

**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS  
AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL**

**UNIDAD ZACATENCO**

**DEPARTAMENTO DE  
MATEMÁTICAS**

**“Extensión de la estrategia de inversión  
por pares a clanes”**

**T E S I S**

Que presenta

**YULIANA MARÍA HINCAPIÉ HINCAPIÉ**

Para obtener el grado de  
**MAESTRA EN CIENCIAS**

**EN LA ESPECIALIDAD DE  
MATEMÁTICAS**

Director de la Tesis:  
**Dr. Feliú Davino Sagols Troncoso**



*A mi hogar, por su inmensa valentía*



# Agradecimientos

Al CONACyT, por el apoyo recibido durante el periodo de mi maestría, sin el cual no hubiera podido estar en México.

Al Cinvestav, por abrirme las puertas a tan agradables enseñanzas.

Al Dr. Feliú Davino Sagols Troncoso, por su infinita paciencia, por los retos y buenos comentarios. Gracias, no sólo por este trabajo, sino por toda la formación académica y personal que he recibido. Es un honor trabajar con él.

A mis sinodales, el Dr. Onésimo Hernandez Lerma y el Dr. Francisco Ortiz, por el tiempo dedicado para que este proceso llegara a su fin satisfactoriamente.

A mi familia, por el apoyo incondicional que siempre me brindan, y porque día a día me recuerdan que todo es posible y que la distancia nunca es impedimento para lograr lo que se quiere.

A Oscar García, por ver lo mejor de nosotros aún en las situaciones más turbias.

A mis amigos Daniel Muñoz, Kelly Palacio, Maycol Segura, María Teresa LLano, Yohanna Tabares, Isabel Castro, Lidia Frías y Pablo Lira, ya que en estos últimos años han sido un pilar fundamental en mi vida, y no miento al decir que este trabajo no se hubiese podido terminar sin el apoyo de ellos.

Por último, y no menos importante, al Señor pues sin él no soy nada.



# Índice general

<b>1. Preliminares</b>	<b>7</b>
1.1. Series de tiempo . . . . .	7
1.2. Conceptos de finanzas . . . . .	13
<b>2. Modelo de inversión por pares</b>	<b>21</b>
2.1. Identificación de pares de activos . . . . .	21
2.1.1. Modelo de tendencias comunes y la APT . . . . .	24
2.2. Prueba de cointegración . . . . .	29
2.3. Estrategias de inversión . . . . .	33
2.3.1. Primera estrategia de inversión . . . . .	35
2.3.2. Segunda estrategia de inversión . . . . .	36
<b>3. Generalización del modelo inversión por pares</b>	<b>39</b>
3.1. Identificación de clanes de activos. . . . .	39
3.2. Spread para clanes . . . . .	43
<b>4. Otras implementaciones y resultados</b>	<b>47</b>
4.1. Búsqueda de subclanes cointegrantes . . . . .	48
4.1.1. Subclan con mejor coeficiente de determinación $R^2$ . . . . .	48
4.1.2. Búsqueda exhaustiva de subclanes . . . . .	52
4.2. Correlaciones entre pruebas de estacionariedad . . . . .	55
4.2.1. Índice de Correlación . . . . .	55

4.2.2. Pruebas de estacionariedad ADF y PP . . . . .	56
4.2.3. Medida de estacionariedad $M_s$ . . . . .	57
4.2.4. Optimización de la correlación . . . . .	59
<b>A. Calculo del umbral <math>\Delta</math></b>	<b>65</b>
<b>B. Promedio móvil del spread</b>	<b>71</b>
<b>Bibliografía</b>	<b>75</b>

# Resumen

La técnica de inversión por pares, en inglés *pairs trading*, es una estrategia de inversión neutral cuyos portafolios de mercado se construyen utilizando sólo dos instrumentos  $A$  y  $B$ , usando una posición larga (corta) de una unidad en  $A$  y una posición corta (larga) de  $\gamma$  unidades en  $B$ . Para poder aplicar la técnica es necesario que la serie  $s = \log(p_t^A) - \gamma \log(p_t^B)$ , que en adelante será llamada *spread*, sea estacionaria. En otras palabras, la técnica se lleva a cabo cuando  $A$  y  $B$  son cointegrantes. Aquí  $p_t^A$  y  $p_t^B$  son los precios en el tiempo  $t$  de  $A$  y  $B$ , respectivamente. Si usamos como regla ir largo en  $A$  y corto en  $B$  cuando el spread está por debajo de su media y corto en  $A$  y largo en  $B$  cuando el spread está por encima de su media entonces el rendimiento en el intervalo de tiempo que inicia en  $t_1$  y termina en  $t_2$  (sin considerar comisiones) es precisamente  $s(t_2) - s(t_1)$ . Cuando el spread es estacionario éste muestra reversión a la media y se pueden esperar ganancias entrando al mercado siempre que el spread se aleje suficientemente de la media y saliendo al cruzarla.

El hecho de usar pares de activos en esta estrategia pareciera restrictivo. Podemos pensar, por ejemplo, que tenemos a nuestra disposición el espacio de series generado por las combinaciones lineales de los instrumentos básicos del mercado. Entonces, podemos aplicar la estrategia de inversión por pares en el espacio resultante. Encontrar un par de portafolios apropiados para la estrategia equivale a encontrar una combinación lineal de instrumentos básicos que produzca una serie estacionaria. Así, lo que proponemos es generalizar la estrategia de inversión por pares de manera que no sólo intervengan dos instrumentos básicos en el mercado sino conjuntos con dos o más elementos.



# Abstract

Pairs trading is a market neutral strategy whose portfolios are constructed using only two instruments  $A$  and  $B$ , using a long (short) position of a unit of  $A$  and a short (long) position of  $\gamma$  units of  $B$ . In order to apply this strategy it is necessary that the series  $s = \log(p_t^A) - \gamma \log(p_t^B)$ , named spread, be stationary. In other words,  $A$  and  $B$  must be cointegrating. Where  $p_t^A$  and  $p_t^B$  are the prices at time  $t$  of  $A$  and  $B$ , respectively. If we use as a rule to go long in  $A$  and short in  $B$  when the spread is below its mean, and short in  $A$  and long in  $B$  when the spread is above its mean, then the return in the time interval starting at  $t_1$  and ending in  $t_2$  (without considering commissions) is  $s(t_2) - s(t_1)$ . When the spread is stationary this shows a reversion to the mean and thus we can expect a revenue entering the market whenever the spread goes far enough from the average and leaving when crossing it.

The fact of using pairs of assets in this strategy seems restrictive. We can think, for example, that we have at our disposal the series space generated by the linear combinations of the basic market instruments. Then, we can apply the strategy of pairs trading in the resulting space. Finding a pair of appropriate instruments for the strategy is equivalent to finding a linear combination of assets that produces a stationary series. Thus, we propose to generalize the strategy of pairs trading so that not only two basic instruments in the market intervene but also two or more elements.



# Introducción

En el mundo de las inversiones siempre ha existido un profundo interés en conocer a detalle los movimientos del mercado, ya sea por intereses académicos, para construir modelos matemáticos, o por el deseo de construir estrategias de inversión que sean redituables. Históricamente vemos que muchas estrategias que fueron muy exitosas en su momento, poco a poco perdieron su capacidad para generar ganancias por muchas razones: se crearon métodos o regulaciones para contrarrestarlas, o simplemente la práctica masiva de la misma anuló su capacidad para generar ganancias. Así, muchas veces, si algún grupo de personas llega a diseñar una nueva estrategia de inversión, no es común que lo reporten en la literatura; pues en este caso, hay más un interés económico que científico y se prefiere ocultar su existencia.

Una de las estrategias de la cual podemos encontrar bastante información es el modelo de inversión por pares, que como su nombre lo indica, trabaja con sólo dos instrumentos del mercado. Se puede encontrar más información de este modelo en [Whi04], [Pol07] o [Ehr06]. Esta tesis está enteramente basada en el modelo de inversión por pares, en sus partes teórica y práctica. Conoceremos un poco de su historia en el Capítulo 1. En el Capítulo 2 se expone la forma de ejecución del modelo, que en resumen consta de tres pasos:

1. Identificación de posibles pares cointegrantes.
2. Prueba de cointegración.
3. Estrategia de inversión.

Una descripción detallada de la estrategia de inversión por pares aparece en [Vid04]. Lo realmente novedoso de esta tesis es que se implementaron algoritmos para cada uno de los tres pasos ya mencionados.

