos argumentan que la cruce es más importante, mientras que la mutación es solo necesaria para asegurar que soluciones potenciales no se pierdan. Otros argumentan que la cruce en una gran población uniforme solo sirve para propagar innovaciones encontradas originalmente por la mutación, y que en una población no uniforme la cruce es casi siempre equivalente a una mutación muy grande.
- Generalmente, los GA pueden encontrar rápidamente buenas soluciones, incluso para espacios de búsqueda difíciles.
- Como en la solución de cualquier problema, es necesario estimar ciertos parámetros. En el caso de los GA, algunos de estos parámetros son: la probabilidad de mutación, la probabilidad de recombinación y el tamaño de la solución para encontrar ajustes razonables para la clase de problema en la que se está trabajando. Una tasa de mutación muy pequeña puede conducir a derivaciones genéticas o convergencia prematura del algoritmo genético en un óptimo local. Una tasa de mutación que es muy grande puede conducir a perder una buena solución. Hay límites superiores e inferiores teóricos pero no prácticos aún para estos parámetros que pueden ayudar a guiar la selección.

Los problemas que parecen ser particularmente apropiados para ser resueltos por algoritmos genéticos incluyen problemas de horario y programación, también ha sido aplicado a la ingeniería, y generalmente se aplican como una aproximación para resolver problemas de optimización global.

Como regla general los GA pueden ser útiles en dominios de problemas que tengan entornos complejos, dado que operadores como la recombinación se han diseñado para mover la población desde un óptimo local en el cual un método tradicional podría quedarse pegado.

b) Tabú Search

Uno de los más destacados métodos dentro de las metaheurísticas fue introducido y formalizado por Glover (Glover, 1966, 1969; Glover y Laguna, 1977) y es conocido con el nombre de Tabú Search. El algoritmo radica en dos ideas principales, reestructurar múltiples reglas de decisión con el propósito de combinarlas o alternarlas con nuevas decisiones de producción no provistas por ninguna de las reglas aisladas, y llevar patrones de búsqueda más allá del simple óptimo local.

La búsqueda Tabú usa un procedimiento de búsqueda local o de vecindario para moverse de forma iterativa desde una solución x a una solución x' en el vecindario de x , hasta que algún criterio de detención haya sido satisfecho.

Para explorar regiones del espacio de búsqueda que hayan sido dejadas sin explorar por el procedimiento de búsqueda local y que por esto, hayan escapado al óptimo local, la búsqueda tabú modifica la estructura del vecindario de cada solución a medida que la búsqueda progresa. Las soluciones admitidas en $N^*(x)$, el nuevo vecindario, se determinan a través del uso de estructuras de memoria especiales. La búsqueda ahora progresa por movimientos iterativos desde una solución x a una solución x' en $N^*(x)$.

Tal vez el tipo más importante de memoria de corto plazo para determinar las soluciones en $N^*(x)$, también es el que le da su nombre a esta búsqueda, es el uso de la lista tabú, la

cual en su forma más simple, contiene las soluciones que han sido visitadas en el pasado reciente. Las soluciones en la lista tabú se excluyen de $N^*(x)$. Otras estructuras de lista tabú prohíben soluciones que tengan ciertos atributos o previenen ciertos movimientos.

Las listas tabú que contienen atributos son mucho más efectivas, sin embargo crean un nuevo problema. Con la prohibición de un atributo como tabú, típicamente más de una solución es declarada como tabú. Algunas de estas soluciones que deben ahora ser evitadas pueden ser de excelente calidad y no haber sido visitadas aún. Para superar este problema, se ha introducido el criterio de aspiración, el cual permite la eliminación del estado tabú de una solución y por tanto la incluye en el conjunto permitido. Un criterio de aspiración comúnmente utilizado consiste en permitir soluciones que sean mejores que la mejor solución actualmente conocida.

c) Simulated Annealing

Es un algoritmo genérico meta-probabilístico para el problema de optimización global, y se destaca por su capacidad para localizar una buena aproximación al óptimo global de una función dada en un gran espacio de búsqueda. Fue inventado de forma independiente por S. Kirkpatrick, C. D. Gelatt y M. P. Vecchi en 1983, y por V. Cerný en 1985.

El nombre y la inspiración proviene del “recocido” (annealing) en metalúrgica, una técnica que involucra calentamiento y enfriamiento controlado de un material para aumentar el tamaño de sus cristales y reducir sus defectos. El calor ocasiona que los átomos se despeguen de sus posiciones iniciales (un mínimo local de energía interna) y vaguen aleatoriamente a través de estados de mayor energía; el enfriamiento lento les da más oportunidades de encontrar configuraciones con menos energía interna que la inicial.

Análogamente a este proceso físico, en cada paso de un algoritmo de simulated annealing se reemplaza la solución actual por una solución aleatoria cercana, escogida

con una probabilidad que depende de la diferencia entre los valores correspondientes de la función y un parámetro global T (llamado la temperatura), el cual decrece gradualmente durante el proceso. La dependencia es tal que la solución actual cambia casi siempre de manera aleatoria cuando T es grande, efecto que decrece fuertemente cuando T se acerca a cero. El permiso para los movimientos ascendentes evita que este método se pegue en un mínimo local, solución que se obtendría con los métodos del gradiente.

En el método simulated annealing (SA), cada punto s del espacio de búsqueda se compara con un estado de algún sistema físico, y la función $E(s)$ a ser minimizada se interpreta como la energía interna del sistema en ese estado. Por lo tanto la meta es traer al sistema, desde un nivel inicial arbitrario, a un estado con la mínima energía posible.

En cada paso, la heurística de SA considera algunos vecinos s' del estado actual s , y decide probabilísticamente entre mover el sistema al estado s' o permanecer en el estado s . Las probabilidades se eligen de tal forma que el sistema tienda en última instancia a moverse a estados de menor energía. Típicamente este paso se repite hasta que el sistema alcanza un estado que es suficientemente bueno para la aplicación, o hasta que se agota el “presupuesto” asignado en recursos computacionales.

La probabilidad de hacer una transición desde el estado actual s a un estado candidato nuevo s' es una función $P(e, e', T)$ de las energías $e=E(s)$ y $e'=E(s')$ de los dos estados, y de un parámetro global que varía con el tiempo T llamado temperatura. Un requerimiento esencial de la probabilidad de transición P es que debe ser distinta de cero si $e'>e$, lo que significa que el sistema se puede mover a un nuevo estado cuando este es peor (tiene mayor energía) que el actual. Este hecho es el que permite prevenir, a este método de quedar pegado en un mínimo local, un estado en que la energía está lejos de ser mínima, pero aún así posee menos energía que cualquiera de sus vecinos.

Por otra parte, cuando T se acerca a cero, la probabilidad $P(e, e', T)$ debe tender a cero si $e'>e$, y a un valor positivo si $e'<e$. De esa manera, para valores suficientemente pequeños de T , el sistema se moverá a favor de ir hacia a valores de menor energía, y

evitará aquellos movimientos que lleven a valores de energía mayor. En particular, cuando T sea cero, el procedimiento se reducirá al algoritmo del gradiente, el cual realiza el movimiento si y solo si se mueve hacia abajo, es decir, hacia valores de menor energía.

La función P usualmente se escoge de tal forma que la probabilidad de aceptar un movimiento decrece cuando la diferencia $e' - e$ aumenta, esto es, pequeños movimientos hacia arriba son más deseados que grandes movimientos. Sin embargo, este requerimiento no es estrictamente necesario, condicional a que se resuelvan los requerimientos antes mencionados.

Dadas estas propiedades, la evolución del estado s depende crucialmente de la temperatura T .

La temperatura se reduce gradualmente a medida que la simulación avanza. Inicialmente, T se fija en un valor alto (o infinito), y este decrece a cada paso de acuerdo a algún “programa de recocido”, el cual puede ser especificado por el usuario, pero debe terminar con $T=0$ hacia el final del tiempo asignado. En ese sentido, se espera que el sistema vague inicialmente hacia una amplia región del espacio de búsqueda que contiene buenas soluciones, ignorando características pequeñas de la función de energía; luego vuelva hacia las regiones de poca energía que comienzan a hacerse cada vez más estrechas para finalmente moverse hacia abajo de acuerdo a la heurística del método del gradiente.

Este algoritmo parece ser particularmente bueno, incluso puede ser demostrado que, para cualquier problema finito, la probabilidad de que el algoritmo de simulated annealing termine con una solución que sea el óptimo global se acerca a 1 a medida que el tiempo de análisis se extiende. Sin embargo, el tiempo de análisis requerido para asegurar una probabilidad significativa de éxito generalmente excede el tiempo requerido para una búsqueda completa del espacio solución, lo que lo transforma en un algoritmo poco práctico cuando la óptimalidad de la solución es prioritaria.

III. PROBLEMA Y METODOLOGÍA PROPUESTA

III.1 Explicación del Problema

Básicamente, el problema que motiva esta investigación es la necesidad de representar de la mejor manera posible cualquier cartera de inversión. Como se mencionó anteriormente, el problema de la representación de un portafolio de inversión mediante un subconjunto de clases de activos, es enfrentado diariamente dentro del mundo financiero, y se conoce informalmente como proceso de “Mapeo de carteras”.

Para poder abordar el problema, se identifican 2 etapas. La primera etapa consiste en la generación de un procedimiento que permita representar m activos que componen una cartera de inversión, mediante un conjunto de n clases de activos con $n \leq m$. La segunda etapa, consiste en generar una metodología que permita seleccionar un subconjunto k de clases de activos dentro de un set de n clases, con $k < n$. El problema de la segunda etapa se puede describir en términos prácticos como la selección de k variables dentro de un universo de n variables y es el que motiva gran parte de la literatura presentada en el Capítulo 2.

Técnicamente, es posible formalizar el problema de la segunda etapa como:

$$\begin{aligned}
 & \text{Min} \quad f(Y,X) \\
 & \text{s.a} \quad W \cdot 1 = 1 \\
 & \quad \quad W \geq 0 \\
 & \quad \quad W \leq 1 \\
 & \quad \quad \#\{w_i > 0\} \leq K
 \end{aligned}$$

Donde

- $f(Y,X)$: Función que representa el criterio de ajuste utilizado en el problema
- Y : Vector de retornos que representa a la cartera de inversión
- X : Matriz de retornos de las variables que representan las clases de activos de la cartera
- W : Vector de posiciones
- K : Número de final de clases con que se desea realizar la representación.

Una vez establecidas las etapas y las variables que son necesarias para poder comprender el problema que motiva la presente investigación, es posible buscar una metodología que permita solucionar todas las aristas del problema.

La próxima sección describe la metodología propuesta y las distintas interrogantes que se deben resolver para lograr su implementación.

III.2 Discusión de la Propuesta Metodológica

La hipótesis de este trabajo es que es posible desarrollar una metodología simple y robusta que permita representar carteras de mercados emergentes que se caracterizan por tener pocas transacciones, mediante conjuntos de clases de activos de distinta dimensionalidad, y que estas representaciones retengan el comportamiento básico de la cartera de inversión original. En concordancia con la hipótesis propuesta, el objetivo general de esta tesis consiste en diseñar una metodología que defina para cada número m de activos que componen una cartera de inversión, cuál es la composición óptima de k clases, dentro de un universo de n clases de activos, que permite representar un portafolio determinado.

Para lograr generar una metodología que se adapte de la mejor forma posible al inversionista, es necesario resolver ciertas inquietudes referentes a la forma del manejo de datos y carteras que sea acorde al mercado y a las necesidades propias del usuario. A continuación se establecen aquellas interrogantes que se consideran como relevantes a la hora de generar la metodología, explicando para cada una de ellas, el problema que implican en la representación y el impacto de la solución en el diseño de metodología buscada.

III.2.1 Precios para representar la Cartera de Inversión

Como se mencionó anteriormente, el principal objetivo de esta investigación es lograr representar, de la mejor forma posible, una cartera de inversión específica.

Para representar esta cartera de inversión, es necesario establecer qué precios se utilizarán para representar la inversión de la cartera a lo largo del tiempo, es decir, es necesario establecer la serie que se busca imitar con la representación.

La tendencia natural, sería utilizar los precios históricos que ha tenido esta cartera de inversión a lo largo del tiempo, es decir, si la cartera de inversión que se busca imitar fuese por ejemplo, un fondo mutuo, se entiende como precios históricos los valores cuota que el fondo ha obtenido a lo largo de la historia.

Sin embargo, el uso de los precios históricos de la cartera presenta varios problemas. Los tres problemas principales que implica el uso de precios históricos son

- a) La cartera de inversión es activa
- b) Los papeles que componen la cartera cambian sus características con el tiempo
- c) Falta de precios en mercado incompletos

Algunos de éstos, son propios de cualquier mercado de inversión, independiente de la profundidad de éste, y algunos son propios de mercados emergentes, como el mercado nacional, que tienen la característica de ser mercados ilíquidos, de pocas transacciones.

A continuación se explica en detalle cada uno de los problemas.

a) Cartera Activa

Cartera activa, se refiere a la variabilidad que tienen las carteras de inversión a lo largo del tiempo. Por lo general, los *managers* que manejan portafolios de inversión constantemente cambian su cartera, tanto en selección de activos como en exposición, con el objetivo de buscar retornos mayores.

Por lo tanto, el supuesto de que la cartera mantiene sus posiciones estables durante todo el periodo a evaluar es insostenible.

Esta característica es propia de cualquier mercado de inversión y de cualquier cartera de inversión que se gestione en forma activa.

b) Los papeles que componen la cartera cambian sus características con el tiempo

Una particularidad de los papeles de Renta Fija, es que cambian sus características a lo largo del tiempo. Muchas de las características propias de los papeles de Renta Fija van variando, por ejemplo, la duración de los papeles varía a lo largo del tiempo, lo mismo ocurre con la clasificación de riesgo del papel.

Esta característica es propia de cualquier mercado de Renta Fija, independiente de la completitud de este.

c) Falta de precios en mercado incompletos:

Aunque un papel no cambiase de características, como es el caso de las acciones, existe el problema de la falta de precios para todos los días cuando el papel es poco líquido.

Esta es una característica propia de mercados emergentes que presentan el problema de mercados incompletos debido a la baja frecuencia de transacciones de algunos papeles.

Para solucionar los problemas anteriormente descritos se propone realizar un proceso que se denominará como “Pre-Mapeo” Este proceso consistirá básicamente en replicar el comportamiento de la cartera actual en el pasado. Para lograr este objetivo, lo primero es mantener fijas las posiciones actuales de la cartera en cada uno de los activos de inversión. Con esto se soluciona el problema de la cartera activa.

Por otra parte, para representar a cada activo de la cartera, se propone utilizar una serie de índices, los cuales se definen como un activo o portafolio de activos que poseen la característica de ser estables en su definición y completos en precios.

Estables en su definición, se refiere a la característica que posee el índice de ser capaz de representar al activo en sus condiciones actuales, aún mejor que una serie de precios de ese mismo papel. Por ejemplo un bono va cambiando su duración a través del tiempo, mientras que en el caso del índice que representa al mismo bono, la duración puede mantenerse constante, por lo que para representar ese activo en particular es mejor utilizar una secuencia de “precios” de un índice que mantenga constantes sus características en lugar de los precios reales del mismo activo. Con esto se soluciona el problema del cambio de características del papel.

Completos en precios, se refiere a la característica de poseer precios diarios, es decir, un activo o conjunto de activos para el cual se ha resuelto el problema de la ausencia de datos mediante una metodología determinada, independiente para cada clase de activo. Con esto se soluciona el problema de la falta de precios.

III.2.2 Diferenciación entre tipos de Índices

Una vez definido el proceso de Pre-Mapeo y establecida la necesidad de utilizar índices para representar a los activos presentes en la cartera, es necesario determinar el grado de detalle o generalidad del índice.

Dentro del universo de índices disponibles en el mercado, claramente es posible diferenciar entre aquellos índices que representan a un sector de inversión amplio y que se caracterizan por estar compuestos por un gran número de activos, de aquellos que se utilizan para representar a un sector de inversión específico o incluso a veces sólo a un activo específico.

En base a esta diferenciación se propone generar una clasificación que llamaremos “global/específico” o “padre/hijo” y consiste en distinguir entre aquellos índices compuestos por un gran número de activos, los que se denominarán como “globales o padres” y aquellos que están compuestos por un número inferior de activos que se puede considerar subconjunto del universo de alguno de los índices globales, los que se denominarán como “específicos o hijos”.

La idea que se persigue con esta clasificación es permitir la presencia de índices globales en la representación final. Estos índices globales no necesariamente estarán presentes en la cartera “Pre-Mapeada” prefiriendo el uso de índices específicos ya que se busca una representación lo más detallada posible.

III.2.3 Función objetivo o criterio de ajuste de la representación

Para decidir entre dos o más subconjuntos de variables, es fundamental disponer de medidas que establezcan que tan bueno es el ajuste que proporciona el conjunto de variables seleccionadas, es decir, es necesario establecer una adecuada función objetivo para escoger un subset óptimo.

En el capítulo 2 se describen 3 criterios que permiten medir el ajuste de una selección de variables, el R^2 , la varianza residual s^2 y el Cp de Mallows.

Como la premisa básica de esta investigación es generar una metodología sencilla, la selección se enfoca en los criterios ampliamente utilizados, es decir, en R^2 y en la varianza residual s^2 .

Como se demostró anteriormente, maximizar R^2 , es análogo a minimizar la varianza residual, y en esta misma línea, minimizar la varianza residual, es análogo a minimizar la suma de los errores al cuadrado (SE). Por lo tanto, para determinar un subconjunto como óptimo, se escogerá aquél que presente el mínimo SE.

Esta elección parece la más sensata, ya que es intuitiva para el usuario. Una forma sencilla y práctica de seguir el comportamiento de una serie a través de un conjunto de variables es utilizando una regresión. Cuando se resuelve una regresión utilizando el método de mínimos cuadrados, la función objetivo que se busca minimizar corresponde

a la suma de errores al cuadrado $f(Y,X) = \sum (Y-WX)^2$.

Además, la elección de esta función objetivo se reafirma con la elección de la función comúnmente utilizada en los problemas de index tracking, el Tracking error (TE), ya que como es posible ver, minimizar el TE es equivalente a minimizar la suma de errores al cuadrado⁷.

⁷ Min TE = Min Var(e^2) = Min E(e^2)-E(e)² = Min E(e^2)-0 = Min $\Sigma e^2/n$ = Min Σe^2

III.2.4 Selección de variables “Compuestas” o variables “Puras”

Esta sección analiza el uso de métodos que seleccionan variables “Compuestas” en comparación con los métodos que seleccionan variables “Puras”.

Recordemos que para toda esta investigación se entiende como variable “Compuesta”, aquella que se genera como alguna combinación de variables originales, y que no es directamente observable. Análogamente, se entiende por variable “Pura” aquella variable que se obtiene de una selección dentro del universo total de variables disponibles, y que es, por lo tanto, observable.

Para poder definir este punto, es crucial determinar el objetivo que el usuario persigue con la representación, y lo aceptable o no que es el uso de variables “Compuestas” en la representación final. Es importante señalar que la representación mediante variables “Compuestas” sólo se considera como aceptable, cuando no es importante la interpretación de las variables que representan la cartera de inversión.

Según el objetivo, se presentan dos opciones. Si se desea representar mediante variables compuestas, el método consiste en obtener los k componentes principales más representativos. Como el objetivo final es lograr el mayor ajuste a la serie Pre-Mapeada, se debe seleccionar aquellos PCs más significativos en la regresión entre la serie Pre-Mapeada y los PCs.

Por otra parte, si se busca la mejor representación mediante el uso de variables puras, el método consiste en seleccionar k variables del universo total, mediante algún proceso algebraico, entregando finalmente la selección que mejor representa a la cartera específica.

El sentido de utilizar variables Compuestas, radica básicamente en que la obtención de la selección de variables utilizando PC presenta 2 ventajas:

- La obtención de la solución mediante el método de PC es rápida, en términos de tiempo computacional.
- Los PCs son no correlacionados por lo que se reduce el problema de multicolinealidad de la solución.