La estrategia de inversión por pares desde el inicio parecía un método restrictivo, en el sentido de que se trabaja con sólo dos instrumentos, así es que nos empezamos a plantear algunas preguntas como: ¿Necesariamente se debe trabajar con dos instrumentos?, ¿se tendría mayor maleabilidad si se hiciera uso de tres o más instrumentos? Es así, que en el Capítulo 3 se generaliza la estrategia de inversión por pares; donde se intenta hacer un símil con la estrategia de pares, empezando con crear un algoritmo que pueda identificar posibles clanes cointegrantes, es decir, identificar un conjunto de clanes de tamaño mayor o igual a dos que cumplan con la propiedad de ser cointegrantes. A partir de ahí, se crean dos portafolios  $A$  y  $B$  que en adelante actuarán como un par, continuando con los pasos 2 y 3 de la estrategia de inversión por pares.

Finalmente, el Capítulo 4 se divide en dos partes. En la primera se crea un algoritmo que busca exhaustivamente subclanes cointegrantes de un clan arbitrario. Por último, se explica un método para ajustar parámetros de la prueba de cointegración introducidos en el Capítulo 2. Esto se hace por medio de un algoritmo genético que ajusta estos parámetros para que exista una mejor correlación entre esta prueba y otras pruebas de estacionariedad conocidas, como son la prueba de Dickey-Fuller aumentada y la prueba de Phillip-Perron, de las cuales se puede encontrar información en [MK98].

# Capítulo 1

## Preliminares

Este capítulo está dividido en dos secciones. La primera da un recorrido breve por los términos y conceptos de probabilidad necesarios para entender el modelo de inversión por pares. Se hace a manera de información para las personas que no tengan conocimientos de probabilidad y procesos estocásticos. En la segunda sección, mencionaremos algunos conceptos relacionados con finanzas de forma muy puntual, para al final introducir con detalle el método de inversión por pares que es la base de esta tesis.

### 1.1. Series de tiempo

Desde el punto de vista teórico una *serie de tiempo* es una colección de variables aleatorias  $\{X_t\}$ . Una colección de tales variables aleatorias ordenadas en el tiempo es un *proceso estocástico*. La palabra estocástico es de origen griego y significa *perteneciente al azar*. Si se trata de una variable continua se denota la variable por  $X(t)$ , y si  $t$  es una variable discreta, es costumbre denotar la variable por  $X_t$ . Un ejemplo de variable aleatoria continua  $X(t)$  es el registro de un electrocardiograma. Ejemplos de variables aleatorias discretas son los datos de desempleo, ofertas de dinero, cierre de precios de instrumentos financieros, entre otros. Aquí sólo vamos a considerar procesos discretos. Por lo tanto, sin motivo a confusión

usaremos cualquiera de las dos notaciones  $X_t$  o  $X(t)$ .

La estructura de probabilidad de la sucesión de variables aleatorias  $\{X_t\}$  queda determinada por la distribución conjunta del proceso estocástico. Se plantea, sin embargo, la cuestión: si el tiempo  $T$  es un conjunto infinito, ¿necesitamos una distribución dimensionalmente infinita para definir la estructura de probabilidad del proceso estocástico? Kolmogorov en 1933 mostró que cuando el proceso estocástico satisface ciertas condiciones de regularidad, éste se describe por una colección de distribuciones de dimensión finita. Es decir, bajo ciertas condiciones, la estructura probabilística del proceso estocástico  $\{X_t\}$  queda completamente especificada por las distribuciones conjuntas  $F(X_{t_1}, \dots, X_{t_n})$  para todos los valores de  $n$  (enteros positivos) y cualquier subconjunto  $(t_1, \dots, t_n)$  de  $T$ . Una de las condiciones de regularidad es la *simetría*, que consiste en que al reacomodar los índices permanece invariante la distribución. La otra condición es la de *compatibilidad*, es decir, la dimensión de la distribución conjunta se puede reducir tomando la distribución marginal.

Usar distribuciones conjuntas para definir procesos estocásticos es apropiado sólo desde una perspectiva teórica. En algunos casos, es posible y mucho más práctico hacerlo en términos del primero y segundo momentos. Si para un  $t$  específico,  $X_t$  es una variable aleatoria, podemos denotar sus funciones de distribución y densidad con  $F(X_t)$  y  $f(X_t)$ , respectivamente. Así, la familia paramétrica de densidades está determinada por:

$$\begin{aligned} \text{Media } \mu_t &= E(X_t) \\ \text{Varianza } \sigma_t^2 &= \text{var}(X_t) \\ \text{Covarianza } \gamma_{t_1, t_2} &= \text{cov}(X_{t_1}, X_{t_2}) \end{aligned}$$

La distribución de un proceso estocástico se caracteriza por su primer y segundo momentos, ambos como funciones de  $t$ . Note que si  $X(t)$  sigue una distribución normal, la distribución de  $X(t)$  queda completamente caracterizada por su media y varianza. Esto se conoce como *proceso gaussiano*.

El hecho de que los parámetros desconocidos cambian con el tiempo  $t$  re-

presenta un problema difícil. Hay un número finito pero demasiado grande de parámetros a estimar. Sin embargo, tenemos sólo una muestra de tamaño uno en cada una de las variables aleatorias. Por ejemplo, si decimos que la tasa de desempleo al final de esta semana es una variable aleatoria, tenemos sólo una observación de esta variable aleatoria particular en una semana. No hay forma de conseguir otra observación, por lo que tenemos lo que se llama una *única realización*. Esta característica nos obliga a especificar algunos modelos altamente restrictivos para la estructura estadística del proceso estocástico. Dada una sola realización, necesitamos reducir el número de parámetros  $\mu_t, \sigma_t^2, \gamma_{t_1, t_2}$  para  $t \in T$ . Esto se puede hacer imponiendo la condición de *estacionariedad*, donde se restringe la heterogeneidad en el tiempo del proceso.

**Estacionariedad.** Una serie de tiempo se dice que es estrictamente estacionaria si la distribución conjunta de  $X_{t_1}, \dots, X_{t_n}$  es la misma que la distribución conjunta de  $X_{t_1+\tau}, \dots, X_{t_n+\tau}$  para todo  $t_1, \dots, t_n$  y  $\tau$ . La distribución del proceso estacionario permanece sin cambios cuando es desplazada en el tiempo por un valor arbitrario de  $\tau$ . Así, los parámetros que caracterizan la distribución del proceso no dependen de  $t$ , en el lapso o *lag*  $\tau$ . El concepto de estacionariedad es difícil de verificar en la práctica por que está definido en términos de la función de distribución. Por esta razón, el concepto de estacionariedad definido en términos de los momentos se prefiere habitualmente.

Un proceso estocástico  $\{X_t, t \in T\}$  se dice estacionario de orden  $l$ -ésimo si para cualquier subconjunto  $(t_1, \dots, t_n)$  de  $T$  y cualquier  $\tau$  los momentos son

$$E(X_{t_1}^{l_1}, \dots, X_{t_n}^{l_n}) = E(X_{t_1+\tau}^{l_1}, \dots, X_{t_n+\tau}^{l_n}),$$

donde  $l_1 + l_2 + \dots + l_n \leq l$ , tomando enteros positivos para  $l_1, \dots, l_n$  y  $l$ .

Cuando  $l = 1$ , es decir,  $l_1 = 1$ , se tiene que  $E(X_t) = E(X_{t+\tau}) = \mu(\text{constante})$ , y el proceso  $\{X_t\}$  se dice que es *estacionario de primer orden*.

Cuando  $l = 2$ , los posibles casos son  $(l_1 = 1, l_2 = 0)$ ,  $(l_1 = 2, l_2 = 0)$  y  $(l_1 = 1, l_2 = 1)$ . De acuerdo con los tres casos el proceso  $\{X_t\}$  tendría los siguientes

tes momentos:

$$\begin{aligned} E(X_t) &= E(X_{t+\tau}) = \mu \text{ (una constante)} \\ E(X_t^2) &= E(X_{t+\tau}^2) = \sigma^2 \text{ (una constante)} \\ \text{cov}(X_{t_1}, X_{t_2}) &= \text{cov}(X_{t_1+\tau}, X_{t_2+\tau}) = \gamma_{t_1, t_2} = \gamma_\tau \end{aligned}$$

donde  $t_1 - t_2 = \tau$ . La media y la varianza de  $X_t$  son constantes y la covarianza de  $X_t$  depende sólo del *lag* o intervalo  $\tau = t_1 - t_2$ , no de  $t_1$  o  $t_2$ . En estas condiciones se dice que  $X_t$  es *estacionario de segundo orden*. La estacionariedad de segundo orden es también llamada *débil*, de *sentido amplio* o de *covarianza estacionaria*. En el modelado de series de tiempo, la estacionariedad de segundo orden es comúnmente la más utilizada para estacionariedad. Esto se debe en parte al hecho de que en un proceso estacionario normal (o gaussiano), estacionariedad de segundo orden es equivalente a estacionariedad estricta, para mayor información ver [MK98]. Si  $X_t$  sigue una distribución normal multivariada, ésta queda completamente caracterizada por el primero y segundo momentos, los dos conceptos de estacionariedad estricta y estacionariedad débil son equivalentes (recordemos caso  $n = 2$ ). Para otras distribuciones esto no es así.

Una de las formas más simples de analizar series de tiempo no estacionarias (que es lo más común) es volverlas estacionarias por diferenciación; de [MK98] se puede ver que si  $y_t$  representa una *caminata aleatoria*, ésta no es estacionaria, pero tal condición puede cambiarse diferenciando una vez, es decir, calculando  $\Delta y_t = y_t - y_{t-1}$ .

Cuando las series no estacionarias se convierten en estacionarias al diferenciar una vez, la serie se dice que es *integrada de orden 1* y se escribe por  $I(1)$ . Si la serie necesita ser diferenciada  $k$  veces para alcanzar estacionariedad, entonces la serie es  $I(k)$ . En nuestro ejemplo la variable  $y_t$  es  $I(1)$ . Algunas veces corremos con la suerte de partir de una serie  $x_t$  estacionaria, en tal caso escribimos que  $x_t$  es  $I(0)$ . A las series  $I(k)$  con ( $k \neq 0$ ) se les llama *procesos diferencialmente estacionarios*.

**Cointegración.** A pesar de que dos series de tiempo son no estacionarias, es posible que una combinación lineal específica de ellas lo sea; es decir, las dos series se mueven juntas en un tiempo dado. Los econométristas Engle y Granger

propusieron para este fenómeno el término de *cointegración*. Formalmente, sean  $x_t, y_t$  dos variables  $I(1)$ , se dice que son *cointegradas* si existe un número  $\beta$ , llamado *coeficiente de cointegración*, tal que  $y_t - \beta x_t$  es  $I(0)$  y escribimos que  $x_t$  y  $y_t$  son  $CI(1, 1)$ . Más generalmente, si  $x_t$  es  $I(b)$  y  $y_t$  es  $I(d)$ , entonces  $x_t$  y  $y_t$  son  $CI(b, d)$  si  $y_t - \beta x_t$  es  $I(d - b)$  con  $b > 0$ ; esto significa que la ecuación de regresión

$$y_t = \beta x_t + \mu_t,$$

tiene sentido porque las variables  $x_t, y_t$  no se alejan mucho la una de la otra con el tiempo. Así, hay una relación de equilibrio a largo plazo entre ellos. Si  $x_t$  e  $y_t$  no son cointegradas, es decir,  $y_t - \beta x_t = \mu_t$  es también  $I(1)$ , entonces las series serán divergentes con el tiempo para todo  $\beta$ . En este caso, la relación entre las variables  $y_t, x_t$  que obtenemos mediante la regresión de  $y_t$  y  $x_t$  sería *espuria*; pero nosotros no abundaremos sobre el tema (ver [MK98]).

Ahora se hará un breve recorrido por un modelo que captura de manera óptima la noción de cointegración que fue desarrollado en 1981 por Granger, proporcionándonos una interesante relación entre cointegración y el modelo que a continuación presentaremos, llamado *modelo de corrección de errores* que en adelante abreviaremos por MCE.

**Modelo de Corrección de Errores (MCE).** Sean  $\{x_t\}, \{y_t\}$  series de tiempo y  $\varepsilon_{x_t}, \varepsilon_{y_t}$  procesos de ruido blanco. Las series  $\{x_t\}, \{y_t\}$  mantienen un MCE si existen constantes  $\alpha_x, \alpha_y$ , y  $\gamma$  tales que, para todo  $t = 1, 2, \dots$ , se tiene que

$$y_t - y_{t-1} = \alpha_y(y_{t-1} - \gamma x_{t-1}) + \varepsilon_{y_t},$$

$$x_t - x_{t-1} = \alpha_x(y_{t-1} - \gamma x_{t-1}) + \varepsilon_{x_t},$$

donde el lado izquierdo de estas expresiones es el incremento de las series de tiempo en cada periodo. El lado derecho consta de dos partes, la parte de corrección de error y la parte del ruido blanco, respectivamente. Veamos la parte de corrección de error  $\alpha_y(y_{t-1} - \gamma x_{t-1})$  de la primera ecuación. El término  $y_{t-1} - \gamma x_{t-1}$  es representativo de la desviación del equilibrio a largo plazo (el valor de equilibrio es cero en este caso), y  $\gamma$  es el coeficiente de cointegración. La constante  $\alpha_y$  es la tasa de corrección de error, indicativo de la velocidad de corrección pa-

ra mantener el equilibrio. Así, conforme las series evolucionan en el tiempo, las desviaciones del equilibrio provienen del ruido blanco.

**Teorema 1.1** (Teorema de Representación de Granger [Vid04]). *Sean  $x_t, y_t$  dos series de tiempo  $I(1)$ . Ambas series son cointegrantes si y solo si ellas tienen una representación MCE.*

Un enfoque más directo del modelo de cointegración es atribuido a Stock y Watson, se denomina *Modelo de Tendencias Comunes* el cual presentaremos a continuación.

**Modelo de Tendencias Comunes.** La idea principal es que una serie de tiempo se expresa como una simple suma de dos series componente: un componente estacionario y un componente no estacionario común. Si las dos series cointegran, entonces la composición de cointegración lineal actúa para anular los componentes no estacionarios, dejando sólo los componentes estacionarios. Para ver esto consideremos dos series de tiempo:

$$\begin{aligned}y_t &= n_{y_t} + \varepsilon_{y_t}, \\z_t &= n_{z_t} + \varepsilon_{z_t},\end{aligned}$$

donde las componentes  $n_{y_t}, n_{z_t}$  son caminatas aleatorias (no estacionarias) de las dos series de tiempo, y  $\varepsilon_{y_t}, \varepsilon_{z_t}$  son las componentes estacionarias de las series de tiempo. Por otra parte, si  $y_t, z_t$  son cointegrantes, entonces la combinación lineal  $y_t - \gamma z_t$  debe ser estacionaria para alguna  $\gamma$  apropiada. Expandiendo y reorganizando términos tenemos

$$y_t - \gamma z_t = (n_{y_t} - \gamma n_{z_t}) + (\varepsilon_{y_t} - \gamma \varepsilon_{z_t}).$$

Si la combinación en la anterior ecuación debe ser estacionaria, la componente no estacionaria debe ser igual a cero, implicando que  $n_{y_t} = \gamma n_{z_t}$ , o sea que el componente de tendencia de una serie debe ser un múltiplo escalar del componente de tendencia en la otra serie. Por tanto, para que dos series sean cointegrantes, las tendencias deben ser idénticas salvo por un escalar.

## 1.2. Conceptos de finanzas

En esta sección se presenta el marco teórico que envuelve el modelo de inversión por pares, sus modelos matemáticos, y claro está, la interpretación financiera adecuada para comprender los próximos capítulos.

**Modelo CAPM.** CAPM es el acrónimo de *Capital Asset Pricing Model*, originalmente propuesto por William T. Sharpe. Es muy importante porque de él se desprenden muchos otros modelos en el campo de finanzas. Aquí es donde se formaliza la noción de un *portafolio de mercado*, el cual es un conjunto de activos que actúan como un medio integrado de inversión.

El modelo CAPM explica los rendimientos<sup>1</sup> de los activos como la suma de dos componentes, uno es el componente de mercado o sistemático y el otro es el componente residual o no sistemático. Más concretamente, si  $r_p$  es el rendimiento del activo y  $r_m$  es del rendimiento del portafolio de mercado, la fórmula que relaciona los dos rendimientos es:

$$r_p = \beta r_m + \theta_p. \quad (1.1)$$

donde,  $\beta r_m$  es la componente del rendimiento del mercado y la variable  $\theta_p$  es el componente residual o rendimiento residual del portafolio. Más específicamente,  $\theta_p$  es la porción del rendimiento de los activos que no es explicable por el rendimiento del mercado.

Una vez establecida la separación de los rendimientos de los activos en dos componentes, el modelo CAPM procede a elaborar una suposición clave con respecto a la relación entre ellos. La aserción del modelo es que el componente de mercado y el componente residual no están correlacionados

La variable  $\beta$  es muy importante en este modelo y observemos que es realmente la pendiente de la Ecuación (1.1). Por lo tanto, esta puede ser estimada como la pendiente de la línea de regresión entre los rendimientos del mercado y los rendimientos de los activos. Aplicando la fórmula de regresión estándar para la

---

<sup>1</sup> sea  $\{p_t, t \in T\}$  la serie de precios de un activo cualquiera, el  $i$ -ésimo rendimiento se define como  $\log(p_{i+1}) - \log(p_i) \approx \frac{p_{i+1} - p_i}{p_i}$

estimación de la pendiente, tenemos que es la covarianza entre el rendimiento del activo y el rendimiento del mercado dividido por la varianza de los rendimientos del mercado, es decir,

$$\beta = \frac{\text{cov}(r_p, r_m)}{\text{var}(r_m)}.$$

Habiendo presentado el modelo CAPM, ahora tenemos la maquinaria requerida para definir estrategias de mercado neutral.