En resumen, la elección de variables “Puras” o “Compuestas” radicará básicamente en 2 criterios, el objetivo final que el usuario busca con la representación y el sacrificio que está dispuesto a hacer en términos de precisión vs. tiempo computacional.

III.2.5 Algoritmo de Optimización

El último punto a analizar corresponde a la selección del método de inspección de subconjuntos para llegar al óptimo.

Si las variables seleccionadas son “Compuestas”, el algoritmo de selección consiste en regresionar la serie de la cartera Pre-Mapeada contra los PCs y seleccionar aquellos que presenten un mayor ajuste.

Por otra parte, si las variables seleccionadas son “Puras”, el método de inspección de subconjuntos debe ser más complejo, seleccionando para este fin un método estadístico.

Dentro de los métodos de selección, aquel que provee la solución óptima, es el análisis de todos los subset. Sin embargo, el tiempo computacional usado cuando el número de variables examinadas es grande, puede llegar a ser enorme.

Técnicamente hablando dentro de un universo de n variables, si deseo seleccionar k , el número de subconjuntos que se debe inspeccionar es de

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (2.14)$$

Obteniéndose un valor máximo cuando $k = \frac{n}{2}$, donde

$$\binom{n}{k} = \binom{n}{\frac{n}{2}} = \frac{n!}{\frac{n}{2}! \frac{n}{2}!} = \frac{\prod_{i=\frac{n}{2}+1}^n i}{\frac{n}{2}!} \quad (2.15)$$

Cuando el número de subconjuntos a inspeccionar es muy grande, parece razonable elegir alguno de los procedimientos de selección que acortan el tiempo de ejecución, analizando un número menor de subconjuntos. Sin embargo, estos métodos presentan el inconveniente de la disminución de la precisión de la respuesta, en algunos casos.

Como el objetivo es lograr una solución con la mayor precisión posible, se propone la idea de reducir de alguna forma el universo total de variables a analizar cuando el número de subconjuntos que debe ser inspeccionado (y por lo tanto, el tiempo computacional utilizado) parece muy grande.

Dentro de los métodos estadísticos descritos en el Capítulo 2, existe uno que elimina variables en forma iterativa hasta llegar a una solución óptima con algún criterio de impuesto por el usuario. Este método es el descrito bajo el nombre de eliminación backward

La idea que se quiere presentar, es modificar de cierta forma el método para ser utilizado sólo como una forma de reducción del universo total de variables. Es decir, la idea es ir eliminando de forma iterativa variables hasta que el universo se reduzca a un tamaño de n^* ⁸, para posteriormente realizar el método de analizar todos los subconjuntos sobre este universo reducido.

Este método de reducción lo denominaremos como “backward modificado” y consiste en reducir el universo inicial de variables a un número F de variables, en base a la eliminación de aquella que menos aporta a la regresión reduciendo el universo en forma paulatina, iterativamente.

⁸ Donde $k \leq n^* \leq n$. Para ver en detalle la selección de n^* ver III.3 Descripción de la Metodología Propuesta. .

Más precisamente, el método “backward modificado” se puede explicar de la siguiente forma:

1. Parte del modelo de regresión que incluye a todas las n variables
2. Se realizan todas las regresiones posibles utilizando todas las variables disponibles menos una y se calcula el ajuste R^2 para cada una de estas alternativas
3. Se selecciona la regresión con una variable eliminada que posea mayor R^2
4. Si $n^* > F$, se reingresa a la etapa 2
5. Si $n^* = F$, se debe seleccionar a al subconjunto de n^* variables como el universo reducido

Es importante señalar que la utilización del método backward para reducir el universo inicial dependerá exclusivamente del número de subconjuntos que el usuario este dispuesto a inspeccionar, es decir, del tiempo computacional disponible. Utilizando el método de reducción se gana eficacia, sobre todo en términos de eficiencia computacional, a costa de la precisión.

Una vez resueltas las principales interrogantes sobre la implementación, es posible definir la metodología para resolver el problema en cuestión.

III.3 Descripción de la Metodología Propuesta

El primer paso de la metodología consiste en realizar el proceso definido en la sección anterior como “Pre-Mapeo”, es decir, asignar a cada activo de la cartera, tenga éste precios diarios o no, un “índice representativo” que sí posea precios diarios. Y posteriormente generar la serie que represente los retornos de la cartera en el tiempo dejando fija la cartera que se desea representar. Es decir, se genera una serie en base a los retornos de los índices seleccionados, que representa el retorno que habría obtenido la cartera representada si se hubiese mantenido estática durante el período de análisis.

El siguiente paso consiste en escoger el número k de clases con la que se quiere realizar la representación⁹, donde $k \leq n$, con n representando el número inicial de clases de activos que representan la cartera.

Posteriormente es necesario, determinar, si el usuario está dispuesto a utilizar índices “Compuestos”, es decir, combinaciones de varios de los índices iniciales o si necesita seleccionar específicamente un subconjunto de las variables originales o índices “Puros”.

Si se desean índices compuestos, se debe encontrar los Componentes Principales asociados al set de datos inicial. Posteriormente, se utiliza el método de componentes principales en regresión utilizando como variables predictoras los componentes obtenidos del set de datos compuesto por los índices, y como variable dependiente la serie generada en el proceso de pre-mapeo. Se entrega finalmente, los k vectores generados más significativos en la regresión, es decir, aquellos que posean los mayores estadísticos¹⁰ t.

Si se desean índices puros, el primer paso consiste en clasificar cada índice como global o específico (padre/hijo), según el criterio señalado anteriormente. Luego, se determina el número máximo de iteraciones I que el usuario está dispuesto a realizar.

⁹ Para ver métodos de selección de la variable k ver Salas (2007)

¹⁰ Para ver más detalles del proceso de regresión con componentes principales, ver Capítulo II.1 Métodos de selección utilizando variables compuestas

Posteriormente se calcula el número total de combinaciones posibles T para el set de clases de activos inicial de la cartera en cuestión, (n sobre k).

Una vez realizadas las fases anteriores, si el número total de combinaciones T es menor o igual que el número de iteraciones máximas impuestas por el usuario I , se testean todas las combinaciones posibles, de k índices, regresionando en cada caso a través de una minimización de la suma de errores al cuadrado entre la serie replicada (con intercepto) y la serie de retornos replicada inicialmente. De esta forma, se elige aquel subconjunto que entrega la menor suma de errores al cuadrado, y por tanto que tenga el mayor ajuste determinado por R^2 de entre todos los analizados.

Por otra parte, si el número total de combinaciones T es mayor que el número de iteraciones máximas impuestas por el usuario I , se utiliza un algoritmo de eliminación sobre el universo de índices iniciales “específicos” (hijos), hasta que el número n^* donde $k \leq n^* \leq n$, donde $n^* = (\text{n}^\circ \text{ índices globales} + \text{n}^\circ \text{ índices específicos reducidos})$, de modo tal que el total de combinaciones posibles T^* para el set de índices que representan el universo reducido, (n^* sobre k) sea menor o igual que I menos I_B . Donde I_B , es el número de iteraciones realizadas en el proceso backward.

Posteriormente, se testean todas las combinaciones posibles de k índices, sobre el universo reducido n^* , regresionando en cada caso a través de una minimización de la suma de errores al cuadrado entre la serie replicada (con intercepto) y la serie de retornos replicada inicialmente. Análogamente al caso anterior, se elige aquel subconjunto que entrega la menor suma de errores al cuadrado, y por tanto que tenga el mayor ajuste determinado por R^2 de entre todos los inspeccionados en el universo reducido. A continuación se precisa el algoritmo recién descrito:

1. Se Pre-Mapea la cartera
 - a. Se asigna a cada activo de la cartera un “índice representativo”, generando n clases de activos (n° de índices diferentes).
 - b. Se genera la serie que represente los retornos de la cartera en el tiempo en función de la cartera fija y los retornos de los índices seleccionados.
2. Se escoge el número k clases con la que se quiere realizar la representación, donde $k \leq n$
3. Si se desea representar mediante índices “Compuestos”
 - a. Se encuentra PCs asociados al set de datos de n índices.
 - b. Se regresa PCs contra serie Pre-Mapeada
 - c. Se calcula t estadístico para cada PC presente en la regresión.
 - d. Se seleccionan los k PCs con t estadístico mayor.
4. Si se desea representar mediante índices “Puros”
 - a. Se realiza clasificación padre/hijo de los índices.
 - b. Se determina el número máximo de iteraciones I que el usuario está dispuesto a realizar.
 - c. Se calcula el número total de combinaciones posibles T para el conjunto de clases de activos inicial de la cartera $\binom{n}{k}$.
 - d. Si $T < I$
 - i. Se regresa todas las combinaciones posibles de k en n
 - ii. Se escoge la combinación de k con mejor ajuste R^2

e. Si $T > I$

- i. Se realiza eliminación backward de variables hijos hasta reducir el universo a n^* , donde $k \leq n^* \leq n$, tal que $\binom{n^*}{k} + I_B < I$. Donde I_B , es el número de iteraciones realizadas en el proceso backward.
- ii. Se regresiona todas las combinaciones posibles de k en n^* , donde $n^* = (\text{n}^\circ \text{ índices globales} + \text{n}^\circ \text{ índices específicos reducidos})$ y $k \leq n^* \leq n$
- iii. Se escoge la combinación de k con mejor ajuste R^2

IV. EJEMPLO DEL USO DE LA METODOLOGÍA EN UNA CARTERA DE RENTA FIJA NACIONAL

IV.1 Descripción de Datos

Como se mencionó anteriormente, para la implementación de la metodología es necesario seleccionar un mercado, una cartera específica de mercado y un conjunto de índices que sean capaces de representar de la mejor manera posible el universo de mercado. Con el fin de analizar la implementación se busca representar la cartera de los Saldos Nominales de los instrumentos de Renta Fija presentes en la custodia del Depósito Central de Valores (DCV) utilizando para este fin, índices de Renta Fija Nacional.

A continuación presentamos una breve descripción de cada uno de los datos a considerar.

IV.1.1 Mercado de Renta Fija en Chile

Dentro del mercado financiero nacional, los instrumentos de renta fija tienen una participación bastante importante, correspondiendo a esta clase los mayores montos transados diariamente. A pesar de su tamaño, este mercado tiene pocas transacciones para la gran mayoría de sus instrumentos, característica general del mercado chileno, lo que dificulta la representación de carteras de inversión al no existir precios diarios de todos los instrumentos.

Dentro del mercado de Renta Fija Nacional se pueden distinguir 5 tipos de familias de papeles: Bonos de Gobierno, Bonos Corporativos, Depósitos a Plazo, Letras

Hipotecarias y Bonos de Reconocimiento. A continuación se presenta una breve descripción de cada una de ellas.¹¹

a) Bonos de Gobierno

Se considerará bonos de Gobierno a todo papel emitido por Instituciones del Gobierno de Chile (Banco Central y Tesorería General de la República.). A continuación se presenta una breve descripción de algunos de los papeles más importantes de esta familia:

i. Pagarés reajustables del Banco Central de Chile con pago en cupones (PRC)

Bono amotizable emitido por el Banco Central de Chile, que paga cupones semestrales que comprenden capital e intereses salvo el último cupón que puede ser diferente. El plazo de vencimiento de estos títulos es desde cuatro hasta veinte años y devengan un interés a una tasa anual vencida, calculado sobre el capital expresado en UF.

ii. Cupones de emisión reajustables opcionales (CERO) en UF

Cupones emitidos con la finalidad de dar la opción de sustitución y canje a los tenedores de Pagarés reajustables del Banco Central de Chile con pago en cupones (PRC). Esta sustitución posibilita el fraccionamiento de los cupones de los PRC, permitiendo disociar los pagos del instrumento original, teniendo de esta manera un tratamiento como valores independientes y entregando una mayor y más eficaz información de la curva de rendimiento.

iii. Bonos del Banco Central de Chile en pesos y en UF (BCP) (BCU)

Los BCP y BCU se emiten con la finalidad de reemplazar a los instrumentos del mismo emisor: PDBC y PRC, respectivamente, que circulan actualmente en el mercado financiero local, procurando, entre otros objetivos, estandarizar los instrumentos a formato *Bullet*.

¹¹ Para ver más detalles en la descripción de los instrumentos vea Anexo A: Características de los instrumentos del mercado de Renta Fija en Chile

Estos bonos pagan cupones con vencimientos semestrales iguales y sucesivos, que incluyen los intereses devengados, salvo el último cupón que comprende capital e intereses. Los cupones se pagan en moneda corriente nacional, a la fecha de los correspondientes vencimientos.

iv. Bonos de la Tesorería General de la República en UF (BTU)

Bonos Bullet emitidos por la tesorería General de la República

De los bonos que se han colocado hasta la fecha, unos devengan interés de 4,5% anual vencido y otros devengan interés de hasta 3% anual vencido, que se calcula en forma simple sobre la base de períodos semestrales de 180 días y anuales de 360 días. Los intereses que se vayan devengando por la posesión de estos títulos, se pagan semestralmente. Cada pago semestral de intereses se hace constar mediante un cupón que forma parte del bono respectivo.

La amortización del capital de los bonos se realiza en una sola cuota al vencimiento de los mismos.

b) Bono Corporativo

Títulos de deuda emitidos por empresas, principalmente sociedades anónimas abiertas, que se emiten para financiar proyectos de inversión de largo plazo. También se emiten para satisfacer compromisos financieros, como el refinanciamiento de pasivos.

c) Depósitos a plazo

Título de deuda que se emite para certificar el depósito de dinero recibido de una persona, sea esta natural o jurídica, bajo la obligación de restitución del dinero dentro de un plazo preestablecido, incluyendo reajustes e intereses convenidos, según corresponda.

Estos instrumentos son considerados como instrumentos de Intermediación Financiera.

d) Letras Hipotecarias

Instrumento para el otorgamiento de préstamos con garantía hipotecaria, emitidos por empresas bancarias, sociedades financieras y aquellas cooperativas de ahorro y crédito sometidas a la fiscalización y control de la Superintendencia de Bancos e Instituciones Financieras.

e) Bonos de Reconocimiento

Títulos de deuda emitidos por el Instituto de Normalización Previsional (INP) para reconocer períodos de cotizaciones previsionales que registraron imponentes del antiguo sistema previsional y que optaron u optan por afiliarse al nuevo sistema de previsión en alguna Administradora de Fondos de Pensiones (AFP), de acuerdo con lo establecido en el Decreto Ley N.º3500.

IV.1.2 Cartera de Saldos Nominales de Instrumentos de Renta Fija en Depósito Central de Valores DCV

El Depósito Central de Valores es una empresa cuyo objeto es recibir en depósitos valores de oferta pública, facilitar la transferencia de éstos a través del procesamiento y registro electrónico de transacciones y proveer el servicio de administración de los Valores custodiados.

DCV procesa y registra electrónicamente las operaciones de transferencias efectuadas en las bolsas de valores como también en el mercado extra bursátil y facilita la información necesaria para el pago de los derechos relativos a los valores depositados.

En el Depósito Central de Valores (DCV) permanecen guardados instrumentos financieros por más de US\$ 66.135 millones. Los instrumentos custodiados son depósitos a plazo, letras de crédito, acciones, efectos de comercio y pagarés, entre otros que pertenecen a todas las AFP, Compañías de Seguros, Corredores de Bolsa, Administradoras de Fondos Mutuos, Bancos y sus respectivos clientes.

Además de llevar un registro de los montos nominales que se mantiene en depósito, es posible encontrar información de valorización de los valores mantenidos, conforme al valor par en el caso de instrumentos de renta fija y al valor final en el caso de instrumentos de intermediación financiera. Tratándose de instrumentos de renta variable éstos son valorizados según los precios informados por la Superintendencia de Administradoras de Fondos de Pensiones y en ausencia de dichos precios, se valorizan utilizando el último precio de cierre informado por la Bolsa de Comercio de Santiago o en su defecto por la Bolsa Electrónica de Chile o la Bolsa de Corredores. Todas las valorizaciones mencionadas precedentemente son expresadas en Unidades de Fomento.

Para este ejemplo se busca representar la cartera de los Saldos Nominales de los instrumentos de Renta Fija presentes en el DCV el día 30/06/2006.

Debido a la gran magnitud de la cartera representada, no es posible mostrar la cartera en detalle papel por papel, compuesta por más de 50.000 papeles. Sin embargo, La Figura 4-1 permite observar la composición de esta cartera por familias de papeles.

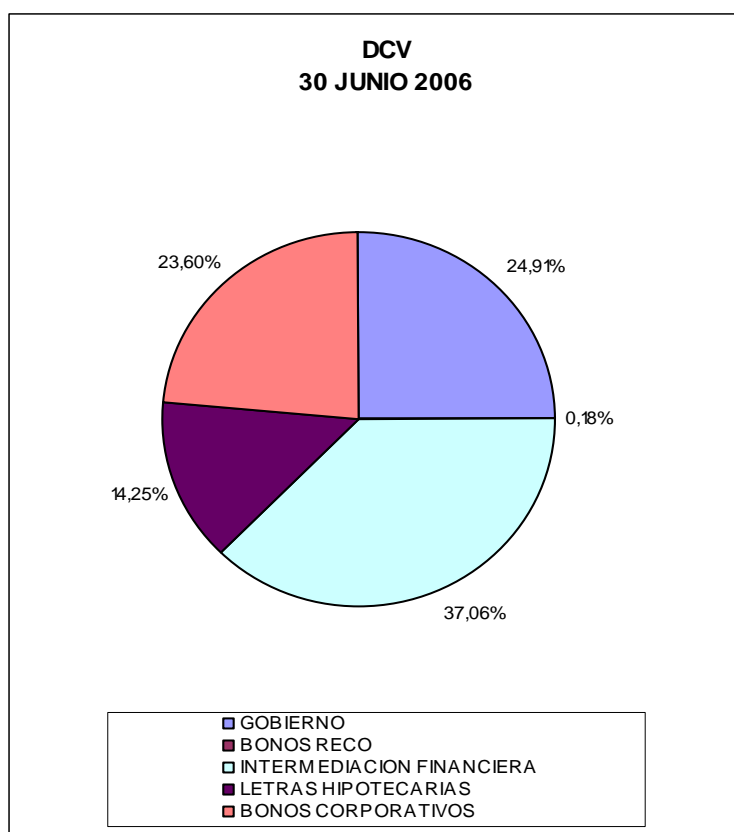


Figura 4- 1: Composición Cartera Saldos Nominales RF DCV el 30/06/2006

IV.1.3 Índices de Renta Fija RA

Como se señaló anteriormente, para representar las inversiones de una cartera en un mercado emergente, se utilizarán índices pasivos que de cierta forma reflejen el comportamiento de los activos presentes en la cartera de inversión.

Como la mayoría de los mercados emergentes, el mercado chileno de instrumentos de renta fija presenta una baja frecuencia de transacciones, es decir es poco profundo, lo que va en contra de la necesidad de tener diariamente precios de los diferentes activos que componen este mercado. Muchos de estos instrumentos se transan un par de veces al mes por lo que la obtención de precios diarios para todos los instrumentos se vuelve una tarea sumamente compleja.

Por este motivo, se hace necesaria la correcta elección de algún modelo de valorización que entregue diariamente estimaciones de precios adecuadas para los distintos tipos de instrumentos.

Con el fin de solucionar estos problemas y de proporcionar herramientas que apoyen las decisiones de inversión y gestión del riesgo financiero, DICTUC S.A., filial de la Pontificia Universidad Católica de Chile (UC), ofrece un servicio diario de valorización y asignación de tasas para los instrumentos de renta fija del mercado nacional, expresados en pesos y en UF, basado en un Modelo Referencial de Precios (MRP) desarrollado por FINlab-UC.

Este modelo se caracteriza por utilizar toda la información de mercado disponible, a través de información de transacciones del día, y además de fechas pasadas. Esta última es utilizada en las consultas de datos históricos y en la estimación de Estructuras de Referencia mediante un modelo dinámico de no arbitraje tipo Vasicek¹², obteniéndose así valorizaciones que se ajustan a la realidad del mercado nacional, y que poseen

¹² Para ver más detalles acerca de modelo ver Cortazar et al. (2007). Term Structure Estimation in Markets with Infrequent Trading.

volatilidades estables y consistentes con las volatilidades observadas para los distintos papeles.

A través de este modelo es posible generar una cinta de precios de instrumentos de renta fija para todos los días, precios que actualmente permiten valorizar diariamente la cartera de Renta Fija de todos los Fondos Mutuos del Mercado Nacional.

En base a esta cinta de precios y como parte del proyecto RiskPortfolio¹³, se generan índices de Renta Fija para el mercado Nacional.

Para la construcción de estos índices, los instrumentos presentes en cada uno de éstos se ponderan en base a sus saldos en el DCV, de manera de crear un índice lo más representativo del mercado. Los instrumentos que componen un determinado índice se revisan una vez al mes a principios de éste, eliminando los que ya no cumplan con los requisitos del índice, y agregando los que posean las características necesarias para pertenecer a éste. Todos los pagos efectuados por los instrumentos, ya sean intereses, amortizaciones o prepagos, se reinvierten en el mismo índice.

Actualmente, se han desarrollado más de 100 índices de Renta Fija Nacional, que basan su composición principalmente en las características particulares de cada instrumento. Es así como se pueden obtener índices globales, por familia, por papeles, por duración, por moneda o por clasificación de riesgo, dependiendo de las características de los papeles que componen cada uno de ellos.

Para este ejemplo, se selecciona la segmentación de índices por Duración. El motivo de esta selección radica principalmente en 2 factores. El primero es que se busca seleccionar un criterio que sea válido para todas las familias de índices y el segundo, que las diferencias en el comportamiento de los índices según está característica se determinen como más significativa que el resto.¹⁴ La Figura 4-2 presenta un mapa de

¹³ RiskPortfolio es un Proyecto Fondef desarrollado por el Departamento de Ingeniería Industrial y de Sistemas de la Pontificia Universidad Católica de Chile, que contó con la participación y el apoyo de AFP Habitat y cuyos resultados actualmente son parte de sitio web: www.RiskAmerica.com

¹⁴ Para más detalles acerca de la elección de índices ver Salas (2007)

los índices actualmente disponibles y sus clasificaciones. Señalando con círculo rojo los índices seleccionados para fines de esta investigación.

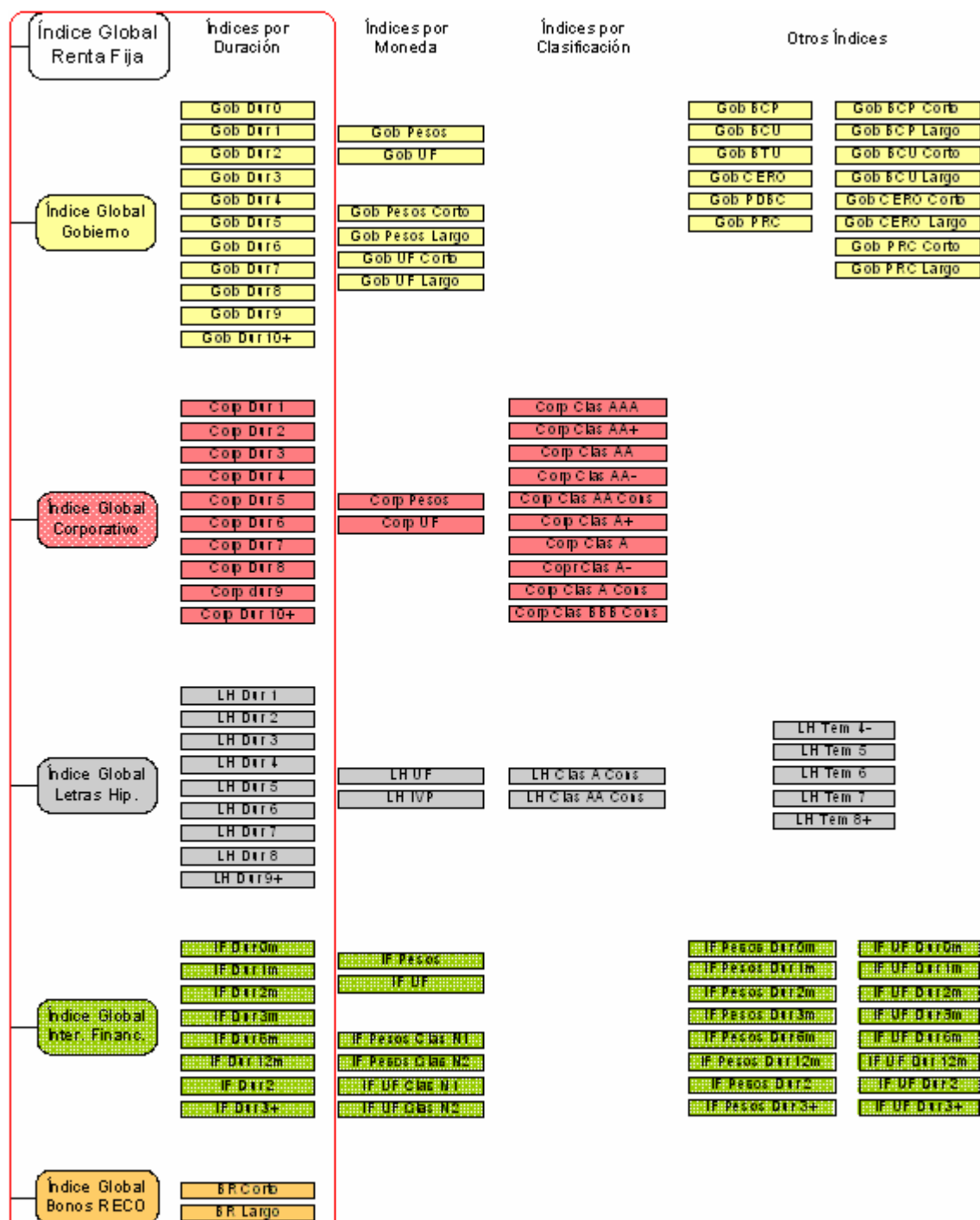


Figura 4-2: Mapa Índices Renta Fija RA

De esta forma, se consideran para el Pre-mapeo inicial 46 índices de Renta Fija, caracterizados por duración. La Tabla IV-1 permite ver una breve descripción de cada uno de los índices utilizados para la representación.

Tabla IV- 1: Descripción Índices Renta Fija Nacional Seleccionados

Familia	Nemotécnico	Descripción
GLOBAL	RACL	Índice constituido por todos papeles de Renta Fija del mercado Chileno
GOBIERNO	RACLGOB	Índice constituido por todos los papeles emitidos por Instituciones del Gobierno de Chile. Incluye a los siguientes papeles: PRC,BCP,BCU, BTU,PDBC,CERO y BTP
	RACLGOB_D0	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración inferior a 6 meses
	RACLGOB_D1	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 6 meses y 1.5 años
	RACLGOB_D2	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 1.5 y 2.5 años
	RACLGOB_D3	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 2.5 y 3.5 años
	RACLGOB_D4	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 3.5 y 4.5 años
	RACLGOB_D5	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 4.5 y 5.5 años
	RACLGOB_D6	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 5.5 y 6.5 años
	RACLGOB_D7	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 6.5 y 7.5 años
	RACLGOB_D8	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 7.5 y 8.5 años
	RACLGOB_D9	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración entre 8.5 y 9.5 años
	RACLGOB_D10P	Índice constituido por todos los papeles emitidos por Instituciones de Gobierno, con duración mayor a 9.5 años
	CORPORATIVO	RACLCO
RACLCO_D1		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 6 meses y 1.5 años
RACLCO_D2		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 1.5 y 2.5 años
RACLCO_D3		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 2.5 y 3.5 años
RACLCO_D4		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 3.5 y 4.5 años
RACLCO_D5		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 4.5 y 5.5 años
RACLCO_D6		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 5.5 y 6.5 años
RACLCO_D7		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 6.5 y 7.5 años
RACLCO_D8		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 7.5 y 8.5 años
RACLCO_D9		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración entre 8.5 y 9.5 años
RACLCO_D10P		Índice constituido por todos los Bonos de empresas, bancos y securitizadoras, con duración mayor a 9.5 años

LETRAS HIPOTECARIAS	RACLLH	Índice constituido por todas las Letras Hipotecarias del mercado Chileno
	RACLLH_D1	Índice constituido por todas las Letras Hipotecarias, con duración entre 6 meses y 1.5 años
	RACLLH_D2	Índice constituido por todas las Letras Hipotecarias, con duración entre 1.5 y 2.5 años
	RACLLH_D3	Índice constituido por todas las Letras Hipotecarias, con duración entre 2.5 y 3.5 años
	RACLLH_D4	Índice constituido por todas las Letras Hipotecarias, con duración entre 3.5 y 4.5 años
	RACLLH_D5	Índice constituido por todas las Letras Hipotecarias, con duración entre 4.5 y 5.5 años
	RACLLH_D6	Índice constituido por todas las Letras Hipotecarias, con duración entre 5.5 y 6.5 años
	RACLLH_D7	Índice constituido por todas las Letras Hipotecarias, con duración entre 6.5 y 7.5 años
	RACLLH_D8	Índice constituido por todas las Letras Hipotecarias, con duración entre 7.5 y 8.5 años
	RACLLH_D9P	Índice constituido por todas las Letras Hipotecarias, con duración mayor a 8.5 años
INTERMEDIACION FINANCIERA	RACLIF	Índice constituido por todos los Depósitos y Efectos de Comercio del mercado Chileno
	RACLIF_DM0	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 1 y 15 días
	RACLIF_DM1	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 16 y 45 días
	RACLIF_DM2	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 46 y 75 días
	RACLIF_DM3	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 76 y 105 días
	RACLIF_DM6	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 136 y 225 días
	RACLIF_DM12	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 271 y 450 días
	RACLIF_D1	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 1.5 y 2.5 años
	RACLIF_D2	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración entre 2.5 y 3.5 años
	RACLIF_D3P	Índice constituido por todos los Depósitos y Efectos de Comercio, con duración mayor a 3.5 años
BONOS DE RECONOCIMIENTO	RACLBR	Índice constituido por todos los Bonos de Reconocimiento del mercado Chileno.
	RACLBR_DC	Índice constituido por todos los Bonos de Reconocimiento, con duración menor a 3 años
	RACLBR_DL	Índice constituido por todos los Bonos de Reconocimiento, con duración mayor a 3 años

IV.2 Resultados

IV.2.1 Pruebas realizadas

En esta sección, se busca mostrar el uso de la metodología aplicándola a un mercado emergente, seleccionando para este fin, el mercado de Renta Fija Nacional y su cartera representada por los saldos nominales en las custodias del DCV el día 30 de Junio de 2006.

Se elijen distintas representaciones de la cartera inicial, variando tanto el número de índices k con que se desea representar la cartera, como el esfuerzo computacional que se requiere en encontrar una solución (n° iteraciones).¹⁵

Finalmente, se evalúan las diferencias en comportamiento de las distintas representaciones y se comparan con la representación intuitiva que se generaría con el método tradicional.

Como conjunto de índices representativos, se escoge el segmento de índices de renta fija, caracterizados por duración. Se considera un período de 2 años de datos diarios para cada uno de los índices, abarcado el periodo entre el 01 de Julio de 2004 y el 30 de Junio de 2006, lo cual implica 504 datos de retornos diarios para cada índice, para el análisis *in sample*. Se considera además, un período de 6 meses de datos diarios para cada uno de los índices, abarcado el periodo entre el 01 de Julio de 2006 y el 29 de Diciembre de 2006, para el análisis *out of sample*.

Se realizan dos pruebas, para cada una de las cuales se busca representar 3 carteras de distinta dimensionalidad ($k= 3, 5$ y 8).

La primera prueba consiste en buscar la mejor representación para la cartera mediante el uso de variables compuestas. La segunda prueba consiste en representar la cartera utilizando variables puras, es decir, seleccionando específicamente índices dentro del universo inicial.

¹⁵ Para ver en detalle como seleccionar k y número de iteraciones adecuado, ver Salas (2007)

A continuación se presentan cada una de las pruebas realizadas y los principales resultados obtenidos en cada una de ellas.

a) Representación Mediante Variables Compuestas

En esta sección se busca representar la cartera mediante la selección de variables compuestas, es decir, se busca la mejor representación utilizando para este fin Componentes Principales.

El primer paso de la metodología consiste en asignar a cada activo de la cartera, tenga éste precios diarios o no, un índice representativo. En el Anexo B se muestra la composición, instrumento por instrumento, de la cartera del DCV y la asignación de índices generada en base a los datos disponibles.

Siguiendo los criterios establecidos anteriormente, si se desea representar la cartera mediante un conjunto de índices “Compuestos”, es decir, combinaciones de varios de los índices iniciales, se debe utilizar la metodología de componentes principales.

El Anexo C posee información detallada de la serie de datos de cada componente y su composición.

La Tabla IV-2 muestra el porcentaje de la varianza acumulada explicado por cada set de componentes principales:

Tabla IV- 2: % de la Varianza explicada por los componentes principales

De los resultados obtenidos, es posible inferir que los 3 primeros factores permiten explicar el 76,93% de la varianza total del set de datos. De la misma forma con 5 u 8 primeros componentes principales, explican el 85,45% y 91,22% de la varianza total del conjunto de datos, respectivamente.

Sin embargo, como se señaló en el Capítulo 2¹⁶, si se busca representar una serie de datos específica y además el objetivo prioritario de la representación es lograr el mejor ajuste posible, la selección de los componentes que se debe utilizar debe ser hecha seleccionando aquellos más significativos para la serie.

Para lograr este objetivo, es necesario seguir con la metodología, es decir se debe regresionar todos los componentes principales vs la serie Pre-mapeada, calculando además, los t estadísticos para cada PC.

El Anexo D muestra el resultado de la regresión y el estadístico t para cada uno de los PCs.

Siguiendo con la metodología es necesario ordenar los PCs de acuerdo a su estadístico t en orden decreciente con el fin de seleccionar sólo aquellos componentes que más aportan a la regresión.

La Tabla IV-3 muestra los primeros 10 componentes en orden decreciente según su estadístico t

Tabla IV- 3: 10 PCs con mayor estadístico t

		<i>Estadístico t</i>
1	PC1	48100,0
2	PC3	8872,0
3	PC2	7816,8
4	PC5	3720,9
5	PC7	3453,0
6	PC10	3381,1
7	PC6	3262,0
8	PC25	2903,2
9	PC15	1890,9
10	PC14	1808,3

¹⁶ Capítulo 2, Sección II.1.2 “Estrategias para seleccionar componentes en una regresión utilizando componentes principales”

Confirmando la idea presentada en el Capítulo 2, la selección de componentes principales que son más significativos en la regresión no siempre coincide con la selección de aquellos componentes que más aportan a la varianza total de los datos.

En el caso de la primera selección, es decir cuando, represento la cartera con 3 PCs ambos objetivos se alinean y la selección de los componentes más significativos es la misma de la de los más variables. La Figura 4-3 muestra la composición de la cartera utilizando 3 PCs

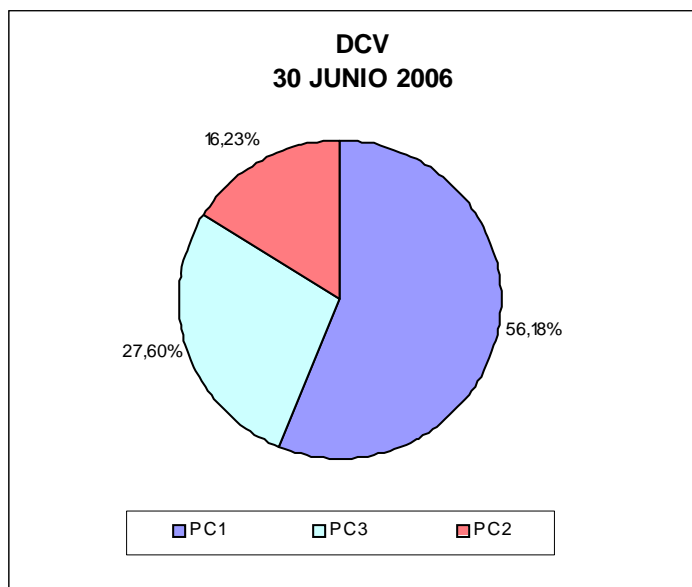


Figura 4-3: Composición cartera con 3 PCs

Con esta representación es posible lograr un ajuste de R^2 de 96.62%.

Por otra parte, en el caso de la segunda selección, es decir cuando, represento la cartera con 5 PCs los objetivos no son alcanzables de manera simultánea y la selección de los componentes más significativos, no es la misma de la de los más variables. La Tabla IV-3 muestra como los 5 PCs más significativos, son en orden decreciente, el primero, el tercero, el segundo, el quinto y el séptimo. Para que ambos objetivos estuviesen alineados, se debería seleccionar al cuarto PC en vez del séptimo. La Figura 4-4 muestra la composición de la cartera utilizando 5 PCs.

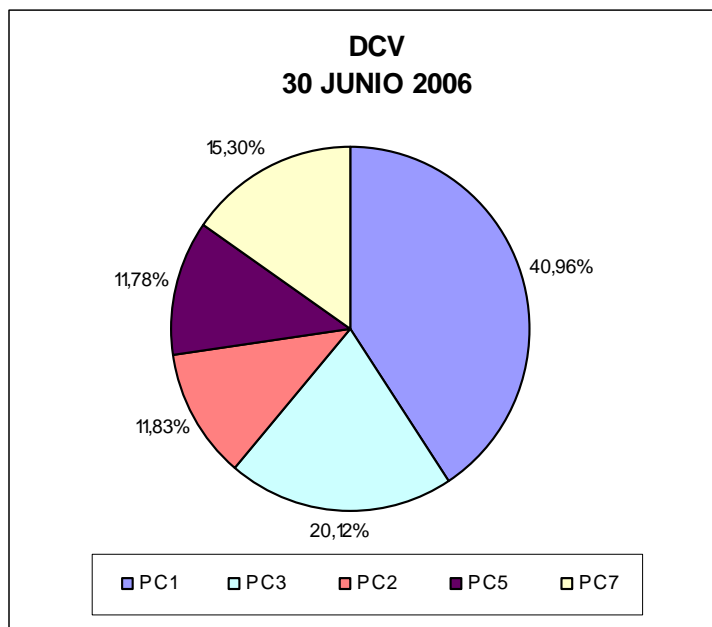


Figura 4-4: Composición cartera con 5 PCs

Con esta representación es posible lograr un ajuste de R^2 de 97.64%.

Mientras más aumenta el número de variables seleccionadas, mas difícil resulta, en este ejemplo, conciliar los objetivos. En el último caso de selección, cuando represento la cartera con 8 PCs claramente los objetivos no son alcanzables de manera simultánea. La Tabla V-3 muestra como los 8 PCs más significativos, incluyen incluso al componente 25. Con esta representación es posible lograr un ajuste de R^2 de 98.84%.

La Figura 4-5 muestra la composición de la cartera utilizando 8 PCs.

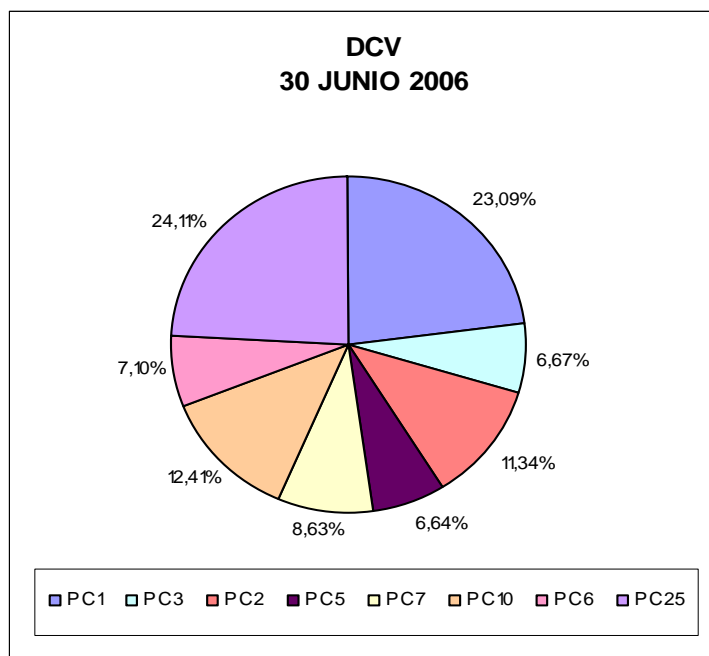


Figura 4-5: Composición cartera con 8 PCs

b) Representación Mediante Variables Puras

La segunda prueba consiste en representar la cartera de RF del DCV, mediante la selección de variables originales.