**Estrategias de Mercado Neutral.** Son aquellas donde el rendimiento de la estrategia no está correlacionado con el rendimiento del mercado. Independientemente de si el mercado sube o baja, en tiempos buenos y malos, la estrategia neutral del mercado se desempeña de manera estable y los resultados se logran normalmente con una menor volatilidad<sup>2</sup>. Este resultado se logra mediante la creación de portafolios neutrales del mercado.

En el contexto del modelo CAPM, los portafolios de mercado neutral pueden definirse como portafolios cuyo  $\beta$  es cero. Para examinar las implicaciones, apliquemos un valor  $\beta$  de cero a la ecuación (1.1). Es fácil ver que el rendimiento del portafolio deja de tener un componente de mercado y está completamente determinado por  $\theta_p$ , el componente residual. En el modelo CAPM, se supone que el componente residual no está correlacionado con los rendimientos del mercado, por lo que el rendimiento del portafolio es neutral para el mercado. Por lo tanto, un portafolio con  $\beta$  cero califica como un portafolio de mercado neutral.

Al trabajar con portafolios neutrales de mercado, podemos concentrarnos en la predicción y la negociación de los rendimientos residuales. Dado que la media sobre el rendimiento residual es cero, es razonable esperar un comportamiento fuerte de *reversión a la media*<sup>3</sup> de la serie de tiempo residual. Este comportamiento de reversión a la media puede entonces ser explotado en el proceso de predicción del rendimiento.

**Modelos de Factores.** Los modelos de factores son usados para explicar las características de los rendimientos de los activos, donde todos los modelos com-

---

<sup>2</sup>La volatilidad es una medida de la frecuencia e intensidad de los cambios del precio de un activo. En términos matemáticos es lo mismo que la desviación estándar.

<sup>3</sup>Alto grado de oscilación sobre el valor medio.

parten la característica común que pueden ser vistos como extensiones al modelo CAPM. La premisa del modelo CAPM es que los rendimientos de los activos son explicados casi por completo por el comportamiento del mercado global. Cada activo es sensible al mercado en su propia forma característica, y esta sensibilidad se denomina beta.

En el modelo CAMP hay un solo factor explicativo y un valor de exposición; llamados, el rendimiento del mercado y el beta, respectivamente. Una extensión natural a esta idea sería tener múltiples factores explicativos y valores de exposición. En este contexto de múltiples factores explicativos, el rendimiento de un activo sería entonces un agregado de las contribuciones de rendimientos de los factores de exposición.

El fundamento teórico de los modelos de factores es proporcionado por la teoría de arbitraje de precios (la APT) que explicamos a continuación.

**Teoría de Arbitraje de Precios o la APT [Vid04].** La teoría de arbitraje de precios fue originalmente propuesta por Stephen A. Ross en 1976. De acuerdo a la APT, si dos valores tienen exactamente los mismos factores de riesgo de exposición, entonces se espera que el rendimiento de los dos valores para un determinado marco de tiempo sea el mismo. El rendimiento actual puede, sin embargo, variar ligeramente debido a los diferentes rendimientos específicos para los dos valores.

En el marco multifactor, un activo se caracteriza totalmente por su perfil expresado mediante los factores de exposición. La contribución del rendimiento en general del activo por cada factor es proporcional a la exposición del activo a los diferentes factores. El rendimiento total es el agregado de las contribuciones. Por lo tanto, si la APT se resumiera en una frase, probablemente sería algo como esto: "Déme el perfil del factor de riesgo de una acción, y le contaré todo acerca de sus características de riesgo y rendimiento". Mostraremos alguna terminología y notación que rodea la APT.

Empezaremos primero con la exposición a factores de riesgo. Siguiendo con la idea de que la APT es una extensión del modelo CAPM, denotaremos las exposiciones de los factores como  $(\beta_1, \beta_2, \beta_3, \dots, \beta_k)$ . Si  $(r_1, r_2, r_3, \dots, r_k)$  denota las

contribuciones de rendimiento de cada factor, entonces el rendimiento del activo se da como

$$r = \beta_1 r_1 + \beta_2 r_2 + \beta_3 r_3 + \dots + \beta_k r_k + r_e,$$

donde  $r_e$  es el rendimiento idiosincrásico o rendimiento específico del activo que no es explicable por los factores del modelo. Una de las suposiciones clave de la APT es que el rendimiento específico para una acción dada no está correlacionado con los rendimientos de los factores ni con los rendimientos específicos de cualquier otra acción.

**Spread.** Sean  $p_t^A$ ,  $p_t^B$  los precios en el tiempo  $t$  de los instrumentos  $A$  y  $B$ , respectivamente. El spread del portafolio  $A$ - $B$  se define como:

$$\log(p_t^A) - \gamma \log(p_t^B).$$

Se refiere a la diferencia entre el precio de compra (venta) de una unidad del instrumento  $A$  y el precio de venta (compra) de  $\gamma$  unidades del instrumento  $B$ .

**Inversión por pares.** La técnica de inversión por pares se refinó y popularizó a mediados de los años 1980's, cuando el *quant* Nunzio Tartaglia conformó un equipo de físicos, matemáticos y expertos en análisis computacional con el objetivo de identificar oportunidades de arbitraje en los mercados de capital. El llamado Tartaglia Group dentro de Morgan Stanley, diseñó un sistema de negociación automatizado que incorporaba la habilidad e intuición de los inversionistas para identificar oportunidades de arbitraje y lo empaquetó en un conjunto de reglas simples de decisión que les permitía identificar pares de acciones cuyos precios se movían conjuntamente. Mediante el empleo de dicha estrategia lograron obtener en 1987 beneficios por 50 millones de dólares. Gerry Bamberger quien era parte del Tartaglia Group fue el responsable de popularizar el modelo de inversión por pares y de implementarlo a través de la modelación cuantitativa y la ejecución computacional.

La inversión por pares se define como una estrategia que consiste en identificar dos activos con trayectorias de precios similares. Si los precios de los activos se mueven conjuntamente, se asume que el spread entre estos, es decir, su distancia relativa, presenta reversión a la media. Cuando el precio de uno de los activos

se incrementa relativamente con respecto del “par”, la estrategia sugiere ponerse corto<sup>4</sup> en el activo sobrevaluado y ponerse largo<sup>5</sup> en el activo subvaluado de forma simultánea.

La inversión por pares ha despertado gran interés en la literatura en los años recientes, con lo cual se han generado un sinnúmero de metodologías para la identificación de los pares y la construcción de reglas de decisión. La presente tesis plantea la aplicación del análisis de cointegración basado en un modelo de arbitraje estadístico para la identificación de activos a fin de obtener un spread con reversión a la media.

La clave del éxito en la inversión por pares radica en la identificación de pares de instrumentos financieros cointegrantes. En un estudio por Gatev et al., se adoptó un enfoque puramente empírico para este fin. Ellos metodológicamente escogían pares basados enteramente en el movimiento de precios históricos de los activos. Además del conjunto de pares elegidos utilizando los precios históricos, otro conjunto de pares fue creado haciendo emparejamientos aleatorios entre los valores. Luego, se compararon las ganancias obtenidas con cada conjunto de pares, donde los rendimientos generados por la metodología de emparejamiento fue mejor que el conjunto de pares de la muestra al azar.

Ahora ajustamos el modelo de cointegración al logaritmo de los precios de los activos. Supondremos que el logaritmo de los precios de los activos es una caminata aleatoria (se lee como no estacionario). Esta suposición es bastante estándar, se ha usado extensamente en modelos de valoración de opciones con resultados satisfactorios. Sean  $\{\log(p_t^A)\}$  y  $\{\log(p_t^B)\}$  series de tiempo no estacionarias, correspondientes a dos activos o instrumentos cointegrantes  $A$  y  $B$  respectivamente. Aplicando el MCE (descrito en la Sección 1.1) tenemos,

$$\log(p_t^A) - \log(p_{t-1}^A) = \alpha_A[\log(p_{t-1}^A) - \gamma \log(p_{t-1}^B)] + \varepsilon_A, \quad (1.2)$$

$$\log(p_t^B) - \log(p_{t-1}^B) = \alpha_B[\log(p_{t-1}^A) - \gamma \log(p_{t-1}^B)] + \varepsilon_B. \quad (1.3)$$

<sup>4</sup>Una posición corta es una estrategia de inversión en la que el inversor vende acciones prestadas. La expectativa del inversionista es que el precio de la acción disminuirá con el tiempo, momento en el cual él comprará las acciones en el mercado abierto y devolverá las acciones al corredor del cual las tomó prestadas.