Siguiendo la metodología propuesta, se debe seleccionar un conjunto de índices puros, para lo cual, primero es necesario realizar la clasificación “global/específico” de los índices. Observando el mapa presentado en la Figura 4-1, claramente se pueden distinguir 6 índices globales: Renta Fija Global, Gobierno Global, Corporativo Global, Letras Hipotecarias Global, Intermediación Financiera Global y Bonos de Reconocimiento Global. Los 40 índices restantes son considerados subconjuntos de alguno de estos índices globales y por tanto son clasificados como índices específicos.

El paso siguiente implica analizar el número total de combinaciones posibles para el set de activos inicial de la cartera en cuestión. La Figura 4-6 muestra la variación del número de combinaciones y por tanto del número de iteraciones a realizar, en función

del número de índices que se desea utilizar para la representación, considerando 46 índices iniciales

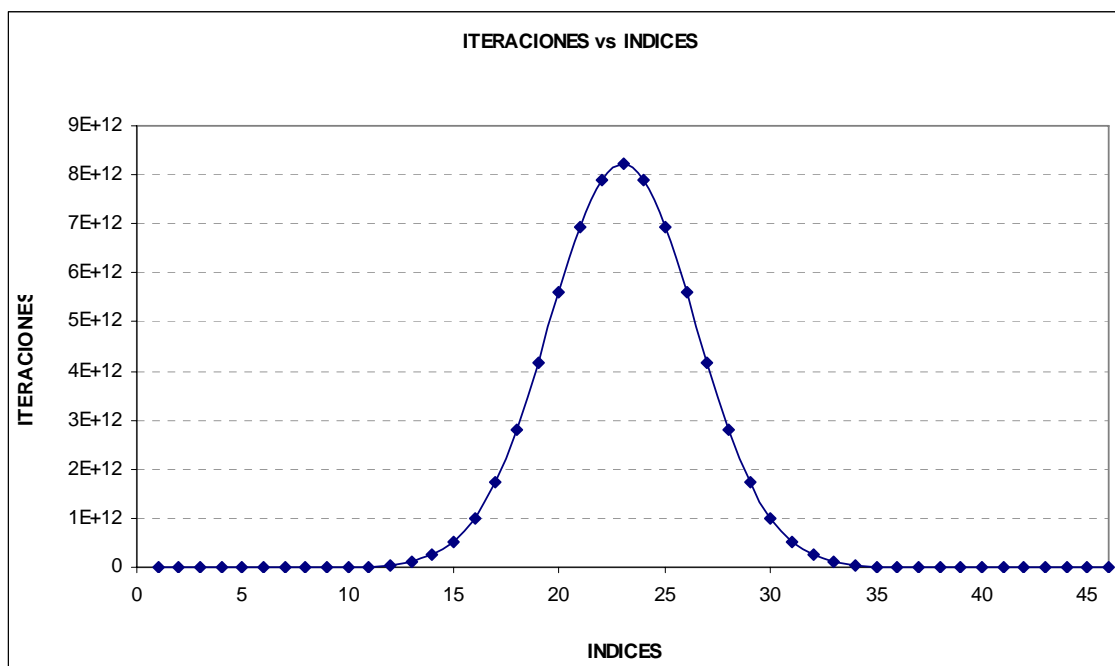


Figura 4-6: N° de iteraciones vs N° de índices

Para representar la cartera con 3 índices, es necesario analizar 1.518 combinaciones y por lo tanto implica realizar 1.518 iteraciones. Sin embargo, si la representación desea ser realizada con 5 y 8 índices, el número de combinaciones a analizar aumenta explosivamente, llegando a 1.370.800 y 260.930.000 iteraciones respectivamente, alcanzando un máximo de 8.233.400.000.000 iteraciones cuando se desea representar la cartera con 23 índices, exactamente la mitad del universo disponible.

Continuando con la metodología, es necesario analizar si el número de iteraciones a realizar será suficiente para revisar todas las combinaciones posibles, o si será necesario hacer un proceso de reducción mediante el método Backward.

Para $k=3$ el número de iteraciones es razonable, por lo que se realiza la inspección de todas las combinaciones sin reducción del universo inicial. A continuación se analizan todas las combinaciones posibles seleccionando aquella que entregue el mayor ajuste.

El algoritmo desarrollado, entrega tanto los índices seleccionados como sus ponderaciones dentro de la cartera. Además entrega información acerca del intercepto (Alfa) y del ajuste conseguido con la solución. La Tabla IV-4 presenta la información del ajuste para $k=3$.

Tabla IV- 4: Ajuste para $k=3$

Nº Índices	3
Iteraciones	15180
Alfa	0,00000418
Σe^2	0,00000781
R^2	98,72%

El valor de R^2 es de 98,72%, es decir la representación permite explicar casi un 99% de la varianza de la serie.

Es importante señalar que el valor de alfa es estadísticamente no significativo, es decir sólo se utiliza como ajuste dentro de la muestra pero no es una variable utilizada cuando se busca una representación con fines predictivos. La variable alfa presenta un estadístico t de 0,74, valor muy inferior al valor t crítico, de dos colas, con 503 grados de libertad y $\alpha = 0,05$ de 1,965.

La Figura 4-7 presenta la composición de la cartera solución seleccionada como óptima.

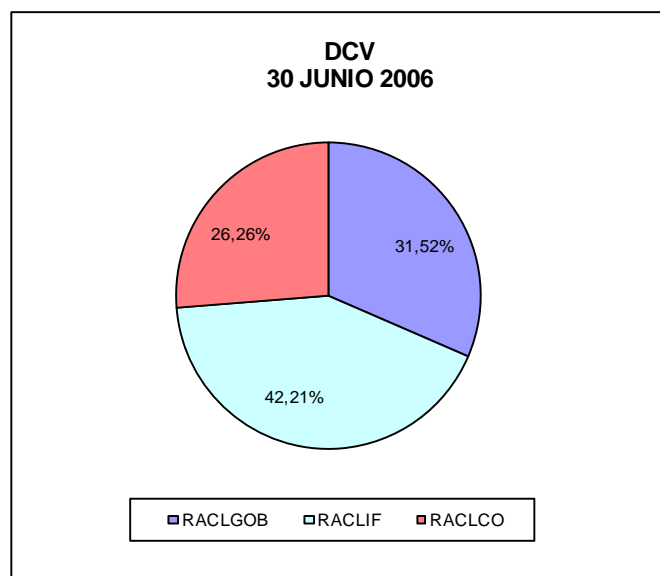


Figura 4-7: Composición cartera óptima para $k=3$

La solución obtenida parece razonable e intuitiva. La selección de 3 índices globales parece ser la solución más sensata cuando el universo total desea ser representado por un número pequeño de índices, ya que éstos engloban la información de los índices específicos.

Para analizar en forma detallada el comportamiento de la solución y su similitud en relación a los datos originales, se generan series que muestran el comportamiento de la cartera “Pre-mapeada” (compuesta por todos los índices) y la “Mapeada” (cartera solución de 3 índices) Las Figuras 4-8 y 4-9 y las Tablas IV-5 y IV-6 permiten analizar el comportamiento de ambas series tanto dentro del periodo de análisis como fuera de este.

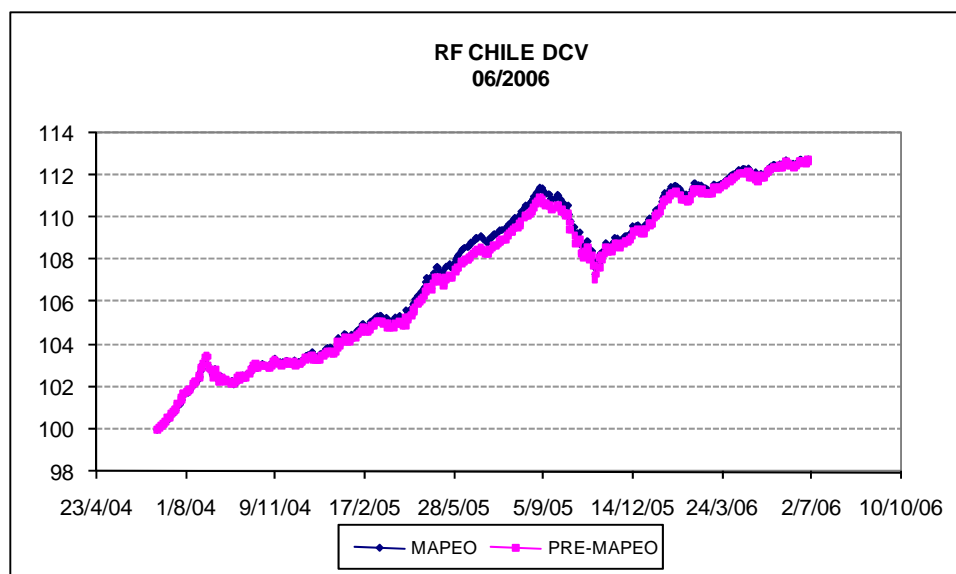


Figura 4-8: Series In-sample

Tabla IV- 5: Análisis estadístico Series In-sample

		PRE-MAPEO	MAPEO
TOTAL	σ	1,754%	1,752%
	μ	6,224%	6,224%
AÑO 1	σ	1,636%	1,585%
	μ	8,510%	9,078%
AÑO 2	σ	1,858%	1,893%
	μ	3,986%	3,444%

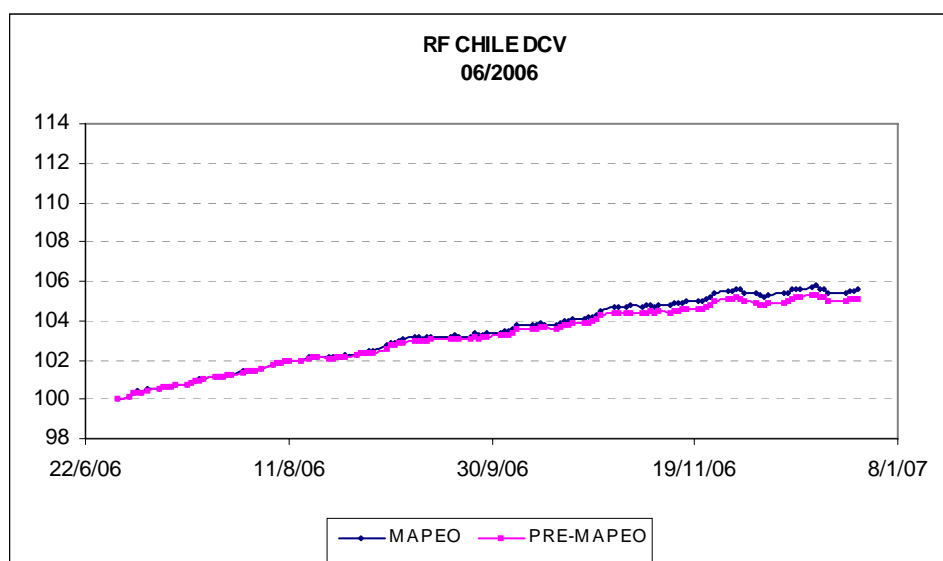


Figura 4-9 : Series Out-of-sample

Tabla IV- 6: Análisis estadístico Series Out-of-sample

		PRE-MAPEO	MAPEO
TOTAL	σ	1,039%	1,108%
	μ	10,853%	11,752%

Finalmente, y con el fin de validar el resultado obtenido se analiza la multicolinealidad de la solución. Para $k=3$ se obtiene un $VIF=5,41$ menor que 10, es decir la solución no presenta problemas de multicolinealidad, por lo tanto, la varianza de los coeficientes estimados es baja, con lo que la solución es considerada como válida para el análisis¹⁷.

Prosiguiendo con las pruebas, se analiza la solución para $k=5$. En este caso, el número de iteraciones necesarias para analizar todas las combinaciones posibles no es razonable, por lo que se debe hacer un proceso de reducción backward previo a la inspección del universo. Con el fin de obtener distintas soluciones, se fijan 4 niveles de iteraciones diferentes, lo que implicará una reducción mayor o menor del universo inspeccionado según el número de iteraciones fijada. A continuación se analizan todas las combinaciones posibles dentro del universo reducido, seleccionando aquella que entregue el mayor ajuste.

Tabla IV- 7: Ajuste para $k=5$

Nº índices	5
Iteraciones	30000
	80000
	250000
	500000
Alfa	0,00000147
Σe^2	0,00000165
R^2	99,73%

Para todos los niveles de iteraciones fijados se obtiene la misma solución. El valor de R^2 es de 99,73%, es decir la representación permite explicar casi el 100% de la varianza de la serie.

¹⁷ $VIF:1/1-R^2$. Factor de inflación de la varianza. Un VIF mayor que 10 puede implicar multicolinealidad.

Al igual que en la solución anterior, el valor de alfa es estadísticamente no significativo. La variable alfa presenta un estadístico t de 0,467, valor muy inferior al valor t crítico, de dos colas, con 503 grados de libertad y $\alpha = 0,05$ de 1,965.

La Figura 4-10 presenta la composición de la cartera solución seleccionada como óptima.

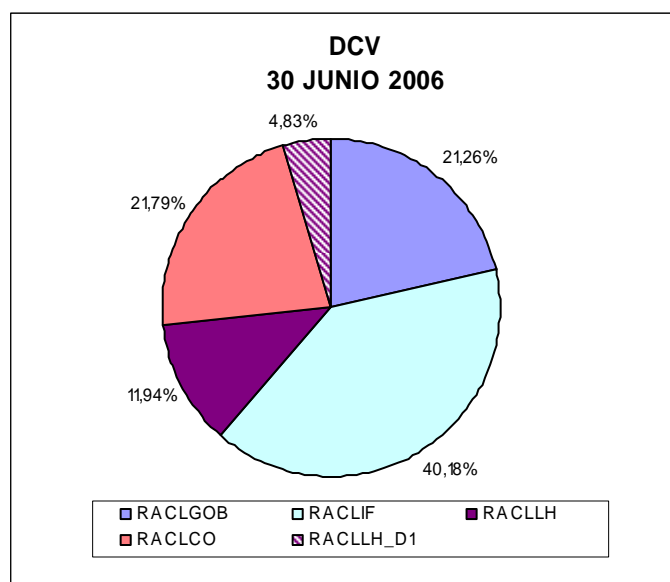


Figura 4-10: Composición cartera óptima para k=5

La solución obtenida parece razonable. La selección de 4 índices globales y uno específico, parece ser una solución sensata, ya que cuando el universo total desea ser representado por un número pequeño de índices, la intuición hace pensar que la optimalidad se encuentra en la representación en gran medida, a través de índices globales.

Por otra parte, el ajuste alcanzado es superior en aproximadamente un 1%, al obtenido si se utilizase la representación intuitiva, es decir asignando a cada familia un índice global representativo en el porcentaje real de la cartera en la familia respectiva, que corresponde a un R^2 de 98,61%.

Al igual que en el ejemplo anterior, se generan series que muestran el comportamiento de la cartera “Pre-mapeada” y la “Mapeada”. Las Figuras 4-11 y 4-12 y las Tablas IV-8 y IV-9 permiten analizar el comportamiento de ambas series tanto dentro del periodo de análisis como fuera de este.

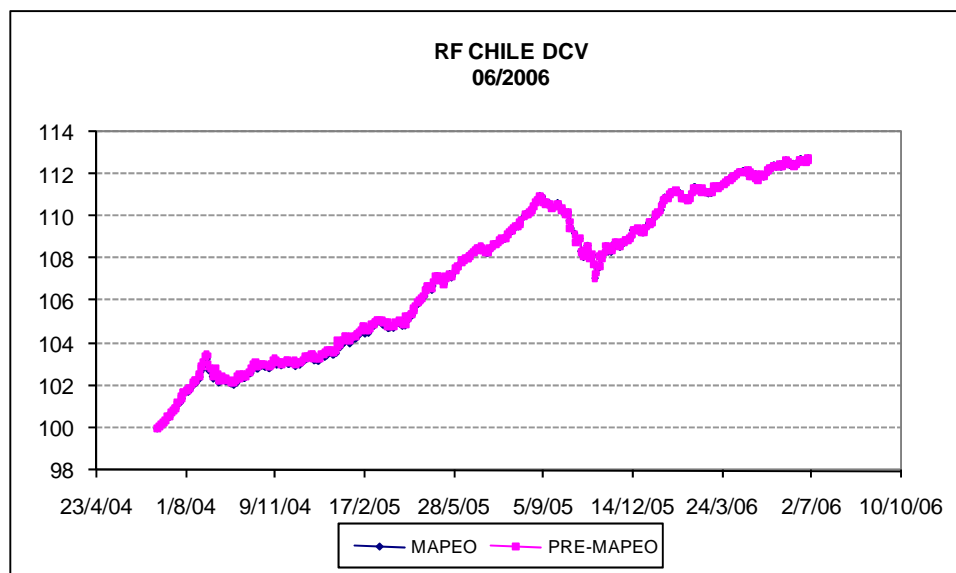


Figura 4-11: Series In sample

Tabla IV- 8: Análisis estadístico Series In sample

		PRE-MAPEO	MAPEO
TOTAL	σ	1,754%	1,757%
	μ	6,224%	6,224%
AÑO 1	σ	1,636%	1,611%
	μ	8,510%	8,525%
AÑO 2	σ	1,858%	1,885%
	μ	3,986%	3,971%

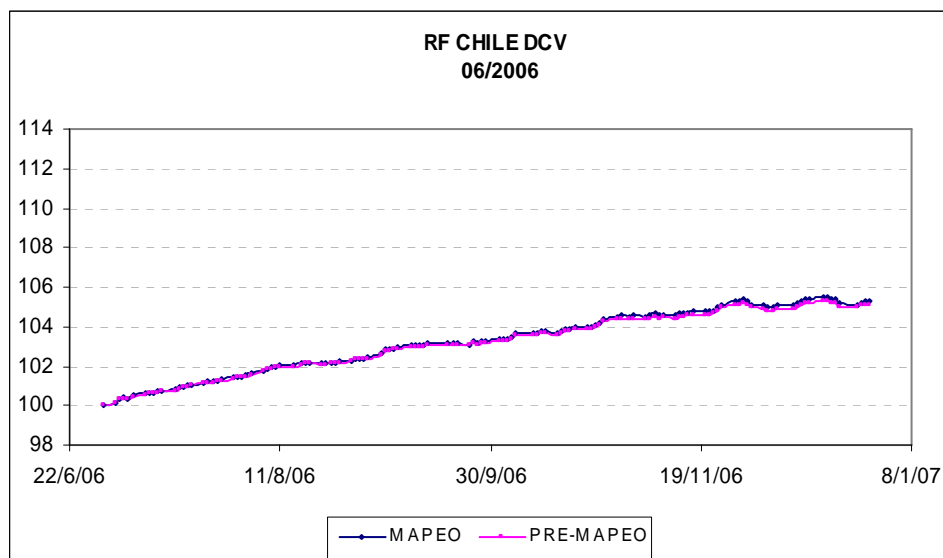


Figura 4-12: Series Out of sample

Tabla IV- 9: Análisis estadístico Series Out of sample

		PRE-MAPEO	MAPEO
TOTAL	σ	1,039%	1,089%
	μ	10,853%	11,223%

Finalmente, y con el objetivo de validar el resultado obtenido se analiza la multicolinealidad de la solución. Para $k=5$ se obtiene un $VIF=6,95$ menor que 10, con lo que la solución es considerada como válida para el análisis.