<sup>5</sup>Una posición larga es la compra de una acción, con la expectativa de que el activo aumentará de valor.

Los parámetros que determinan como único el modelo son el coeficiente de cointegración  $\gamma$  y las dos constantes de la corrección de errores  $\alpha_A$  y  $\alpha_B$ . Por tanto, la estimación del modelo consiste en determinar valores apropiados para  $\alpha_A$ ,  $\alpha_B$  y  $\gamma$ . El lado izquierdo de las ecuaciones (1.2) y (1.3) son los rendimientos de los activos en el periodo de tiempo actual. En el lado derecho, note la expresión para el equilibrio a largo plazo,  $\log(p_{t-1}^A) - \gamma \log(p_{t-1}^B)$ , en ambas ecuaciones. En palabras, es la diferencia a escala del logaritmo de precios. Por cierto, esto coincide con lo que hemos llamado el spread, también note que los subíndices para los precios de los activos en la expresión para el equilibrio a largo plazo es  $t - 1$ . Por tanto, el conocimiento de la realización pasada puede ser usada por nosotros para darnos una predicción de los incrementos en los logaritmos de los precios; es decir, en los rendimientos. A pesar de que ambas poblaciones siguen un proceso log-normal, uno puede ganar cierta previsibilidad en sus declaraciones sobre la base de las realizaciones anteriores.

Ahora nos enfocaremos en la parte de cointegración. Del Teorema de Representación de Granger, sabemos que la serie  $\log(p_{t-1}^A) - \gamma \log(p_{t-1}^B)$  es estacionaria y con media reversible. Hasta aquí, conocemos todo acerca de predecir series de tiempo estacionarias. Si tan solo pudiéramos asociar la serie de tiempo del spread a un portafolio, podríamos considerar negociar el portafolio basados en nuestra predicción de los valores de la serie de tiempo. Se procede a hacer exactamente eso. Considere un portafolio con una acción larga de  $A$  y una acción corta de  $B$ . El rendimiento del portafolio para un periodo de tiempo está dada como

$$[\log(p_{t+i}^A) - \log(p_t^A)] - \gamma [\log(p_{t+i}^B) - \log(p_t^B)].$$

Reorganizando los términos tenemos,

$$[\log(p_{t+i}^A) - \gamma \log(p_{t+i}^B)] - [\log(p_t^A) - \gamma \log(p_t^B)] = \text{spread}_{t+i} - \text{spread}_t. \quad (1.4)$$

Por tanto, el rendimiento del portafolio se calcula a partir del incremento del valor del spread en el lapso de tiempo  $i$ . Lo que hemos hecho es asociar exitosamente un portafolio con una serie de tiempo estacionaria. En la Sección 2.2 daremos una interpretación más puntual del coeficiente de cointegración  $\gamma$ .

La discusión hasta ahora describe brevemente cómo podríamos operar una vez que conocemos dos activos que están cointegrados. Pero, ¿cómo identificamos pares de activos?, ¿podemos comprobar que efectivamente son cointegrantes?, ¿podemos determinar el coeficiente de cointegración? Respondemos a estas preguntas en la Sección 2.2. Ya una vez identificadas las parejas potenciales, se verifica la hipótesis propuesta de que los pares de activos son, en efecto cointegrantes basados en pruebas estadísticas a partir de datos históricos. Luego se construye una estrategia de inversión que se discutirá en la Sección 2.3, por ahora daremos una breve introducción.

La idea es tomar las decisiones de inversión basándonos en la oscilación del spread  $s(t)$  alrededor de su valor medio  $\mu$ . Para esto se fija un valor  $\Delta > 0$  y se realiza alguna de las dos estrategias de inversión que se explican a continuación.

**Estrategia 1.** Si en el momento  $t_0$  se cumple que  $s(t_0) - \mu < -\Delta$  entonces se compra<sup>6</sup> el portafolio. Si en lugar de esto se cumple que  $s(t_0) - \mu > \Delta$  entonces se vende<sup>7</sup> el portafolio. Se da por terminada la inversión en el primer instante  $t_1$  en el que el spread sufra reversión a la media. En esta estrategia el rendimiento por inversión es:

$$\log(p_{t_1}^A) - \gamma \log(p_{t_0}^B) = \Delta$$

**Estrategia 2.** Se compra el portafolio cuando  $s_{t_0} - \mu < -\Delta$  y se vende cuando  $s_{t_1} - \mu > \Delta$ . En este caso la ganancia es:

$$\log(p_{t_1}^A) - \gamma \log(p_{t_0}^B) = 2\Delta.$$

La diferencia es que en la Estrategia 2 se hacen menos inversiones con mayor ganancia. Esto ayuda a reducir las comisiones que se le pagan al intermediario. Sin embargo, para un spread particular la mejor estrategia depende del spread mismo y se determina haciendo simulaciones sobre registros históricos. Ambas estrategias se presentan con mayor detalle en el Capítulo 2, por ahora lo que nos interesa es dar una idea de cómo se determina  $\Delta$ .

El parámetro  $\Delta$  se calcula como el valor que maximiza el rendimiento total

---

<sup>6</sup>Largo en  $A$  y corto en  $B$

<sup>7</sup>Corto en  $A$  y en largo  $B$

sobre los registros históricos recientes del spread considerando el número total de inversiones realizadas. Así, el valor de  $\Delta$  depende de cada spread y se determina mediante un proceso de simulación que se explica con detalle en el Capítulo 8 de [Vid04]. En el Apéndice A se incluye el código en R asociado al método.

Es muy importante entender que estas estrategias funcionan sólo cuando el spread es estacionario. En la práctica puede suceder que en cualquier momento el spread deje de ser estacionario y las estrategias generen pérdidas irremediables.

## Capítulo 2

# Modelo de inversión por pares

En el Capítulo 1, mencionamos que la estrategia de inversión por pares se divide en tres pasos, los cuales son: identificación de los pares de activos, pruebas de cointegración y formulación de las reglas de inversión. En este capítulo, nos centraremos en cada uno de ellos, dando una idea general de cómo proceder en este modelo.

### 2.1. Identificación de pares de activos

En esta parte de la estrategia se filtran los pares de activos disponibles mediante un proceso “rugoso” pero computacionalmente barato. En general, determinar si un par de activos es cointegrante es un proceso computacionalmente costoso. Así, primero se aplica una técnica poco precisa para eliminar pares que de manera evidente no sean cointegrantes. Después, con los pares restantes se aplican métodos mucho más finos de análisis de cointegración.

La parte teórica contenida en esta sección ha sido tomada de [Vid04, págs 85-94, Cap 6]. Se han agregado demostraciones formales y comentarios no incluidos en esta fuente.

La metodología consiste en asociar con cada par de instrumentos una me-

medida de qué tan distantes están de ser cointegrantes. Así, el proceso rugoso de selección de pares descarta aquellos cuya distancia es mayor que cierto umbral pre-establecido. Por lo tanto, una adecuada elección de la medida de la distancia es clave para el proceso de selección de pares. Para escoger la medida de distancia, se explotarán los nexos entre cointegración, teoría de arbitraje de precios (APT) y modelo de tendencias comunes. Vamos a empezar con el modelo de tendencias comunes.

### Tendencias comunes en el modelo de cointegración

Sean las series de tiempo  $y_t$  y  $z_t$  como en la ecuación (2.1).

$$\begin{aligned} y_t &= n_{y_t} + \varepsilon_{y_t} \\ z_t &= n_{z_t} + \varepsilon_{z_t} \\ r_{y_{t+1}} &= n_{y_{t+1}} - n_{y_t} \end{aligned} \quad (2.1)$$

donde  $n_{y_t}$  y  $n_{z_t}$  representan las tendencias comunes;  $\varepsilon_{y_t}$ ,  $\varepsilon_{z_t}$  son las componentes estacionarias y específicas de las series de tiempo y  $r_{y_t}$  es la innovación derivada de la componente de tendencia común de la serie  $y_t$ . Recordemos de la Sección 1.1, que si dos series cointegran entonces sus tendencias comunes deben ser idénticas excepto por un escalar. Por tanto, si  $y_t$  y  $z_t$  son dos series de tiempo como en la ecuación (2.1) y además son cointegrantes, entonces las secuencias de innovación derivadas de las componentes de tendencia común deben estar perfectamente correlacionadas, es decir,  $cor(r_{y_{t+1}}, r_{z_{t+1}}) = \pm 1$ , ahora veamos por qué esto es cierto

Las secuencias de innovación de las series  $y_t$  y  $z_t$  están dadas por:

$$\begin{aligned} n_{y_{t+1}} - n_{y_t} &= r_{y_{t+1}}, \\ n_{z_{t+1}} - n_{z_t} &= r_{z_{t+1}}. \end{aligned} \quad (2.2)$$

De acuerdo al modelo, las tendencias comunes deben ser idénticas excepto por un escalar, es decir, para los marcos de tiempo  $t$  y  $t + 1$  debe suceder lo siguiente:

$$\begin{aligned}n_{y_t} &= \gamma n_{z_t}, \\n_{y_{t+1}} &= \gamma n_{z_{t+1}}.\end{aligned}\tag{2.3}$$

Luego, de las ecuaciones (2.2) y (2.3), se tiene

$$\begin{aligned}n_{y_{t+1}} - n_{y_t} &= \gamma(n_{z_{t+1}} - n_{z_t}) \\r_{y_{t+1}} &= \gamma r_{z_{t+1}}\end{aligned}\tag{2.4}$$

Esto significa que las series  $r_{y_{t+1}}$  y  $r_{z_{t+1}}$  deben ser idénticas excepto por un escalar, el cual será llamado de ahora en adelante *coeficiente de cointegración*. De la Ecuación 2.4, se deduce que el coeficiente de cointegración se puede obtener mediante la realización de una regresión simple. Así, el coeficiente de cointegración está dado como:

$$\gamma = \frac{\text{cov}(r_y, r_z)}{\text{var}(r_z)}.$$

La definición del modelo de tendencias comunes se basa en que las componentes específicas deben ser series estacionarias. Por tanto, para que exista cointegración, es absolutamente necesario que la componente específica sea estacionaria. Así, la primera diferenciación de la componente específica no debe ser un ruido blanco, porque si lo fuera, entonces la serie específica sería una caminata aleatoria, es decir, una serie no estacionaria. Esto contradice la condición de estacionariedad para la componente específica.

Finalmente, concluimos que hay dos condiciones que necesitan ser satisfechas para asegurar cointegración en un modelo de tendencias comunes.

- I Las secuencias de innovación derivadas de las tendencias comunes de dos series deben ser idénticas excepto por un escalar.
- II Las componentes específicas de las dos series deben ser estacionarias.

Habiendo establecido el modelo de tendencias comunes, regresamos a la cuestión de la aplicación de esto a los datos de activos. La aplicación es posible sólo si separamos las series de tiempo en componentes de tendencia común y componente específica/estacionaria. Para hacer esto, buscamos la teoría de arbitraje

de precios (la APT) y establecemos una conexión entre la APT y el modelo de tendencias comunes.

### 2.1.1. Modelo de tendencias comunes y la APT

Consideremos el logaritmo de los precios de los activos como la suma de una caminata aleatoria y una serie estacionaria

$$\log(\text{price}_t) = n_t + \varepsilon_t, \quad (2.5)$$

donde  $n_t$  es una caminata aleatoria, y  $\varepsilon_t$  es la componente estacionaria. Diferenciando el logaritmo de los precios de los activos se obtiene la secuencia de los rendimientos. Por tanto, basados en la ecuación (2.5), el rendimiento  $r_t$  en el tiempo  $t$  puede ser separado en dos partes

$$\begin{aligned} \log(\text{price}_t) - \log(\text{price}_{t-1}) &= n_t - n_{t-1} + (\varepsilon_t - \varepsilon_{t-1}) \\ r_t &= r_t^c + r_t^s \end{aligned}$$

donde  $r_t^c$  es el rendimiento debido a la componente de tendencia no estacionaria, y  $r_t^s$  es el rendimiento debido a la componente estacionaria.

Note que el rendimiento debido a la componente de tendencia es similar a  $r_{y_t}$  en la ecuación (2.1). Por lo tanto, el criterio de cointegración puede reformularse de la siguiente manera: si dos activos son cointegrantes, los rendimientos de sus tendencias comunes deben ser idénticas excepto por un escalar. ¿Pero por qué los activos deben tener rendimientos comunes?, ¿hay una razón financiera para que esto exista entre activos? La respuesta es afirmativa, y la APT da explicación para esto. Recordando las discusiones anteriores sobre APT, los rendimientos de los activos pueden ser separados en rendimientos del factor común (rendimientos basados en la exposición de los activos a diferentes factores de riesgo) y los rendimientos específicos (rendimientos específicos de los activos). Si dos activos comparten el mismo perfil de exposición a factores de riesgo, entonces los rendimientos de los factores comunes para ambos activos deben ser los mismos. Esto justifica por qué es razonable esperar que los activos tengan el mismo rendimiento

en la componente común.

Ahora estamos listos para establecer paralelismos entre el modelo de las tendencias comunes y la APT. De acuerdo con la APT, los rendimientos de activos para un periodo de tiempo pueden ser separados en dos tipos: rendimientos del factor común y rendimientos específicos. Correspondientes a la tendencia común de la innovación y a la primer diferencia de la componente específica en el modelo de tendencias comunes. Para que la correspondencia sea válida, la diferenciación de los rendimientos específicos debe ser una serie de tiempo estacionaria. Alternativamente, como se discutió anteriormente, los rendimientos específicos no deben ser un ruido blanco. Pero APT es sólo un modelo y no puede proporcionarnos garantía alguna sobre la serie de tiempo de los rendimientos específicos. Por lo tanto, tendríamos que asumir que la serie de rendimientos específicos no sea un ruido blanco. Sin embargo, es tranquilizador saber que esta suposición se verifica cuando se corren pruebas de cointegración; la justificación de esta suposición la veremos al final de la sección.

**Observación.** Un par de instrumentos con el mismo perfil de exposición a factores de riesgo satisface las condiciones necesarias para cointegración según el modelo de tendencias comunes.

Veamos que se cumplen las dos condiciones:

Condición 1. Consideraremos dos instrumentos  $A$  y  $B$  con exposición a vectores de riesgo  $\gamma x$  y  $x$ , respectivamente. Los vectores de exposición de los factores en este caso son idénticos excepto por un escalar. Denotemos los factores de exposición como

$$\text{Activo A: } \gamma x = (\gamma x_1, \gamma x_2, \gamma x_3, \dots, \gamma x_n)$$

$$\text{Activo B: } x = (x_1, x_2, \dots, x_n)$$

Geoméricamente, puede interpretarse que los vectores de factores de exposición de los dos activos apuntan en la misma dirección; es decir, el ángulo entre ellos es cero.

Si  $b = (b_1, b_2, \dots, b_n)$  es el vector de factores de rendimiento, y  $r_A^{spec}$  y  $r_B^{spec}$

son los rendimientos específicos para los activos  $A$  y  $B$ , entonces los rendimientos para los activos  $r_A$  y  $r_B$  están dados como

$$\begin{aligned} r_A &= \gamma(x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_n b_n) + r_A^{spec} \\ r_B &= (x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_n b_n) + r_B^{spec} \end{aligned}$$

Los rendimientos para los factores comunes para los activos son por tanto

$$\begin{aligned} r_A^{cf} &= \gamma(x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_n b_n) \\ r_B^{cf} &= x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_n b_n \end{aligned}$$

Así,  $r_A^{cf} = \gamma r_B^{cf}$  las sucesiones son idénticas excepto por un escalar. Esto satisface la primera condición de cointegración según el modelo de tendencias comunes.

Condición 2. Ahora consideremos la combinación lineal de los rendimientos

$$r_A - \gamma r_B = (r_A^{cf} - \gamma r_B^{cf}) + (r_A^{spec} - \gamma r_B^{spec})$$

El lado izquierdo de la ecuación representa el rendimiento de un portafolio en largo de un activo  $A$  y corto de  $\gamma$  veces un activo  $B$ . El lado derecho muestra que este rendimiento es separable en un rendimiento de factor común y un rendimiento específico del portafolio. Por tanto

$$r_{port} = r_{port}^{cf} + r_{port}^{spec}, \quad (2.6)$$

donde

$$\begin{aligned} r_{port}^{cf} &= r_A^{cf} - \gamma r_B^{cf} \\ r_{port}^{spec} &= r_A^{spec} - \gamma r_B^{spec} \end{aligned}$$