Para finalizar, se analiza la solución para $k=8$. En este ejemplo, al igual que en el caso anterior, el número de iteraciones necesario para analizar todas las combinaciones posibles no es razonable, por lo que se debe hacer un proceso de reducción backward previo a la inspección del universo. Se fijan los mismos 4 niveles de iteraciones que para $k=5$. A continuación se analizan todas las combinaciones posibles dentro del universo reducido, seleccionando aquella que entregue el mayor ajuste.

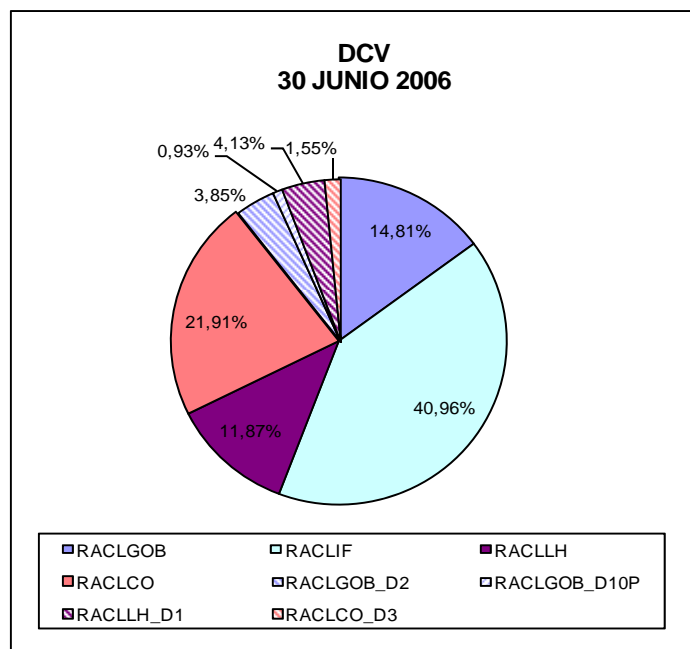
Tabla IV- 10: Ajuste para k=8

Nº Índices	8
Iteraciones	30000
	80000
	250000
	500000
Alfa	0,00000034
Σe^2	0,00000129
R ²	99,79%

Para todos los niveles de iteraciones fijados se obtiene la misma solución. El valor de R² es de 99,79%, es decir la representación permite explicar casi el 100% de la varianza de la serie.

Al igual que en la solución anterior, el valor de alfa es estadísticamente no significativo. La variable alfa presenta un estadístico t de 0,154, valor muy inferior al valor t crítico, de dos colas, con 503 grados de libertad y alfa = 0,05 de 1,965

La Figura 4-13 presenta la composición de la cartera solución seleccionada como óptima.

**Figura 4-13: Composición cartera óptima para k=8**

La solución obtenida parece razonable. Como es posible ver en la Figura 4-1, los papeles que tiene mayor peso en la cartera real están representados tanto por índices globales como por descomposiciones de aquellos más significativos en la cartera de ellos, en la medida en que aumenta el número de índices en la representación.

Al igual que en los casos anteriores, se generan series que muestran el comportamiento de la cartera “Pre-mapeada” y la “Mapeada”. Las Figuras 4-14 y 4-15 y las Tablas IV-11 y V-12 permiten analizar el comportamiento de ambas series tanto dentro del periodo de análisis como fuera de este.

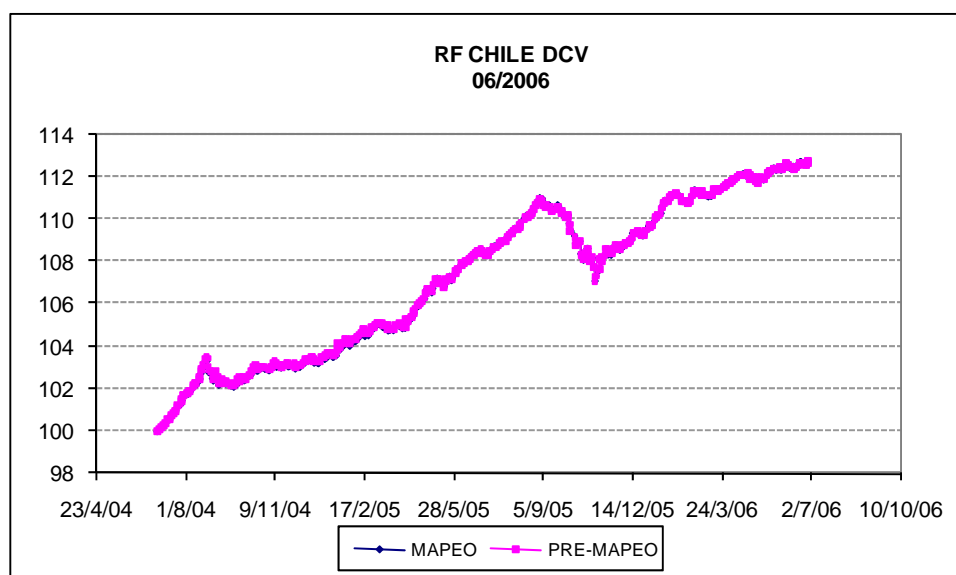


Figura 4-14: Series In-sample

Tabla IV- 11: Análisis estadístico Series In-sample

		PRE-MAPEO	MAPEO
TOTAL	σ	1,754%	1,760%
	μ	6,224%	6,224%
AÑO 1	σ	1,636%	1,623%
	μ	8,510%	8,537%
AÑO 2	σ	1,858%	1,880%
	μ	3,986%	3,960%

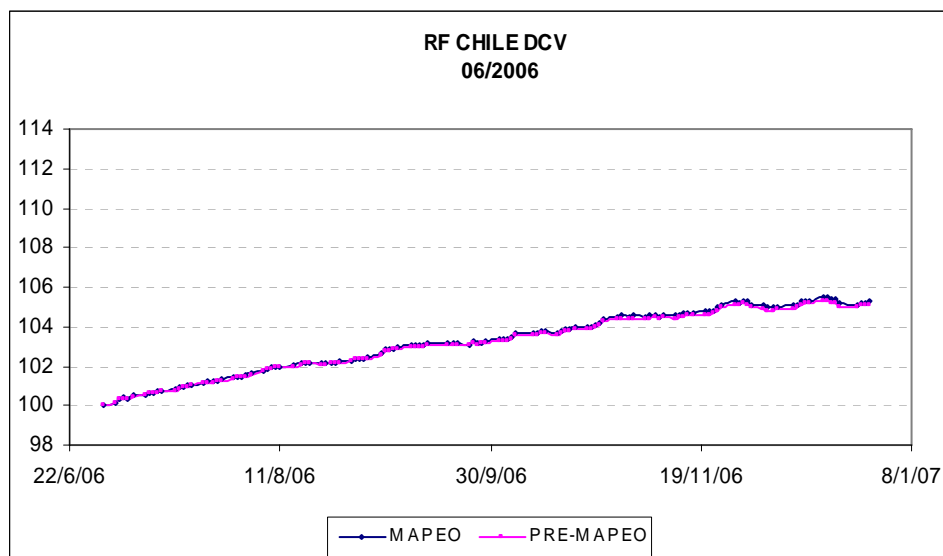


Figura 4-15: Series Out-of-sample

Tabla IV-12: Análisis estadístico Series Out-of-sample

		PRE-MAPEO	MAPEO
TOTAL	σ	1,039%	1,083%
	μ	10,853%	11,169%

Finalmente, y con el fin de validar el resultado obtenido se analiza la multicolinealidad de la solución. Para $k=8$ se obtiene un $VIF=23$ mayor que 10, con lo que la solución puede presentar el problema de multicolinealidad debiendo tener precaución con la confiabilidad de los ponderadores estimados por la metodología. Si se quiere evitar o reducir el problema, se puede elegir la representación utilizando componentes principales.

Como una forma de resumir los resultados obtenidos la Tabla IV-13 muestra la composición de cada representación en conjunto con la representación de la cartera hecha en el Pre- Mapeo.

Tabla IV-13: Análisis estadístico Series Out-of-sample

	PRE MAPEO	k=3	k=5	k=8
RACL	0,000%	-	-	-
RACLG0B	0,000%	31,52%	21,26%	14,81%
RACLBR	0,000%	-	-	-
RACLIF	0,000%	42,21%	40,18%	40,96%
RACLLH	0,000%	-	11,94%	11,87%
RACLCO	0,000%	26,26%	21,79%	21,91%
RACLG0B_D0	5,253%	-	-	-
RACLG0B_D1	5,354%	-	-	-
RACLG0B_D2	5,074%	-	-	3,85%
RACLG0B_D3	2,615%	-	-	-
RACLG0B_D4	1,008%	-	-	-
RACLG0B_D5	0,861%	-	-	-
RACLG0B_D6	1,090%	-	-	-
RACLG0B_D7	0,723%	-	-	-
RACLG0B_D8	0,608%	-	-	-
RACLG0B_D9	0,076%	-	-	-
RACLG0B_D10P	2,246%	-	-	0,93%
RACLBR_DC	0,091%	-	-	-
RACLBR_DL	0,093%	-	-	-
RACLIF_DM0	3,228%	-	-	-
RACLIF_DM1	6,245%	-	-	-
RACLIF_DM2	3,919%	-	-	-
RACLIF_DM3	5,166%	-	-	-
RACLIF_DM6	6,728%	-	-	-
RACLIF_DM12	8,618%	-	-	-
RACLIF_D2	1,788%	-	-	-
RACLIF_D3P	1,369%	-	-	-
RACLLH_D1	1,870%	-	4,83%	4,13%
RACLLH_D2	0,733%	-	-	-
RACLLH_D3	1,377%	-	-	-
RACLLH_D4	2,480%	-	-	-
RACLLH_D5	2,345%	-	-	-
RACLLH_D6	1,790%	-	-	-
RACLLH_D7	1,727%	-	-	-
RACLLH_D8	1,388%	-	-	-
RACLLH_D9P	0,535%	-	-	-
RACLCO_D1	2,220%	-	-	-
RACLCO_D2	1,226%	-	-	-
RACLCO_D3	3,068%	-	-	1,55%
RACLCO_D4	1,378%	-	-	-
RACLCO_D5	1,185%	-	-	-
RACLCO_D6	1,315%	-	-	-
RACLCO_D7	2,502%	-	-	-
RACLCO_D8	2,192%	-	-	-
RACLCO_D9	2,081%	-	-	-
RACLCO_D10P	6,436%	-	-	-

IV.2.2 Discusión de los Resultados

En la primera parte de este ejemplo se obtiene la mejor representación para la cartera de los saldos nominales del DCV del mercado de Renta Fija Nacional, a través del método de Componentes Principales, logrando capturar el 97.64% de la varianza total de la cartera Pre-Mapeada para la representación mayor de $k=8$.

Además se reafirma la idea presentada en el Capítulo 2 de la discordancia entre el objetivo de lograr un buen ajuste en la representación y lograr explicar el mayor porcentaje de la varianza de los datos originales evitando así el problema de multicolinealidad.

Sin embargo, aún si la selección de un conjunto de variables “Compuestas” se hace en base al criterio del mayor ajuste, no necesariamente será mejor que la obtenida utilizando variables “Puras”.

Llevando el análisis al plano del problema en cuestión, la utilización de PCs que explican un alto % de la varianza total de los datos para representar la serie Pre-Mapeada presenta el problema de no incorporar la información de la composición de la cartera en la selección de los componentes. De esta forma, si por ejemplo, una cartera estuviese compuesta casi en un 100 % por sólo uno de los índices, y éste no tuviese una alta ponderación en el primer componente principal o en al menos alguno de los componentes, esta cartera quedaría mejor representada por el índice puro, que por el PC.

Así mismo, en la representación de la cartera utilizando 8 variables “Puras” el ajuste obtenido es de 99.79%, mayor que el obtenido utilizando PCs.

En la segunda parte, se busca la representación de la cartera específica del DCV, a través del algoritmo que busca la mejor solución cuyo criterio de selección es el ajuste medido con la suma de errores al cuadrado de la regresión.

Para una representación con 3 activos, el número de combinaciones a analizar, si se desea inspeccionar todas las combinaciones factibles, es considerado como razonable, y por lo tanto, no es necesaria la reducción del universo inicial a través del método

Backward. El resultado obtenido para esta representación es satisfactorio, ya que, por una parte permite alcanza un nivel de ajuste de $R^2=98,72\%$ y además satisface la intuición de la importancia de la utilización de índices globales cuando la representación es pequeña.

En términos del comportamiento de la serie obtenida con la representación y la general, es posible concluir que la solución obtenida permite lograr un excelente ajuste, tanto in-sample (retornos promedio de serie Pre-mapeada y serie Mapeada muy similares con 1,754% y 1,752% anual; y volatilidades de 6,224% anual para ambas series), como out-of-sample (retornos promedio de serie Pre-mapeada y serie Mapeada muy similares con 1,039% y 1,108% anual; y volatilidades de 10,853% y 11,752% anual respectivamente).

Por otra parte, para una representación con 5 y 8 activos, el número de combinaciones a analizar si se desea inspeccionar todas las combinaciones factibles, es considerado como poco razonable, y por lo tanto, es necesaria la reducción del universo inicial a través del método Backward. Los resultados obtenidos para estas representaciones son igualmente satisfactorios, ya que, por una parte permiten alcanzar niveles de ajuste de $R^2=99,73\%$ y 99.79%, y son capaces de satisfacer la intuición de la importancia de la utilización de índices globales cuando la representación es pequeña y el desglose de éstos en índices más específicos cuando aumenta la dimensionalidad de la representación.

En términos del comportamiento de la serie obtenida con la representación y la general, es posible concluir que las soluciones obtenidas permiten alcanzar excelentes ajustes, tanto in- simple, como out-of-sample.

Es importante mencionar, que a pesar de que el aumento de la dimensionalidad de la representación, mejora el ajuste, implica la posibilidad de existencia del problema de multicolinealidad, sobretodo en índices tan altamente correlacionados como los de Renta Fija Nacional (en términos generales se ven afectados por los mismos factores, hecho que se reafirma con los resultados obtenidos en los Componentes Principales, que indican que es posible explicar un alto porcentaje de la varianza total con unos pocos factores).

Para reducir el problema de la multicolinealidad, es recomendable utilizar la representación utilizando componentes principales. Sin embargo, es importante recalcar que esta representación es recomendable en la medida en que el usuario no deba realizar análisis que dependan de la interpretación de las variables utilizadas en la representación.

Finalmente y en términos generales, es posible concluir que la metodología permite obtener resultados que son capaces de representar de manera fidedigna para una cartera de renta fija chilena.

V. CONCLUSIONES GENERALES

En primer lugar, se debe señalar que existe una gran cantidad de mercados en los cuales es posible aplicar la metodología propuesta, ya que incluso en mercados completos, donde existen transacciones para todos los papeles, todos los días, el poder representar de forma adecuada una cartera de inversión, con el fin de realizar diversos análisis, es fundamental, desde la perspectiva de la robustez y confianza de los resultados obtenidos.

Por otra parte, se observa que la investigación desarrollada abarca muchas ramas de desarrollo diferentes, ya que pese a que al parecer el problema en cuestión no ha sido abordado antes de forma específica, se han desarrollado diversas metodologías, para resolver el problema de encontrar un subset óptimo desde un universo objetivo.

Dichas metodologías se centran en áreas tan diversas como los análisis de espacios muestrales, los análisis heurísticos y los análisis estadísticos, planteando todas, caminos diferentes para resolver un mismo problema.

Debido a las características propias de cada metodología, cada una de ellas ha sido aplicada a diversas áreas de investigación y modelación, siendo utilizadas en áreas tan diversas como la mecánica, la estadística, el transporte, la logística, etc.

La metodología elegida, por su parte, es simple en términos de implementación y entrega respuestas satisfactorias, con tiempos de ejecución razonables y a través de conceptos sencillos. Además, es una metodología flexible que permite al usuario representar carteras de distinta dimensionalidad, entregando además, la posibilidad de limitar el tiempo de ejecución que el usuario está dispuesto a utilizar y el tipo de variable con la que desea realizar la representación.

Los resultados obtenidos por la metodología serán cada vez más apropiados, en la medida que el usuario sea capaz de analizar las implicancias de la dimensionalidad y del esfuerzo computacional en el mercado en que se desarrolla el problema¹⁸.

Finalmente, es posible concluir que uno de los principales aportes de la investigación, radica en el hecho de abordar de manera formal un problema fundamental y de gran importancia en el área de inversión, como es el hecho de representar una cartera de activos, problema que hasta el día de hoy sólo era realizado en forma intuitiva por los analistas financieros. El hecho de implementar una metodología simple, que busca lograr resultados basados en análisis de reducción de variables en espacios muestrales, significa un gran avance en un área, donde se busca cada vez con más fuerza, la introducción de herramientas basadas en análisis empíricos con el objeto de mejorar y por sobretodo formalizar los procesos de inversión.

¹⁸ Para ver las implicancias de la dimensionalidad del problema y la apropiada selección del tamaño de la representación y el tiempo de ejecución, ver Salas (2007)

BIBLIOGRAFIA

- BERK K.N. (1978). Comparing subset regression procedures. *Technometrics*, 20, 1-6.
- BREIMAN L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37, No. 4, 373-384.
- BROERSEN P.M.T. (1986). Subset regression with stepwise directed search. *Applied Statistics*, 35, No. 2, 168-177.
- CHARON I. y HUDRY O. (1996). Mixing Different Components of Metaheuristics. En I. H. Osman and J. P. Kelly (Eds.), *Meta-Heuristics: Theory and Applications* (1a.ed, pp.589-603). Boston, EEUU: Kluwer Academic Publishers.
- CORTÁZAR, G., SCHWARTZ, E.S. AND NARANJO, L.F. (2007). Term Structure Estimation in Markets with Infrequent Trading, *International Journal of Finance and Economics* (forthcoming).
- DENARDO E. V. y FOX B. L. (1979). Shortest Route Methods: 2 Group Knapsacks, Expanded Networks and Branch and Bound. *Operations Research*, 27, 548-566.
- DRAPER N.R. & SMITH H. (1981). *Applied Regression Analysis* (2a.ed.). New York, EEUU: John Wiley.
- FISHER H. y THOMPSON G. L. (1963). Probabilistic Learning Combinations of Local Job Shop Scheduling Rules. En J.F. Muth and G.L. Thompson (Eds.), *Industrial Scheduling*. (1a. ed, pp.225-251). New Jersey, EEUU: Prentice-Hall.
- GENDREAU M., LAPORTE G. y POTVIN J-Y. (1995). Metaheuristics for the Vehicle Routing Problem. En J. K. Lenstra and E. H. L. Aarts (Eds.). *Local Search Algorithms* New York, EEUU: John Wiley & Sons
- GLOVER F. (1963). Parametric Combinations of Local Job Shop Rules. Capítulo IV, ONR Research Memorandum N° 117, GSIA, Carnegie-Mellon University, Pittsburgh, PA.
- GLOVER F. (1965). A Multiphase-dual Algorithm for the Zero-one Integer Programming Problem. *Operations Research*, 13, 879-919.
- GLOVER F. (1966). An Algorithm for Solving the Linear Integer Programming Problem over a Finite Additive Group, with Extensions to Solving General and Certain Non-linear Integer Programs. *CRC* 66-29.