Note que si los instrumentos  $A$  y  $B$  son cointegrantes, entonces el rendimiento del factor común se convierte en cero. Adicionalmente, el rendimiento del portafolio

folio largo-corto puede ser también visto como el resultante de la diferenciación de la serie de tiempo del spread, tal como en la ecuación (1.4). Basados en la separación del rendimiento de las series, la serie del spread puede ser también representada como la suma de las dos componentes. Uno de ellos es el factor común del rendimiento integrado, el cual llamaremos el spread del factor común. El otro es el rendimiento integrado específico, el cual puede ser llamado el spread específico. Escribiendo esto en forma de ecuación, tenemos

$$spread_t^{cf} - spread_{t-1}^{cf} = r_{port}^{cf} \quad (2.7)$$

$$spread_t^{spec} - spread_{t-1}^{spec} = r_{port}^{spec} \quad (2.8)$$

$$spread_t^{port} - spread_{t-1}^{port} = r_{port} \quad (2.9)$$

luego, de las Ecuaciones (2.6) y (2.9), se tiene que

$$spread_t^{port} - spread_{t-1}^{port} = (spread_t^{cf} + spread_t^{spec}) - (spread_{t-1}^{cf} + spread_{t-1}^{spec}),$$

de donde, como  $spread_0^{port} = spread_0^{cf} = spread_0^{spec} = 0$ , entonces, para todo  $t$ ,

$$spread_t^{port} = spread_t^{cf} + spread_t^{spec}$$

Note que si los instrumentos son cointegrantes, entonces  $r_{port}^{cf} = 0$ . Por tanto,  $spread_t^{port} = spread_t^{spec}$ , lo cual quiere decir que el spread del portafolio es el mismo al spread debido a las componentes específicas. Adicionalmente, la serie del spread debe ser estacionaria para que se dé la cointegración. Esto sería verdad si el spread debido al spread específico fuera estacionario. El spread específico será estacionario si la integración del rendimiento específico de los rendimientos específicos individuales de los stocks es estacionario. Esto es ciertamente la suposición que hacemos cuando relacionamos APT con cointegración. Por tanto, si esta sucesión se lleva a cabo, tendremos satisfechas todas las condiciones necesarias para cointegración.

**La medida de la distancia.** Ahora estamos listos para definir la medida de distancia que usaremos en lo que hemos llamado proceso “rugoso” de selección. Recuerde que en los debates sobre el modelo de las tendencias comunes que la condición necesaria para la cointegración es que las secuencias de innovación

derivadas de las tendencias comunes deben estar perfectamente correlacionados. También se estableció que el rendimiento del factor común del modelo APT podría interpretarse como las innovaciones derivadas de las tendencias comunes. Por lo tanto, la correlación entre las secuencias de innovación es la correlación entre los rendimientos de los factores comunes. La medida de la distancia que se propone es:

$$d(A, B) = |cor(r_A, r_B)| = |\rho| = \left| \frac{cov(r_A, r_B)}{\sqrt{var(r_A)var(r_B)}} \right|. \quad (2.10)$$

Finalmente, concluimos que la perfecta alineación de los vectores de exposición a factores, es decir, un ángulo cero entre ellos, es indicativo de cointegración.

**Las pruebas de cointegración validan la afirmación: los rendimientos específicos no son ruidos blancos.** Sean  $A$  y  $B$  dos instrumentos de un mercado arbitrario, supongamos que bajo una prueba de cointegración obtuvimos que el spread específico del par  $A$ - $B$  es estacionario, este spread lo denotaremos por  $spread_t^{spec}$ . Si  $spread_t^{spec}$  es estacionario, entonces  $r_A^{spec} - \gamma r_B^{spec}$  es también estacionario, dado que es la primera diferencia del spread y por la Ecuación 1.4, tenemos  $r_A^{spec} - \gamma r_B^{spec} = spread_{t+i}^{spec} - spread_t^{spec}$ . Ahora, veamos que  $r_A^{spec}$  y  $r_B^{spec}$  no pueden ser ruidos blancos. Razonando por reducción al absurdo supongamos que  $r_A^{spec}$  y  $r_B^{spec}$  son en efecto ruidos blancos. Suponiendo esto podría pasar que los rendimientos sean debidos a caminatas aleatorias, es decir, tanto  $r_A^{spec}$  como  $r_B^{spec}$  son las innovaciones de dos caminatas aleatorias, digamos que estas caminatas aleatorias son  $\theta^A$  y  $\beta^B$  respectivamente, es decir, se cumple que:

$$\begin{aligned} r_A^{spec} &= \theta_{t+i}^A - \theta_t^A \\ r_B^{spec} &= \beta_{t+i}^B - \beta_t^B, \end{aligned}$$

Por tanto,

$$spread_t^{spec} = \theta_t^A - \gamma \beta_t^B,$$

lo cual quiere decir que  $var(spread_t^{spec}) = var(\theta_t^A - \gamma \beta_t^B)$ , pero por definición de caminata aleatoria se tiene que

$$\begin{aligned}\theta_t^A &= r_A^{spec}(t) + r_A^{spec}(t-1) + \dots + r_A^{spec}(1) \\ \beta_t^B &= r_B^{spec}(t) + r_B^{spec}(t-1) + \dots + r_B^{spec}(1)\end{aligned}$$

donde, para  $t$  y  $t-i$  arbitrarios, los rendimientos  $r_A^{spec}(t)$  y  $r_A^{spec}(t-i)$  no están correlacionados, al igual que los rendimientos  $r_B^{spec}(t)$  y  $r_B^{spec}(t-i)$ . Por otro lado, dado que en la APT los rendimientos específicos de dos instrumentos no están correlacionados con los rendimientos específicos de cualquier otro instrumento, entonces  $r_A^{spec}(t)$  y  $r_B^{spec}(t)$  no están correlacionados para cualquier  $t$ . Así, la varianza del spread específico está dada por:

$$\begin{aligned}\text{var}(\text{spread}_t^{spec}) &= \text{var}\left(\sum_{k=1}^t r_A^{spec}(k) - \gamma \sum_{k=1}^t r_B^{spec}(k)\right) \\ &= \sum_{k=1}^t \text{var}(r_A^{spec}(k)) + \gamma^2 \sum_{k=1}^t \text{var}(r_B^{spec}(k)) \\ &= t \text{var}(r_A^{spec}(1)) + \gamma^2 t \text{var}(r_B^{spec}(1)) \\ &= t [\text{var}(r_A^{spec}(1)) + \gamma^2 \text{var}(r_B^{spec}(1))].\end{aligned}$$

Los ruidos blancos son series de tiempo estacionarias, lo cual indica que sus varianzas son constantes en el tiempo, en este caso se tiene que  $\text{var}(r_A^{spec}(1))$  y  $\text{var}(r_B^{spec}(1))$  son constantes. Por tanto, concluimos que  $\text{var}(\text{spread}_t^{spec})$  es una variable que depende del tiempo  $t$ , es decir, no es constante, lo cual es absurdo pues habíamos supuesto que el spread específico era estacionario, es decir, de varianza constante. Finalmente, concluimos que los rendimientos específicos en efecto no son ruidos blancos como lo queríamos demostrar.

## 2.2. Prueba de cointegración

Existen pruebas de cointegración basadas fundamentalmente en el análisis de raíces unitarias [MK98]. Sin embargo una condición importante que debe cumplir el spread es que presente una varianza muy grande ya que de eso depende el monto de la ganancia que se puede esperar. Una cointegración “perfecta” signifi-

ca que los instrumentos que estamos usando son exactamente el mismo y nunca vamos a observar que un instrumento se desvíe del otro. Por esta razón, en principio, decidimos no utilizar las pruebas conocidas sino hacer más que pruebas de cointegración, pruebas del grado de reversión a la media y otras características propias de las series cointegrantes. Incluso no pedimos reversión estricta a la media y permitimos que la media tenga “corrimientos” a lo largo del análisis de la serie, para que la selección de pares de candidatos sea mucho más flexible. De hecho, al analizar los pares de candidatos, podemos casi siempre esperar tener una exposición a un factor residual que causa el fenómeno del promedio móvil que es a lo que nos referimos con “corrimientos”, la implementación para el cálculo de la media móvil se encuentra en el Apéndice B.

En la sección anterior, vimos cómo encontrar pares de instrumentos que posiblemente son cointegrantes, lo que hacemos ahora es construir una prueba más rigurosa de cointegración. Supongamos que dos instrumentos  $A$  y  $B$ , superaron el filtro de buena alineación de sus vectores de exposición a factores, sus precios en el tiempo  $t$  son,  $p_t^A$  y  $p_t^B$ , respectivamente.

Lo primero que se debe hacer es encontrar una relación lineal entre los logaritmos de los precios de  $A$  y  $B$ , para esto se hace una regresión lineal, la cual nos proporciona el *coeficiente de cointegración*  $\gamma$ , siendo así tenemos que el spread de precios ajustados esta dado por:

$$spread := \log(p_t^A) - \gamma \log(p_t^B). \quad (2.11)$$

Al spread se le resta su media  $\mu$ , así reversión a la media se transforma en reversión a cero, y esto lo denotaremos por  $s$ , es decir,

$$s = spread - \mu, \quad (2.12)$$

es el spread con media cero. De aquí en adelante cada vez que se mencione el spread de dos instrumentos supondremos implícitamente que su media es cero. También se calcula el promedio móvil del spread en la forma que se explica en el Apéndice B, esto corresponde al suavizado de  $s$  que denotaremos por  $s_s$ , y por último construimos lo que llamamos spread ajustado,  $s_{adj} := s - s_s$ .