- GLOVER F. (1968). Surrogate Constraints. *Operations Research*, 16, 741-749.
- GLOVER F. (1969). Integer Programming Over a Finite Additive Group. *SIAM Journal on Control*, 7, 213-231.
- GLOVER F. (1977). Heuristics for Integer Programming Using Surrogate Constraints. *Decision Sciences*, 8, 156-166.
- GLOVER F. y LAGUNA M (1977). *Tabu Search*. (1a.ed). Massachusetts, EEUU: Kluwer Academic Publishers.
- GLOVER F. (1986). Future Paths for Integer Programming and Links to Artificial Intelligence. *Computers and Operations Research*, 13, 533-549.
- GRECHANOVSKY E. (1987). Stepwise regression procedures: overview, problems, results, and suggestions. *Annals of New York Academy of Sciences*, 491, 197-232.
- GUPTA F., PRAJOGI R. y STUBBS E. (1999). The Information Ratio and Performance. *Journal of Portfolio Management*, Fall, 33-39.
- GUNST, R.F. (1983). Regression analysis with multicollinear predictor variables: Definition, detection and effects. *Commun. Statist. Theor. Meth.*, 12, 2217-2260.
- HALLERBACH W.G. (1994). Index Tracking: Some Techniques and Results. En L.Peccati & M.Virén (Ed.), *Financial Modelling*. (1a.ed. pp. 113-137). Heidelberg, Alemania: Physica Verlag.
- HILL, R.C., FORMBY, T.B. and JOHNSON, S.R. (1977). Component selection norms for principal components regression. *Commun. Statist.*, A6, 309-334.
- HOCKING R.R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32, 1-49.
- HOERL R.W., SCHUENEMEYER J.H. Y HOERL A.E. (1986). A simulation of biased estimation and subset selection regression techniques. *Technometrics*, 28, 369-380.
- HOTELLING H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 24 417-441, 498-520
- JANSEN R. Y VAN DIJK R. (2002). Optimal benchmark tracking with small portfolios, *Journal of Portfolio Management*, 29, 33-9.

- KERNIGHAN B. W. y LIN S. (1970). An Efficient Heuristic Procedure for Partitioning Graphs. *Bell System Technical Journal*, 49, 291-307.
- KIRKPATRICK S., GELATT JR C. D. y VECCHI M. P. (1983). Optimization by Simulated Annealing. *Science*, 220, 671-680.
- KUNG E.C. and SHARIF T.A. (1980). Multi-regression forecasting of the Indian summer monsoon with antecedent patterns of the large-scale circulation. *WMO Symposium on Probabilistic and Statistical Methods in Wheather Forecasting*, 295-302
- LAPORTE G. y OSMAN I. H. (1995). *Metaheuristics in Combinatorial Optimization. Annals of Operations Research*, 60, J. C. Baltser Science Publishers, Basel, Switzerland.
- MALLOWS, C. (1973). Some comments on Cp. *Technometrics*, 15, 661-675.
- MANDEL J. (1989). Some thoughts on variable-selection in multiple regression. *Journal of Quality Technology*, 21, No.1, 2-6.
- MARKOWITZ, H. (1959) *Portfolio Selection*, New York: John Wiley and Sons.
- MILLER A.J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society A*, 147, 389-425.
- MITCHELL T.J. Y BEAUCHAMP J.J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83, 1023-1032.
- MONTGOMERY D.G. y PECK E.A. (1992). *Introduction to Linear Regression Analysis* (2a Ed.). New York, EEUU: John Wiley
- OSMAN I. H. y KELLY J. P. (1996). *Metaheuristics: Theory and Applications*, Norwell, MA, EEUU: Kluwer Academic Publishers
- PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag.* (6), 2, 559-572.
- PAPADIMITRIOU C. H. Y STEIGLITZ K. (1982). *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, New York.
- POLYA, G. (1957), *How to Solve it*. (2a.ed.). Princeton, EE.UU.: Princeton University Press.

SALAS C. (2007). *Evaluación e Impacto de Representaciones Alternativas de carteras de Inversión: Aplicación a Carteras de AFP*. Tesis de Magister no publicada, Pontificia Universidad Católica de Chile, Santiago, Chile.

SATCHELL S.E. y HWANG S. (2001). Tracking error: Ex-ante versus ex-post measures. *Journal of Asset Management* 2, 241-246.

SHARPE W.F. (1994). The Sharpe Ratio. *Journal of Portfolio Management*, Fall, 45-58.

SIMON H. A. y NEWELL A. (1958). Heuristic Problem Solving: The Next Advance in Operations Research. *Operations Research*, 6, N°1, 1-10.

SPARKS R.S., ZUCCHINI W. Y COUTSOURIDES D. (1985). On variable selection in linear regression. *Communications in Statistics-Theory and Methods*, 14, 1569-1587.

TETKO, I. V., VILLA, A. E. P.; LIVINGSTONE, D. J. (1996). Neural network studies. 2. Variable selection. *J. Chem. Inf. Comput. Sci.*, 36, 794-803.

THOMPSON M.L. (1978a). Selection of variables in multiple regression: Part I. A review and evaluation. *International Statistical Review*, 46, 1-49.

THOMPSON M.L. (1978b). Selection of variables in multiple regression: Part II. Chosen procedures, computations and examples. *International Statistical Review*, 46, 129-146.

ANEXOS

ANEXO A: CARACTERÍSTICAS DE LOS INSTRUMENTOS DEL MERCADO DE RENTA FIJA EN CHILE

A.1 Bonos de Gobierno

Se considerará bonos de Gobierno a todo papel emitido por Instituciones del Gobierno de Chile (Banco Central y Tesorería General de la República.). A continuación se presenta una breve descripción de algunos de los papeles más importantes de esta familia.

A.1.1 Pagarés descontables del Banco Central de Chile (PDBC)

Instrumento emitido para la ejecución de la política monetaria a través del Sistema de Operaciones de Mercado Abierto (SOMA).

El motivo de la emisión se basa en el cumplimiento del objeto del Banco Central de Chile, tendiente a velar por la estabilidad de la moneda y el normal funcionamiento de los pagos internos y externos, teniendo este Órgano del Estado atribuciones para regular la cantidad de dinero y de crédito en circulación, la ejecución de operaciones de crédito y de cambios internacionales, como también la dictación de normas en materia monetaria, crediticia, financiera y de cambios internacionales.

El monto de una emisión se determina de acuerdo con las metas de política monetaria.

Son títulos emitidos en pesos y se emiten en los siguientes cortes: \$5.000.000, \$50.000.000, \$100.000.000 y \$200.000.000.

El plazo de vencimiento de estos títulos es hasta cinco años contados desde su fecha de emisión y no devengan intereses. Son rescatados y pagados por el Banco Central de Chile a su vencimiento, al valor nominal de los títulos presentados a cobro.

A.1.2 Pagarés reajustables del Banco Central de Chile con pago en cupones (PRC)

Instrumento emitido para la ejecución de la política monetaria a través del Sistema de Operaciones de Mercado Abierto (SOMA).

El motivo de la emisión se basa en el cumplimiento del objeto del Banco Central de Chile, tendiente a velar por la estabilidad de la moneda y el normal funcionamiento de los pagos internos y externos, teniendo este Órgano del Estado atribuciones para regular la cantidad de dinero y de crédito en circulación, la ejecución de operaciones de crédito y de cambios internacionales, como también la dictación de normas en materia monetaria, crediticia, financiera y de cambios internacionales.

El monto de una emisión se determina de acuerdo con las metas de política monetaria, considerando, a la vez, que el monto máximo en circulación de PRC no supere los quinientos millones de unidades de fomento (UF).

Son títulos emitidos en UF y se emiten en los siguientes cortes: UF 500, UF 1.000, UF 5.000 y UF 10.000.

El plazo de vencimiento de estos títulos es desde cuatro hasta veinte años y devengan un interés a una tasa anual vencida, calculado sobre el capital expresado en UF.

El Banco Central de Chile paga estos títulos en cupones insertos en el mismo documento, con vencimientos semestrales iguales y sucesivos, comprendiendo éstos capital e intereses, salvo el último cupón que puede ser diferente.

A.1.3 Cupones de emisión reajustables opcionales (CERO) en UF

Cupones emitidos con la finalidad de dar la opción de sustitución y canje a los tenedores de Pagarés reajustables del Banco Central de Chile con pago en cupones (PRC). Esta sustitución posibilita el fraccionamiento de los cupones de los PRC, permitiendo disociar los pagos del instrumento original, teniendo de esta manera un tratamiento como valores independientes y entregando una mayor y más eficaz información de la curva de rendimiento.

El monto de una emisión se determina de acuerdo con las metas de política monetaria.

Son títulos emitidos en UF y sus cupones se emiten en los siguientes cortes: UF 500, UF 1.000, UF 5.000 y UF 10.000. Sin embargo, el Banco Central de Chile puede emitir cortes variables con el solo fin de hacer equivalentes los montos emitidos con los montos presentados a su sustitución y canje.

El plazo de vencimiento de estos títulos corresponderá a los mismos vencimientos establecidos en los instrumentos originales presentados para su sustitución y canje. Este instrumento además, no devenga intereses.

Los cupones son rescatados y pagados por el Banco Central de Chile a su vencimiento, al valor nominal de los títulos presentados a cobro.

A.1.4 Bonos del Banco Central de Chile en pesos y en UF (BCP) (BCU)

Instrumentos para la ejecución de la política monetaria a través del Sistema de Operaciones de Mercado Abierto, (SOMA).

El motivo de la emisión se basa en el cumplimiento del objeto del Banco Central de Chile, tendiente a velar por la estabilidad de la moneda y el normal funcionamiento de los pagos internos y externos.

Los BCP y BCU se emiten con la finalidad de reemplazar a los instrumentos del mismo emisor: PDBC y PRC, respectivamente, que circulan actualmente en el mercado financiero local, procurando, entre otros objetivos, estandarizar los instrumentos a formato *Bullet*.

El monto de la emisión se determina de acuerdo con las metas de política monetaria.

Los BCP son emitidos en pesos y se emiten con los siguientes cortes: \$5.000.000, \$50.000.000, \$100.000.000 y \$200.000.000. Los BCU son emitidos en UF y se emiten en los siguientes cortes: 500, 1.000, 5.000 y 10.000 UF.

No obstante lo anterior, el Banco puede, eventualmente, emitir cortes distintos.

El plazo de vencimiento mínimo de los bonos es de un año, contado desde su fecha de emisión y devengan un interés a una tasa anual vencida, determinada por el Gerente de División Política Financiera. La tasa de interés se determina en forma simple, calculándose sobre la base de períodos semestrales de 180 días y de un año de 360 días.

El Banco Central de Chile paga estos bonos en cupones con vencimientos semestrales iguales y sucesivos, que incluyen los intereses devengados, salvo el último cupón que comprende capital e intereses. Los cupones se pagan en moneda corriente nacional, a la fecha de los correspondientes vencimientos.

A.1.5 Bonos de la Tesorería General de la República en UF (BTU)

Títulos de deuda que se han emitido con la finalidad de financiar obligaciones de la Partida 50 del Tesoro Público: inversión financiera y prepago de deuda pública.

El monto de una emisión está condicionado al margen de endeudamiento establecido en la Ley de Presupuestos del Sector Público que esté vigente.

El valor nominal de los bonos se expresa en UF y se emiten en cortes de mínimos de 500 UF.

El plazo de maduración de los bonos en circulación es de diez y de veinte años contado desde su fecha de emisión.

De los bonos que se han colocado hasta la fecha, unos devengan interés de 4,5% anual vencido y otros devengan interés de hasta 3% anual vencido, que se calcula en forma simple sobre la base de períodos semestrales de 180 días y anuales de 360 días. Los intereses que se vayan devengando por la posesión de estos títulos, se pagan semestralmente. Si la fecha de vencimiento corresponde a un día inhábil bancario, el pago se efectuará el día hábil bancario inmediatamente siguiente, sin devengo de otras sumas, ya sea por intereses o por cualquier otra causa, adicionales al interés devengado efectivamente hasta la fecha originalmente dispuesta para su pago. Cada pago semestral de intereses se hace constar mediante un cupón que forma parte del bono respectivo.

La amortización del capital de los bonos se realiza en una sola cuota al vencimiento de los mismos. Si la fecha de vencimiento corresponde a un día inhábil bancario, el pago se efectuará el día hábil bancario inmediatamente siguiente al valor de la UF vigente en ese día. No hay rescates anticipados de estos bonos.

A.2 Bono Corporativo

Títulos de deuda emitidos por empresas, principalmente sociedades anónimas abiertas, que se emiten para financiar proyectos de inversión de largo plazo. También se emiten para satisfacer compromisos financieros, como el refinanciamiento de pasivos. Tratándose de una línea de bonos que esté vigente, se pueden realizar nuevas emisiones dentro de esta línea, para financiar exclusivamente el pago de los bonos que estén por vencer dentro de esta misma línea.

El monto de la emisión lo determina libremente cada emisor. No obstante, en el caso de que la finalidad de la emisión de bonos fuere la de financiar nuevos proyectos de inversión del emisor, de un monto superior al 40% del valor total de su activo individual existente antes de la emisión y que exijan la aplicación en etapas sucesivas de los recursos captados durante un período superior a un año, se nombrará, adicionalmente, en la escritura de emisión un administrador extraordinario de dichos recursos y un encargado de la custodia de los mismos.

La emisión de bonos puede efectuarse por montos fijos o por líneas de bonos; en este último caso, se entenderá como tal cuando las colocaciones individuales que estén vigentes no superen el monto total y el plazo de la línea inscrita en la Superintendencia de Valores y Seguros.

El valor nominal de los bonos corporativos puede estar expresado en pesos, en moneda extranjera, o utilizando alguna de las unidades de cuenta correspondientes a los sistemas de reajustabilidad autorizados por el Banco Central de Chile.

Los bonos pueden ser o no reajustables. Los bonos reajustables que se han emitido, en su gran mayoría están expresados en UF.

El plazo de la emisión debe ser superior a un año. En todo caso, el plazo para la amortización o pago total de la deuda no puede ser superior al plazo de duración del emisor.

La tasa de interés de una emisión es determinada libremente por la entidad emisora, la cual dejará estipuladas las fechas y lugares de pago de los intereses.

La amortización de este título de deuda puede realizarse en forma ordinaria directa, con pagos programados de capital e intereses, en forma trimestral, semestral o anual. También ocurre que sólo los intereses se pagan en parcialidades y que al momento del vencimiento del título se amortiza el capital; o bien, tanto el capital como los intereses se pagan sólo una vez y simultáneamente al momento de la total extinción del título.

Una emisión de bonos puede contemplar amortizaciones extraordinarias con rescates anticipados, que pueden efectuarse mediante sorteos u otros procedimientos que aseguren un tratamiento equitativo para todos los tenedores de bonos.

A.3 Depósito a plazo

Título de deuda que se emite para certificar el depósito de dinero recibido de una persona, sea esta natural o jurídica, bajo la obligación de restitución del dinero dentro de un plazo preestablecido, incluyendo reajustes e intereses convenidos, según corresponda. Se trata de una operación de captación que representa un pasivo para la entidad depositaria, la que se hace dueña del dinero depositado por tratarse de un depósito irregular, surgiendo la obligación de restitución a favor del depositante según lo precisado.

El monto mínimo de la emisión lo determina libremente la entidad depositaria. El monto máximo está condicionado a las relaciones entre patrimonio y activos de la entidad depositaria. Es decir, en el caso de los bancos, el monto del depósito a plazo que se pretende constituir no puede afectar los porcentajes mínimos exigidos por la ley, del 8% entre el patrimonio efectivo y los activos ponderados por riesgo, y del 3% entre el capital básico y los activos totales.

Los depósitos a plazo que reciban las entidades depositarias pueden expresar su valor en pesos, en moneda extranjera, o utilizando alguna de las unidades de cuenta correspondientes a los sistemas de reajustabilidad autorizados por el Banco Central de Chile.

Los depósitos a plazo, dependiendo de la forma en que se pacte el derecho a la restitución del dinero depositado, pueden clasificarse en: depósitos a plazo fijo, depósitos a plazo renovable, y depósitos a plazo indefinido.

En los depósitos a plazo fijo la entidad depositaria se obliga a pagar en un día prefijado. En los depósitos a plazo renovable se contemplan condiciones similares a los depósitos a plazo fijo, pero con la posibilidad de prorrogar automáticamente el depósito por un nuevo período, de la misma cantidad de días, en caso de que el depositante no retire el dinero. En los depósitos a plazo indefinido no se pacta, al momento de constituirlos, una fecha o plazo determinados de vencimiento, sino que la entidad depositaria se obliga a la restitución en un plazo prefijado a contar de la vista o aviso de su cliente.

No obstante lo anterior, no existe impedimento legal para pactar una forma mixta, es decir, que un depósito a plazo fijo se transforme en uno de plazo indefinido si aquel no se cobra en la época fijada. De acuerdo con las normas generales sobre captación e intermediación, relativas a plazos mínimos para el pago de intereses y reajustes, el plazo que se pacte no puede ser inferior a treinta días para depósitos no reajustables o a noventa días para depósitos reajustables. En el caso de los depósitos a plazo en moneda nacional, reajustables de acuerdo con la variación del valor del tipo de cambio del dólar de los Estados Unidos de América, pueden efectuarse desde un plazo de treinta días.

En los depósitos a plazo indefinido, el período que queda sujeto a las normas generales sobre plazos mínimos es aquel comprendido entre la fecha de la vista o aviso del depositante y la fecha de pago, es decir, ese lapso no puede ser inferior a 30 días para depósitos no reajustables y a 90 días para depósitos reajustables.

Si bien no hay cortes para los depósitos a plazo, cada entidad depositaria fija libremente el monto mínimo susceptible de documentarse en un depósito a plazo.

La amortización de un depósito a plazo es a la fecha de vencimiento pactada.

La tasa de interés de un depósito a plazo es determinada libremente por la entidad depositaria. No obstante, el depositante puede negociar otra tasa de interés, dependiendo del monto de dinero que quiera depositar. Los intereses pueden pagarse al vencimiento del documento o en forma periódica, dentro del período de vigencia del documento.

A.4 Letras Hipotecarias

Instrumento para el otorgamiento de préstamos con garantía hipotecaria, emitidos por empresas bancarias, sociedades financieras y aquellas cooperativas de ahorro y crédito sometidas a la fiscalización y control de la Superintendencia de Bancos e Instituciones Financieras.

Estas letras de crédito se emiten para la adquisición, construcción o ampliación de viviendas por parte de los usuarios finales de tales inmuebles. En los demás préstamos, tales como los que se otorgan a empresas constructoras para que edifiquen una o más viviendas, sólo pueden emitirse letras de crédito para fines generales.

Las entidades emisoras pueden emitir letras de crédito por cantidades iguales a los montos de los préstamos que otorguen en conformidad con el Título XIII de la Ley General de Bancos. Este Título se denomina «Operaciones Hipotecarias con Letras de Crédito». El monto de un préstamo o el de una emisión está condicionado por el valor del bien raíz a dar en garantía y se fija en un porcentaje de la tasación realizada por la entidad emisora. En el caso de préstamos para adquisición de viviendas, los montos corresponderán al menor valor entre el 75% de la tasación y el 75% del valor comercial de la vivienda.

El valor nominal de las letras de crédito puede estar expresado en pesos, en moneda extranjera o bien utilizando alguna de las unidades de cuenta correspondientes a los sistemas de reajustabilidad autorizados por el Banco Central de Chile.

Los cortes de una emisión de letras de crédito los determina el emisor.

Los mecanismos de reajustabilidad posibles de utilizar en una emisión de letras de crédito son los sistemas de reajustabilidad autorizados por el Banco Central de Chile.

El plazo de la emisión no puede ser inferior a un año.

La tasa de interés de cada emisión es determinada libremente por la entidad emisora. Esta tasa puede ser fija o flotante. Tratándose de tasas de interés flotante, estas deben

tener por base la Tasa de Interés Promedio (TIP) para operaciones reajustables entre 90 y 365 días que periódicamente publica el Banco Central de Chile, o la tasa prime o libo.

La amortización puede hacerse en forma ordinaria, ya sea directa o indirectamente, o en forma extraordinaria. La amortización ordinaria directa es aquella en que, periódicamente, el emisor paga parte del capital y de los intereses convenidos, cuyos valores se expresan en el respectivo cupón. La amortización ordinaria indirecta es aquella que se efectúa mediante compra o rescate de letras o por sorteo a la par, hasta por un valor nominal igual al fondo de amortización correspondiente al período respectivo.

La amortización extraordinaria consiste en la aplicación que el emisor debe hacer del pago anticipado en dinero que el deudor ha efectuado del todo o parte de su deuda, retirando de la circulación por compra, rescate o sorteo a la par, letras de crédito por igual valor. La amortización extraordinaria se produce también cuando el deudor paga anticipadamente el todo o parte de su deuda mediante la entrega de letras de crédito.

A.5 Bonos de Reconocimiento

Títulos de deuda emitidos por el Instituto de Normalización Previsional (INP) para reconocer períodos de cotizaciones previsionales que registraron imponentes del antiguo sistema previsional y que optaron u optan por afiliarse al nuevo sistema de previsión en alguna Administradora de Fondos de Pensiones (AFP), de acuerdo con lo establecido en el Decreto Ley N.º3500.

Tienen derecho a bono de reconocimiento, aquellos trabajadores que hayan registrado al menos doce cotizaciones mensuales en alguna institución de previsión en los cinco años anteriores al 13 de noviembre de 1980, que fue la fecha de la publicación del Decreto Ley N.º3500, que estableció el nuevo sistema de pensiones.

El causante de la emisión de un bono es el trabajador que fue imponente en el sistema previsional antiguo. El monto de la emisión o valor nominal del documento es calculado de acuerdo con la historia previsional del causante hasta la fecha de incorporación al nuevo sistema previsional.

El valor nominal de los bonos de reconocimiento se expresa en pesos y su emisión no presenta cortes.

El valor nominal del bono de reconocimiento se reajusta de acuerdo con la variación que experimente el Índice de Precios al Consumidor entre el último día del mes anterior a la fecha de incorporación del afiliado al nuevo sistema previsional y el último día del mes anterior a la fecha de su pago efectivo.

El plazo de vigencia del bono de reconocimiento es hasta que el causante cumpla los 60 años si es mujer y los 65 años si es hombre. Este plazo puede adelantarse por invalidez o fallecimiento del causante.

El valor nominal del bono de reconocimiento devenga un interés compuesto de 4% real anual, que se capitaliza anualmente por año vencido; no obstante, cuando al hacerse exigible el bono ha transcurrido solamente una fracción de año, o un año o más años

completos más una fracción de año, se aplicará una tasa de interés simple mensual de $(4/12)\%$, por cada mes calendario completo de dicha fracción de año.

El bono de reconocimiento se amortiza con una cuota en la fecha de su vencimiento.

ANEXO B: COMPOSICIÓN CARTERA DCV Y ASIGNACIÓN ACTIVOS

La siguiente Tabla contiene la información de la composición de la cartera del DCV para el día 30 de Junio de 2006 y la asignación de índices correspondiente. Debido a la gran cantidad de instrumentos (más de 58.000) que se maneja en la base, se presentará sólo parte de la cartera.

La asignación al índice se hace en base a la pertenencia del papel al de la composición del índice y en su defecto se busca asignar una cartera compuesta de 2 índices cercanos, con el fin de replicar la duración del instrumento.

Tabla B- 14: Información Cartera de Renta Fija del DCV y Asignación de activos

Nemotecnico	Tipo	Saldo MM	Moneda	DurMac	Saldo MMM \$	DurMacR	Indice
BCP0800407	BCP	128280,00	PESOS	0,734475326	128,28	1	RACLGOB_D1
BCP0800408	BCP	60000,00	PESOS	1,645868	60,00	2	RACLGOB_D2
BCP0800614	BCP	220000,00	PESOS	6,083386947	220,00	6	RACLGOB_D6
BCP0800615	BCP	116300,00	PESOS	6,639789414	116,30	7	RACLGOB_D7
BCP0800708	BCP	320000,00	PESOS	1,824956767	320,00	2	RACLGOB_D2
BCP0800709	BCP	255000,00	PESOS	2,63803315	255,00	3	RACLGOB_D3
BCP0800806	BCP	73400,00	PESOS	0,087671233	73,40	0	RACLGOB_D0
BCP0800810	BCP	165350,00	PESOS	3,480818447	165,35	3	RACLGOB_D3
BCP0800907	BCP	232500,00	PESOS	1,116455715	232,50	1	RACLGOB_D1
BCP0801206	BCP	91570,00	PESOS	0,421917808	91,57	0	RACLGOB_D0
BCU0500116	BCU	2,37	UF	7,656860088	43,02	8	RACLGOB_D8
BCU0500308	BCU	13,50	UF	1,600753801	245,04	2	RACLGOB_D2
BCU0500907	BCU	13,92	UF	1,136787266	252,75	1	RACLGOB_D1
BCU0500908	BCU	13,25	UF	2,058959319	240,51	2	RACLGOB_D2
BCU0500909	BCU	12,00	UF	2,937427919	217,82	3	RACLGOB_D3
BCU0500910	BCU	9,75	UF	3,77774237	176,98	4	RACLGOB_D4
BCU0500912	BCU	12,40	UF	5,354592441	225,08	5	RACLGOB_D5
BCU0500922	BCU	8,35	UF	11,56508709	151,56	12	RACLGOB_D10P
BCU0501113	BCU	11,00	UF	6,261611235	199,67	6	RACLGOB_D6
BBALE-AE	BE	0,13	UF	0,17260274	2,41	0	RACLCO_D1 RACLCO_D2
BBCIS-P1A1	BE	0,16	UF	0,492224299	2,90	0	RACLCO_D1 RACLCO_D2
BBCIS-P1A2	BE	0,04	UF	0,492224509	0,64	0	RACLCO_D1 RACLCO_D2
BWATT-B2	BE	0,30	UF	0,421917808	5,45	0	RACLCO_D1 RACLCO_D2
UBOH-C0192	BE	0,68	UF	0,251840206	12,40	0	RACLCO_D1 RACLCO_D2
BANDI-A1	BE	0,80	UF	1,14383438	14,52	1	RACLCO_D1
BANDI-A2	BE	2,50	UF	1,14386452	45,38	1	RACLCO_D1
BSTGL-SGB	BE	0,80	UF	0,539879512	14,52	1	RACLCO_D1
BTRA1-2A	BE	0,14	UF	1,222584955	2,45	1	RACLCO_D1
BTRA1-A	BE	0,16	UF	0,980819712	2,92	1	RACLCO_D1
UBCI-B0693	BE	0,35	UF	1,139221042	6,26	1	RACLCO_D1
UDES-B0793	BE	0,24	UF	0,982542294	4,36	1	RACLCO_D1
UDESC20794	BE	0,35	UF	1,463680786	6,37	1	RACLCO_D1
BAGUA-C1	BE	1,00	UF	2,2912832	18,15	2	RACLCO_D2
BAGUA-C2	BE	3,00	UF	2,291285383	54,45	2	RACLCO_D2

Nemotecnico	Tipo	Saldo MM	Moneda	DurMac	Saldo MMM \$	DurMacR	Indice
BALIB-A2	BE	0,41	UF	1,768585429	7,44	2	RACLCO_D2
BBCIS-P5A	BE	1,14	PESOS	2,48128853	20,69	2	RACLCO_D2
BSTDS-BJ	BE	10065,00	UF	1,591780822	10,07	2	RACLCO_D2
BSTDS-CB	BE	0,66	UF	2,487964482	11,89	2	RACLCO_D2
BSTDS-DB	BE	0,20	UF	2,48792685	3,54	2	RACLCO_D2
BSTDS-DJ	BE	0,05	UF	1,591780822	0,85	2	RACLCO_D2
BSTDS-DL	BE	0,01	UF	2,084931507	0,18	2	RACLCO_D2
BBCIP11204	BE	0,82	UF	3,293272572	14,90	3	RACLCO_D3
BBCIP21204	BE	4,50	UF	3,293272572	81,68	3	RACLCO_D3
BBCI-Q0305	BE	0,50	UF	3,496539796	9,08	3	RACLCO_D3
BBCIS-P6B	BE	5,00	PESOS	3,257534247	90,76	3	RACLCO_D3
BBCIS-P7A	BE	11500,00	PESOS	3,196444143	11,50	3	RACLCO_D3
BBCIS-P9A	BE	50000,00	PESOS	3,406668529	50,00	3	RACLCO_D3
UBOH-E1295	BE	8250,00	UF	4,479850825	8,25	4	RACLCO_D4
UEDWA10494	BE	1,61	UF	3,794292892	29,22	4	RACLCO_D4
UEDWA20995	BE	0,28	UF	4,348084233	5,08	4	RACLCO_D4
USUDB30693	BE	0,99	UF	4,024186189	17,88	4	RACLCO_D4
USUDB40693	BE	0,22	UF	4,024186539	3,90	4	RACLCO_D4
BAGUA-E	BE	0,21	UF	5,315465923	3,76	5	RACLCO_D5
BBCIS-P5B	BE	1,65	PESOS	4,923287671	29,95	5	RACLCO_D5
BBOHB50891	BE	6435,00	UF	4,718749272	6,44	5	RACLCO_D5
BCHIS-P8B	BE	0,45	PESOS	4,923287671	8,17	5	RACLCO_D5
UCHIA20197	BE	2500,00	UF	5,612242541	2,50	6	RACLCO_D6
UCHIB21097	BE	1,25	UF	6,058806793	22,69	6	RACLCO_D6
UEDWA60899	BE	2,50	UF	6,333330595	45,38	6	RACLCO_D6
UJNT-A1105	BE	1,00	UF	6,065657691	18,15	6	RACLCO_D6
BBICS-FC	BE	0,40	UF	6,639466363	7,26	7	RACLCO_D7
BBICS-FD	BE	1,23	UF	6,639465618	22,33	7	RACLCO_D7
BBICS-LA	BE	0,31	UF	7,100220032	5,57	7	RACLCO_D7
BBICS-LB	BE	0,40	UF	7,100234559	7,26	7	RACLCO_D7
BVAPO-A2	BE	0,16	UF	7,171478386	2,81	7	RACLCO_D7
UCHIC20100	BE	1,00	UF	6,762948008	18,15	7	RACLCO_D7
UCOR-Y1197	BE	2,25	UF	7,273224205	40,84	7	RACLCO_D7
UEDWA70899	BE	0,60	UF	7,100646981	10,89	7	RACLCO_D7
USECC30797	BE	1,00	UF	7,054310117	18,15	7	RACLCO_D7
BASOL-C1	BE	0,50	UF	7,662077326	9,08	8	RACLCO_D8
UBCIG20599	BE	0,97	UF	8,262446576	17,61	8	RACLCO_D8
UCHID20402	BE	0,35	UF	7,913764247	6,35	8	RACLCO_D8
UEST-A0799	BE	2,00	UF	7,718407768	36,30	8	RACLCO_D8
USECD10900	BE	4,00	UF	8,155253426	72,61	8	RACLCO_D8
USECD20900	BE	0,15	UF	8,155252537	2,72	8	RACLCO_D8
BAGUA-F	BE	0,60	UF	8,55457615	10,89	9	RACLCO_D9
BAINT-A	BE	5,00	UF	8,729980389	90,76	9	RACLCO_D9
BANDI-B1	BE	0,99	UF	9,250621735	17,97	9	RACLCO_D9
UDESE10999	BE	0,80	UF	8,58814456	14,52	9	RACLCO_D9
UDESE20999	BE	0,34	UF	9,264254779	6,17	9	RACLCO_D9
UEST-B0603	BE	0,07	UF	8,862885694	1,27	9	RACLCO_D9
BAPAC-B1	BE	2,50	UF	9,551359448	45,38	10	RACLCO_D10P
BAPAC-B2	BE	0,40	UF	9,551360129	7,30	10	RACLCO_D10P
BBICS-PN	BE	9,75	UF	10,14082123	176,98	10	RACLCO_D10P
BBICS-PO	BE	0,31	UF	10,1408222	5,54	10	RACLCO_D10P
BBICS-NK	BE	0,08	UF	24,3260274	1,40	24	RACLCO_D10P

Nemotecnico	Tipo	Saldo MM	Moneda	DurMac	Saldo MMM \$	DurMacR	Indice
BR00011115	BR	4,48	PESOS	9,345205479	0,00		RACLBR_DL
BR00020317	BR	5,61	PESOS	10,67945205	0,01	11	RACLBR_DL
BR00020509	BR	13,15	PESOS	2,84109589	0,01	3	RACLBR_DC
BR00140110	BR	3,83	PESOS	3,545205479	0,00	4	RACLBR_DL
BR00210108	BR	4,11	PESOS	1,561643836	0,00	2	RACLBR_DC
BR00211109	BR	3,17	PESOS	3,397260274	0,00	3	RACLBR_DC
BR00220108	BR	4,72	PESOS	1,564383562	0,00	2	RACLBR_DC
BR00220906	BR	8,77	PESOS	0,230136986	0,01	0	RACLBR_DC
BR00231007	BR	5,94	PESOS	1,315068493	0,01	1	RACLBR_DC
BR00250809	BR	6,12	PESOS	3,156164384	0,01	3	RACLBR_DC
BR00300806	BR	3,15	PESOS	0,167123288	0,00	0	RACLBR_DC
BR00311007	BR	6,01	PESOS	1,336986301	0,01	1	RACLBR_DC
BR01010107	BR	1,26	PESOS	0,506849315	0,00	1	RACLBR_DC
BR01010109	BR	1,46	PESOS	2,509589041	0,00	3	RACLBR_DC
BR01010110	BR	1,83	PESOS	3,509589041	0,00	4	RACLBR_DL
BTU0210915	BTU	11,00	UF	8,284625727	199,67	8	RACLGOB_D8
BTU0260925	BTU	11,00	UF	14,72641046	199,67	15	RACLGOB_D10P
BTU0450824	BTU	18,00	UF	12,73101946	326,73	13	RACLGOB_D10P
BTU0451023	BTU	19,86	UF	12,41648302	360,49	12	RACLGOB_D10P
BVL2AU0912	BVL	0,00	UF	3,083763739	0,04	3	RACLGOB_D3
BVL2AU0913	BVL	0,00	UF	3,548538894	0,07	4	RACLGOB_D4
BVL2AV0916	BVL	0,00	UF	4,907647301	0,04	5	RACLGOB_D5
BVL2AV0917	BVL	0,01	UF	5,348162183	0,11	5	RACLGOB_D5
BVL2BU0612	BVL	0,00	UF	2,966479066	0,01	3	RACLGOB_D3
BVL2BU0613	BVL	0,00	UF	3,433027706	0,04	3	RACLGOB_D3
BVL2BV0616	BVL	0,00	UF	4,79589822	0,01	5	RACLGOB_D5
BVL2BV0617	BVL	0,00	UF	5,238845873	0,06	5	RACLGOB_D5
BVL2BW0920	BVL	0,01	UF	6,63416365	0,18	7	RACLGOB_D7
BVL2BX0925	BVL	0,09	UF	8,647516604	1,66	9	RACLGOB_D9
BVL2CU0313	BVL	0,00	UF	3,316822229	0,02	3	RACLGOB_D3
BVL2CV0317	BVL	0,00	UF	5,129140657	0,06	5	RACLGOB_D5
BVL2CW0620	BVL	0,01	UF	6,528218767	0,16	7	RACLGOB_D7
BVL2CW1220	BVL	0,01	UF	6,738086175	0,12	7	RACLGOB_D7
BVL2CX0625	BVL	0,11	UF	8,551962584	2,00	9	RACLGOB_D9
BVL2CX1225	BVL	0,06	UF	8,745578502	1,03	9	RACLGOB_D9
BVL2DU1212	BVL	0,00	UF	3,200352537	0,01	3	RACLGOB_D3
BVL2DV1216	BVL	0,00	UF	5,018165972	0,03	5	RACLGOB_D5
BVL3AU0314	BVL	0,00	UF	3,77875429	0,02	4	RACLGOB_D4
BVL3AV0318	BVL	0,00	UF	5,566560374	0,04	6	RACLGOB_D6
BVL3AW0320	BVL	0,01	UF	6,423469578	0,10	6	RACLGOB_D6
CERO010107	CERO	0,37	UF	0,506849315	6,75	1	RACLGOB_D1
CERO010108	CERO	0,22	UF	1,506849315	4,05	2	RACLGOB_D2
CERO010109	CERO	0,13	UF	2,509589041	2,35	3	RACLGOB_D3
CERO010110	CERO	0,22	UF	3,509589041	4,00	4	RACLGOB_D4
CERO010111	CERO	0,59	UF	4,509589041	10,76	5	RACLGOB_D5
CERO010112	CERO	0,55	UF	5,509589041	9,93	6	RACLGOB_D6
CERO010113	CERO	0,24	UF	6,512328767	4,36	7	RACLGOB_D7
CERO010114	CERO	0,22	UF	7,512328767	3,95	8	RACLGOB_D8
CERO010115	CERO	0,25	UF	8,512328767	4,53	9	RACLGOB_D9
CERO010116	CERO	0,11	UF	9,512328767	2,00	10	RACLGOB_D10P
CERO010117	CERO	0,22	UF	10,51506849	3,99	11	RACLGOB_D10P
CERO011215	CERO	0,10	UF	9,42739726	1,84	9	RACLGOB_D9

Nemotecnico	Tipo	Saldo MM	Moneda	DurMac	Saldo MMM \$	DurMacR	Indice
CERO011216	CERO	0,13	UF	10,43013699	2,43	10	RACLG0B_D10P
CERO011217	CERO	0,10	UF	11,43013699	1,84	11	RACLG0B_D10P
CERO011218	CERO	0,06	UF	12,43013699	1,06	12	RACLG0B_D10P
CERO011219	CERO	0,02	UF	13,43013699	0,41	13	RACLG0B_D10P
CERO011220	CERO	0,01	UF	14,43287671	0,23	14	RACLG0B_D10P
FNBBV-030706	CLP	7281,21	PESOS	0,008219178	7,28	3	RACLIF_DM0
FNEST-140706	CLP	1633,02	PESOS	0,038356164	1,63	14	RACLIF_DM0
FNSTD-140706	CLP	11907,39	PESOS	0,038356164	11,91	14	RACLIF_DM0
FNBBV-170706	CLP	7980,33	PESOS	0,046575342	7,98	17	RACLIF_DM1
FNBCI-170706	CLP	80296,90	PESOS	0,046575342	80,30	17	RACLIF_DM1
FNHSB-140806	CLP	6552,18	PESOS	0,123287671	6,55	45	RACLIF_DM1
FUDES-140806	DUF	0,01	UF	0,123287671	0,18	45	RACLIF_DM1
FUSEC-140806	DUF	0,14	UF	0,123287671	2,51	45	RACLIF_DM1
FUSTD-140806	DUF	0,04	UF	0,123287671	0,76	45	RACLIF_DM1
FNBBV-150806	CLP	3145,08	PESOS	0,126027397	3,15	46	RACLIF_DM2
FNCOR-150806	CLP	9244,67	PESOS	0,126027397	9,24	46	RACLIF_DM2
FUSTD-130906	DUF	0,32	UF	0,205479452	5,78	75	RACLIF_DM2
FNBBV-140906	CLP	5476,42	PESOS	0,208219178	5,48	76	RACLIF_DM3
FNBCI-140906	CLP	2533,00	PESOS	0,208219178	2,53	76	RACLIF_DM3
FNBCI-140906	CLP	1942,53	PESOS	0,208219178	1,94	76	RACLIF_DM3
FNBCI-131006	CLP	17343,67	PESOS	0,287671233	17,34	105	RACLIF_DM3
FNBBV-161006	CLP	4712,90	PESOS	0,295890411	4,71	108	RACLIF_DM3 RACLIF_DM6
FNBCI-161006	CLP	468,00	PESOS	0,295890411	0,47	108	RACLIF_DM3 RACLIF_DM6
FNBCI-161006	CLP	3059,85	PESOS	0,295890411	3,06	108	RACLIF_DM3 RACLIF_DM6
FUSEC-080207	DUF	0,01	UF	0,610958904	0,21	223	RACLIF_DM6
FNBBV-090207	CLP	2122,67	PESOS	0,61369863	2,12	224	RACLIF_DM6
FNEST-090207	CLP	8339,50	PESOS	0,61369863	8,34	224	RACLIF_DM6
SNFACT090207	CLP	1000,00	PESOS	0,61369863	1,00	224	RACLIF_DM6
FNBBV-120207	CLP	20650,10	PESOS	0,621917808	20,65	227	RACLIF_DM6 RACLIF_DM12
FNBCI-120207	CLP	5304,03	PESOS	0,621917808	5,30	227	RACLIF_DM6 RACLIF_DM12
FNBBV-270307	CLP	5264,57	PESOS	0,739726027	5,26	270	RACLIF_DM6 RACLIF_DM12
FNRPL-270307	CLP	892,93	PESOS	0,739726027	0,89	270	RACLIF_DM6 RACLIF_DM12
FNBBV-280307	CLP	17072,66	PESOS	0,742465753	17,07	271	RACLIF_DM12
FNBCI-280307	CLP	9553,50	PESOS	0,742465753	9,55	271	RACLIF_DM12
FUSTG-281207	DUF	0,00	UF	1,495890411	0,04	546	RACLIF_DM12
FNCOO-030108	CLP	390,00	PESOS	1,512328767	0,39	2	RACLIF_D2
FUDES-221208	DUF	0,00	UF	2,482191781	0,00	2	RACLIF_D2
FUSTD-221208	DUF	0,00	UF	2,482191781	0,01	2	RACLIF_D2
FNRPL-050109	CLP	1303,66	PESOS	2,520547945	1,30	3	RACLIF_D3P
FUBBV-300924	DUF	0,06	UF	18,26575342	1,03	18	RACLIF_D3P
ABND050801	LH	0,00	UF	0,086111111	0,00	0	RACLH_D1 RACLH_D2
ATLA550195	LH	0,00	UF	0,256639281	0,03	0	RACLH_D1 RACLH_D2
SUD6030794	LH	0,00	UF	0,002777778	0,05	0	RACLH_D1 RACLH_D2
SUD6031194	LH	0,00	UF	0,336111111	0,02	0	RACLH_D1 RACLH_D2
ABNA080101	LH	0,00	UF	1,225301699	0,03	1	RACLH_D1
BCI-N71093	LH	0,00	UF	1,086853177	0,03	1	RACLH_D1
BCI-N71193	LH	0,00	UF	1,170042423	0,03	1	RACLH_D1
SUD-A 0699	LH	0,00	UF	2,433916528	0,01	2	RACLH_D2
SUD-A 0998	LH	0,00	UF	2,106121103	0,02	2	RACLH_D2
SUD-A 1198	LH	0,00	UF	2,139672288	0,04	2	RACLH_D2
SUD-A 1298	LH	0,00	UF	2,210171876	0,01	2	RACLH_D2
ABNA120100	LH	0,00	UF	2,620394274	0,02	3	RACLH_D3

Nemotecnico	Tipo	Saldo MM	Moneda	DurMac	Saldo MMM \$	DurMacR	Indice
ABNA120101	LH	0,00	UF	3,059238838	0,01	3	RACLLH_D3
SUD1041202	LH	0,01	UF	4,006904374	0,10	4	RACLLH_D4
SUD1090102	LH	0,00	UF	3,501800844	0,05	4	RACLLH_D4
ABNC150601	LH	0,00	UF	4,603689902	0,04	5	RACLLH_D5
ABNC150801	LH	0,01	UF	4,622150667	0,20	5	RACLLH_D5
SUD0860100	LH	0,00	UF	5,841652357	0,05	6	RACLLH_D6
SUD0860999	LH	0,00	UF	5,825138781	0,02	6	RACLLH_D6
ABNA200901	LH	0,01	UF	6,56617533	0,13	7	RACLLH_D7
ABNC200601	LH	0,01	UF	6,542112408	0,16	7	RACLLH_D7
ABNC200801	LH	0,04	UF	6,54976193	0,64	7	RACLLH_D7
STD52T1103	LH	0,02	UF	7,561981387	0,39	8	RACLLH_D8
STD52T1203	LH	0,00	UF	7,64531472	0,02	8	RACLLH_D8
STD53T1103	LH	0,00	UF	7,533769562	0,07	8	RACLLH_D8
BBVAA60506	LH	0,00	UF	8,522151275	0,07	9	RACLLH_D9P
BCIR420306	LH	0,00	UF	8,538373216	0,03	9	RACLLH_D9P
BCIR44 &04	LH	0,00	UF	8,613721432	0,02	9	RACLLH_D9P
FAL53T**06	LH	75000,00	UF	11,63791404	75,00	12	RACLLH_D9P
BNPDBC010806	PDBC	50000,00	PESOS	0,087671233	50,00	0	RACLGLOB_D0
BNPDBC040706	PDBC	50000,00	PESOS	0,010958904	50,00	0	RACLGLOB_D0
BNPDBC050706	PDBC	45000,00	PESOS	0,01369863	45,00	0	RACLGLOB_D0
BNPDBC050906	PDBC	95000,00	PESOS	0,183561644	95,00	0	RACLGLOB_D0
BNPDBC060706	PDBC	60000,00	PESOS	0,016438356	60,00	0	RACLGLOB_D0
BNPDBC280906	PDBC	45000,00	PESOS	0,246575342	45,00	0	RACLGLOB_D0
BNPDBC290806	PDBC	80000,00	PESOS	0,164383562	80,00	0	RACLGLOB_D0
BNPDBC310706	PDBC	45000,00	PESOS	0,084931507	45,00	0	RACLGLOB_D0
BNPDBC310806	PDBC	0,02	PESOS	0,169863014	0,38	0	RACLGLOB_D0
PRC-1A0100	PRC	0,01	UF	1,740919994	0,10	2	RACLGLOB_D2
PRC-1A0197	PRC	0,01	UF	0,266537164	0,18	0	RACLGLOB_D0
PRC-7B0699	PRC	0,20	UF	6,34676499	3,58	6	RACLGLOB_D6
PRC-7B0793	PRC	0,01	UF	3,487120945	0,18	3	RACLGLOB_D3
PRC-7B0794	PRC	0,03	UF	3,950594283	0,60	4	RACLGLOB_D4
PRC-7B0795	PRC	0,02	UF	4,404678033	0,34	4	RACLGLOB_D4
PRC-7B0796	PRC	0,03	UF	4,84160676	0,45	5	RACLGLOB_D5
PRC-7B0797	PRC	0,00	UF	5,277511938	0,02	5	RACLGLOB_D5
PRC-7B0798	PRC	0,01	UF	5,713869167	0,09	6	RACLGLOB_D6
PRC-7B1001	PRC	0,04	UF	7,237700743	0,76	7	RACLGLOB_D7
PRC-7B1093	PRC	0,10	UF	3,735911406	1,82	4	RACLGLOB_D4
PRC-7B1094	PRC	0,79	UF	4,197728603	14,34	4	RACLGLOB_D4
PRC-7D0502	PRC	0,00	UF	7,530532636	0,01	8	RACLGLOB_D8

ANEXO C: COMPOSICIÓN COMPONENTES PRINCIPALES

La Tabla muestra la composición de los 15 primeros componente principal en función de los índices utilizados.

Tabla C- 15: Composición primeros 15 Componentes Principales

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RAQL	1,86%	11,44%	3,52%	1,06%	1,54%	1,60%	0,63%	-5,52%	-24,07%	5,10%	3,21%	8,64%	-31,53%	4,18%	4,68%
RAQLGOB	1,90%	-15,97%	2,44%	-1,74%	-0,54%	-2,59%	5,97%	-4,09%	91,14%	2,03%	4,07%	2,06%	58,02%	4,86%	1,93%
RAQLBR	3,17%	352,92%	0,42%	3,88%	0,69%	-1,20%	-0,35%	2,19%	-19,40%	1,39%	-0,44%	-3,94%	4,51%	-1,01%	-0,18%
RAQLJF	0,02%	0,96%	0,60%	0,98%	5,23%	1,32%	0,92%	6,07%	2,55%	5,24%	5,75%	17,19%	-50,73%	7,87%	10,34%
RAQLLH	2,61%	-25,79%	5,81%	2,06%	-3,40%	6,51%	3,72%	46,78%	-189,16%	-15,46%	-1,20%	-11,27%	-88,12%	3,81%	-8,60%
RAQLOO	3,49%	-38,10%	8,34%	2,77%	1,10%	2,91%	-5,79%	-55,27%	-66,94%	22,63%	2,39%	17,10%	-56,57%	4,22%	6,30%
RAQLGOB_D0	0,03%	-0,17%	0,19%	0,17%	0,52%	0,48%	0,67%	2,99%	-8,32%	2,82%	1,91%	1,61%	7,27%	0,85%	5,49%
RAQLGOB_D1	0,43%	-4,69%	2,37%	0,03%	0,71%	5,45%	6,33%	9,93%	-14,45%	22,13%	10,04%	-4,25%	147,12%	13,08%	29,60%
RAQLGOB_D2	1,08%	-11,19%	4,18%	0,61%	-1,06%	7,37%	9,50%	-6,01%	67,66%	14,23%	8,85%	7,09%	100,14%	11,06%	1,45%
RAQLGOB_D3	1,63%	-10,77%	3,22%	0,52%	-1,90%	5,63%	7,58%	-11,99%	172,94%	16,84%	4,18%	15,99%	-38,54%	4,05%	-23,75%
RAQLGOB_D4	2,13%	-15,80%	3,69%	-0,20%	-2,15%	1,77%	5,72%	-6,48%	212,90%	-7,61%	3,15%	-4,86%	0,62%	18,29%	-29,06%
RAQLGOB_D5	3,02%	-29,54%	7,17%	-0,96%	0,65%	7,18%	12,76%	-16,87%	323,34%	-6,17%	-62,70%	48,47%	-805,63%	-16,68%	-2,34%
RAQLGOB_D6	3,27%	-25,62%	3,00%	0,06%	3,56%	1,82%	5,58%	-33,33%	436,32%	-51,00%	-0,15%	26,64%	-115,22%	-20,78%	38,23%
RAQLGOB_D7	3,58%	-17,10%	-0,88%	0,53%	-2,49%	-3,96%	-1,71%	-49,22%	574,46%	-74,71%	12,11%	-52,29%	308,10%	31,89%	12,84%
RAQLGOB_D8	4,63%	-52,78%	6,56%	3,78%	-2,41%	3,16%	-10,09%	-7,61%	17,25%	-51,88%	-6,28%	-7,47%	536,68%	22,68%	66,22%
RAQLGOB_D9	6,14%	-58,09%	-81,73%	4,67%	4,67%	8,09%	1,14%	-8,83%	-70,42%	12,91%	-2,95%	12,20%	-6,50%	1,00%	-0,18%
RAQLGOB_D10P	6,29%	-67,89%	2,95%	-26,45%	3,23%	-59,65%	13,73%	33,74%	-125,02%	17,19%	2,02%	-7,28%	91,03%	-9,44%	-4,59%
RAQLBR_DC	1,40%	203,27%	0,35%	77,38%	-13,57%	-19,56%	5,36%	-5,77%	-32,13%	3,64%	-1,25%	4,78%	7,10%	-3,23%	1,03%
RAQLBR_DL	4,18%	437,04%	0,56%	-37,60%	8,72%	9,31%	-3,53%	6,54%	-14,20%	1,14%	0,17%	-7,82%	-12,55%	-0,02%	-1,61%
RAQLJF_DM0	0,01%	0,19%	0,10%	0,11%	0,43%	0,30%	0,38%	1,39%	-3,19%	1,68%	0,87%	2,52%	-8,47%	0,67%	2,74%
RAQLJF_DM1	0,02%	0,46%	0,10%	0,18%	0,46%	0,67%	0,54%	2,13%	1,31%	2,84%	1,46%	6,02%	0,51%	2,83%	4,88%
RAQLJF_DM2	0,02%	1,78%	0,22%	0,27%	-0,41%	1,15%	0,53%	1,90%	-0,07%	5,94%	2,81%	14,40%	7,57%	5,57%	8,13%
RAQLJF_DM3	0,02%	1,80%	0,05%	-0,13%	-0,52%	1,41%	0,27%	5,38%	8,89%	6,88%	3,85%	13,44%	-17,42%	10,61%	6,21%
RAQLJF_DM6	0,02%	2,15%	0,45%	0,25%	1,97%	1,43%	0,65%	9,30%	16,16%	5,11%	4,68%	22,10%	-60,33%	13,46%	9,59%
RAQLJF_DM12	0,07%	1,63%	0,94%	1,01%	6,80%	1,59%	1,26%	7,81%	14,45%	8,56%	9,43%	28,08%	-103,10%	9,81%	18,49%
RAQLJF_D2	0,03%	-9,27%	2,50%	5,05%	25,87%	1,32%	2,66%	25,86%	-87,41%	15,22%	28,63%	43,69%	-401,17%	1,64%	52,51%
RAQLJF_D3P	-0,50%	-9,46%	4,80%	15,97%	79,26%	-0,15%	4,87%	-18,14%	-55,96%	-18,07%	-14,06%	-27,91%	199,96%	5,06%	-27,01%
RAQLLH_D1	0,94%	-8,44%	4,40%	0,97%	0,31%	10,01%	12,63%	25,07%	-85,32%	32,07%	10,89%	-22,38%	216,12%	10,67%	60,13%
RAQLLH_D2	1,36%	-12,09%	5,50%	0,36%	-2,26%	9,55%	11,77%	12,59%	-46,80%	14,63%	5,04%	-17,32%	161,81%	5,73%	18,90%
RAQLLH_D3	1,84%	-16,60%	5,47%	1,07%	-3,69%	8,25%	9,65%	16,73%	-61,74%	6,55%	-0,13%	-10,25%	83,60%	8,55%	3,24%
RAQLLH_D4	2,22%	-17,55%	5,51%	0,68%	-5,82%	6,41%	7,06%	9,26%	-116,36%	1,38%	-5,11%	-29,84%	68,65%	13,95%	8,71%
RAQLLH_D5	2,47%	-19,62%	6,20%	1,16%	-5,27%	6,35%	6,08%	12,52%	-92,90%	0,60%	-13,24%	-21,89%	-102,35%	5,48%	-21,48%
RAQLLH_D6	2,83%	-27,57%	7,27%	1,99%	-3,85%	6,47%	3,69%	19,47%	-176,14%	-11,55%	-19,08%	8,00%	-82,91%	2,86%	-17,43%
RAQLLH_D7	3,09%	-42,61%	5,68%	2,63%	-5,22%	9,30%	1,60%	89,25%	-690,41%	-50,37%	-34,95%	-73,49%	99,53%	25,42%	-16,18%
RAQLLH_D8	3,19%	-30,32%	4,04%	3,62%	0,75%	2,99%	-2,18%	91,58%	-250,02%	-67,67%	62,60%	53,80%	-627,06%	-26,85%	-7,92%
RAQLLH_D9P	3,93%	-50,41%	6,03%	9,26%	6,43%	3,08%	-17,42%	268,35%	507,00%	49,53%	-7,14%	-22,86%	88,68%	-1,91%	-17,23%
RAQLOO_D1	0,66%	-7,29%	3,57%	0,28%	1,42%	8,63%	10,74%	13,60%	-61,17%	32,27%	14,46%	-2,62%	198,16%	11,56%	51,44%
RAQLOO_D2	1,22%	-11,09%	4,32%	-0,23%	-1,83%	8,94%	11,51%	2,09%	40,06%	15,58%	11,25%	12,26%	81,94%	0,49%	7,30%
RAQLOO_D3	1,68%	-14,23%	4,86%	1,39%	-2,16%	6,99%	10,10%	-2,60%	126,92%	10,17%	4,52%	26,00%	-13,52%	-14,97%	4,53%
RAQLOO_D4	1,94%	-10,03%	4,44%	5,07%	-1,97%	6,84%	6,48%	-15,92%	119,05%	1,54%	4,00%	27,37%	-7,12%	3,99%	-15,76%
RAQLOO_D5	2,42%	-31,53%	4,46%	1,88%	-1,10%	5,80%	2,51%	-55,87%	66,24%	13,98%	42,10%	-35,70%	-69,72%	73,42%	-117,72%
RAQLOO_D6	2,97%	-35,64%	6,16%	1,93%	-0,26%	10,07%	-0,14%	-33,90%	-95,88%	0,82%	10,31%	39,65%	464,68%	-134,84%	-44,24%
RAQLOO_D7	3,41%	-29,35%	10,77%	1,33%	-0,52%	4,33%	-5,38%	-47,14%	-38,15%	19,63%	-17,19%	139,41%	449,12%	10,59%	-46,27%
RAQLOO_D8	3,85%	-51,26%	7,62%	4,77%	0,95%	4,00%	-7,69%	-96,27%	40,11%	36,19%	5,96%	-238,78%	-338,84%	-45,28%	9,44%
RAQLOO_D9	4,10%	-45,70%	9,62%	4,10%	3,99%	2,71%	-9,45%	-41,30%	-97,71%	11,65%	14,15%	33,46%	91,19%	-16,93%	22,04%
RAQLOO_D10P	5,37%	-60,07%	12,17%	3,49%	3,17%	-4,13%	-20,85%	-100,41%	-215,39%	34,35%	-9,02%	70,27%	-334,70%	41,75%	35,16%

ANEXO D: REGRESIÓN CON COMPONENTES PRINCIPALES

La Tabla D-16 muestra el resultado de la regresión de la cartera Pre-Mapeada en función de todos los PCs.

Tabla D- 16: Coeficientes y estadístico t de la regresión entre la cartera Pre-Mapea y los PCs

	<i>Coeficientes</i>	<i>Estadístico t</i>
alfa	0,00023868	10918
PC1	0,081414	48100
PC2	-0,023515	-7816,8
PC3	-0,039993	-8872
PC4	0,0088202	1555,5
PC5	-0,023421	-3720,9
PC6	0,025021	3262
PC7	-0,030412	-3453
PC8	0,0108	983,13
PC9	-0,012999	-1046,9
PC10	0,043764	3381,1
PC11	-0,025436	-1741
PC12	0,027529	1728,4
PC13	-0,0091124	-560,96
PC14	0,031109	1808,3
PC15	-0,033779	-1890,9
PC16	-0,0023	-123,72
PC17	-0,02486	-1306,7
PC18	-0,0075333	-377,86
PC19	-0,02737	-1318,7
PC20	0,0249	1191,1
PC21	-0,019349	-900,39
PC22	0,02935	1266,6
PC23	0,012852	504,97
PC24	0,031841	1154,7
PC25	-0,084981	-2903,2
PC26	0,036156	1184,6
PC27	-0,015954	-483,27
PC28	0,00092491	27,577
PC29	0,0073072	186,56
PC30	0,009192	209,17
PC31	-0,029007	-589,72
PC32	0,005441	101,92
PC33	0,019868	356,57
PC34	0,017648	281,62
PC35	0,0028637	37,513
PC36	-0,043982	-450,3
PC37	-0,031889	-216,28
PC38	-0,045703	-293,35
PC39	0,029585	125,98
PC40	-0,01004	-33,602
PC41	0,0456	137,94
PC42	-0,017168	-44,931
PC43	0,022526	51,356
PC44	0,012069	23,257
PC45	-0,008193	-14,56
PC46	0,030763	48,723

Análogamente, la Tabla D-17 ordena de forma decreciente a los PCs según valor absoluto del estadístico t .

Tabla D- 17: PCs ordenados de forma decreciente según valor absoluto del estadístico t .

	<i>Estadístico t</i>
PC1	48100
PC3	8872
PC2	7816,8
PC5	3720,9
PC7	3453
PC10	3381,1
PC6	3262
PC25	2903,2
PC15	1890,9
PC14	1808,3
PC11	1741
PC12	1728,4
PC4	1555,5
PC19	1318,7
PC17	1306,7
PC22	1266,6
PC20	1191,1
PC26	1184,6
PC24	1154,7
PC9	1046,9
PC8	983,13
PC21	900,39
PC31	589,72
PC13	560,96
PC23	504,97
PC27	483,27
PC36	450,3
PC18	377,86
PC33	356,57
PC38	293,35
PC34	281,62
PC37	216,28
PC30	209,17
PC29	186,56
PC41	137,94
PC39	125,98
PC16	123,72
PC32	101,92
PC43	51,356
PC46	48,723
PC42	44,931
PC35	37,513
PC40	33,602
PC28	27,577
PC44	23,257
PC45	14,56