La prueba para asegurar estacionariedad en el spread y concluir que los pares cointegran debe pasar por los siguientes filtros:

1. Se debe asegurar que el spread tenga un mínimo necesario de registros históricos, el parámetro  $A$  en el Algoritmo 2.1 controla esta situación.
2. Se revisa que la variación entre el máximo de  $s_t$  y su mínimo no sea demasiado grande. En el Algoritmo 2.1,  $B$  es la variación máxima permitida.
3. Una característica fundamental en series estacionarias es una fuerte reversión a la media, es por eso que se calcula la serie de intervalos de tiempo entre reversiones. El último intervalo se calcula como la diferencia entre el momento del último cruce y el final de la serie. Así, las longitudes de los intervalos primero y último son sólo cotas inferiores de reversión, de hecho, éstas no se pueden precisar sin tener más registros del spread. Lo que al final nos importa es que estas cotas no sean demasiado grandes. Finalmente, si la tasa de reversión a la media es demasiado pequeña o si algún lapso de reversión es demasiado grande, entonces el par se descarta como cointegrante, en el Algoritmo 2.1 se usan los parámetros  $C$  y  $D$  para la cotas de tasa y lapso de reversión a la media, respectivamente.
4. Se revisa que el percentil 90 en la serie de lapsos de reversión a la media no se aleje mucho de cero. Como hay muy pocos datos en la serie, el riesgo de sesgo se reduce recurriendo al remuestreo con reemplazo *Bootstrap* ver en [DH97], donde buscamos estimar la distribución de tiempo entre dos cruces. El valor 90 se almacena en el parámetro  $F$  y se puede cambiar por cualquier otro valor entre 0 y 100.
5. Por último, sean  $mean$  y  $sd$  la media y la desviación estándar de la distribución muestral del estadístico  $S$  (en este caso el estadístico serán los percentiles que son medidas estadísticas de posición que tienen la propiedad de dividir los términos de la serie en cien grupos iguales). Entonces, si la distribución es normal (que es cierto para muchos estadísticos, si el tamaño de la muestra es  $n \geq 30$ ), cabe esperar que el estadístico se encuentre en el intervalo  $(mean - 2sd, mean + 2sd)$  el 95.45 % de las veces.

Finalmente controlamos el limite superior del intervalo con la cota  $E$  como en el Algoritmo 2.1.

---

**Algoritmo 2.1** Prueba de estacionariedad
 

---

```

1: procedure STATIONARITYTEST( $s, s_s, s_{adj}$ )
2:    $A := 100$ 
3:    $B := 0.15$ 
4:    $C := 0.08$ 
5:    $D := 60$ 
6:    $E := 25$ 
7:    $F := 0.90$ 
8:   lengthSpread := La longitud de  $s$ 
9:   if lengthSpread <  $A$  then
10:    return (FALSE)
11:   diffSpread :=  $\max(s_s) - \min(s_s)$ 
12:   if diffSpread >  $B$  then
13:    return (FALSE)
14:    $cValue := s_{adj}[1]$ 
15:    $pastCrossing := 0$ 
16:    $zeroCrossings := []$ 
17:    $m :=$  Número de filas del vector  $s.adjusted$ 
18:   for  $i = 2$  to  $m$  do
19:     if  $\left( \begin{array}{l} s_{adj}[i] = 0 \text{ or} \\ ((cValue > 0) \text{ and } (s_{adj}[i] < 0)) \text{ or} \\ ((cValue < 0) \text{ and } (s_{adj}[i] > 0)) \end{array} \right)$  then
20:        $i - pastCrossing$  se inserta al final de la lista  $zeroCrossings$ 
21:        $pastCrossing = i$ 
22:        $cValue = s_{adj}[i]$ 
23:   if  $i - pastCrossing > 0$  then
24:      $i - pastCrossing$  se inserta al final de la lista  $zeroCrossings$ 
25:    $n :=$  Tamaño de  $zeroCrossings$ 
26:    $rata := n/m$ 
27:    $lapso := \max(zeroCrossings)$ 
28:   if ( $rata < C$  or  $lapso > D$ ) then
29:    return (FALSE)
30:    $l := 1000$ 
31:    $boot.samples :=$  Matriz de tamaño  $l \times n$ , donde cada fila tiene un
     remuestreo con reemplazamiento de  $zeroCrossings$ .
32:    $boot.bsp :=$  Contiene el percentil 90 ( $F$ ) de cada fila de  $boot.samples$ 
33:    $mean =$  Media de  $boot.bsp$ 

```

---

---

```
34:   sd = Desviación estándar de boot.bsp
35:   booting := mean + 2sd
36:   if booting > E then
37:     return (FALSE)
38:   return (TRUE)
39: end procedure
```

---

Las parámetros  $A$ ,  $B$ ,  $C$ ,  $D$  y  $E$  en este momento tienen valores arbitrarios, pero más adelante, en la Sección 4.2 daremos un método para encontrar el mejor ajuste posible de acuerdo a pruebas de estacionariedad ampliamente probadas.

### 2.3. Estrategias de inversión

Hasta ahora hemos estudiado la forma de obtener pares de instrumentos cointegrantes de un mercado dado. Recapitulando un poco, recordemos que se parte de un mercado  $\mathcal{M}$  de  $n$  instrumentos. A este conjunto se le seleccionaron al azar pares de instrumentos, los cuales eran filtrados por la medida de la distancia para asegurar cierta alineación entre estos, logrando así, obtener un subconjunto  $\mathcal{T}$  de  $\mathcal{M}$ , tal que  $|\mathcal{T}| \leq n$ . Luego, con ayuda del Algoritmo 2.1, se filtran las parejas en el conjunto  $\mathcal{T}$  que son cointegrantes, obteniendo el conjunto  $\mathcal{C}$ , que contiene pares de instrumentos cointegrantes. Note que, hasta ahora, dado un mercado arbitrario, se logra identificar pares de instrumentos cointegrantes. Dado que se quiere llevar a la práctica todas estas teorías, es interesante conocer en términos financieros, si un par cointegrante de  $\mathcal{C}$  permite algún movimiento en el mercado, o más concretamente obtener alguna ganancia de él.

Es por esto que al nuevo conjunto  $\mathcal{C}$  le aplicaremos un nuevo filtro, el cual, permite conocer qué pares de instrumentos en  $\mathcal{C}$  entregan algún porcentaje de ganancia. Para lograrlo se implementan dos estrategias de inversión que se pueden encontrar en el Capítulo 8 de [Vid04], cuyas implementaciones veremos en las Secciones 2.3.1 y 2.3.2. Por ahora, vamos a definir algunos conceptos que son necesarios para ambas estrategias.

Se trabaja con un año financiero, que corresponde a 252 días, donde a cada

instrumento del mercado se le debe averiguar el historial de precios, el cual, para nuestros propósitos, mínimamente debe tener por día, los precios alto, bajo y ajustado, para de esta forma tener tres series de tiempo, que corresponden a los precios del año en cuestión.

Supongamos que tenemos dos instrumentos  $A$  y  $B$  que cointegran, es decir, la pareja  $\{A - B\} \subseteq \mathcal{C}$ , lo que se hace es descargar sus series de precios alto, bajo y ajustado, es decir, al instrumento  $A$ , le corresponde su serie de precios alto  $ph_t^A$ , la serie de precios bajo  $pl_t^A$  y la serie de precios ajustada  $p_t^A$ ; e igualmente a  $B$ , le corresponden las series de tiempo  $ph_t^B$ ,  $pl_t^B$  y  $p_t^B$  para alto, bajo y ajustado, respectivamente.

Ahora con ayuda de la ecuación (2.11) se calcula el spread alto del par  $A - B$ , de la siguiente manera:

$$spread_h = \log(ph_t^A) - \gamma \log(ph_t^B). \quad (2.13)$$

Igualmente el spread bajo para el par  $A - B$  es:

$$spread_l = \log(pl_t^A) - \gamma \log(pl_t^B), \quad (2.14)$$

donde, claramente  $\gamma$  es el coeficiente de cointegración del par  $A - B$ . El spread de los precios ajustados es tal cual como en la ecuación (2.11). Finalmente a cada una de las series le aplicamos reversión a cero, siguiendo la ecuación (2.12), tenemos que los spreads alto y bajo con los que vamos a trabajar son, respectivamente:

$$\begin{aligned} sh &= spread_h - \mu_h, \\ sl &= spread_l - \mu_l, \end{aligned}$$

donde  $\mu_h$  y  $\mu_l$ , corresponden a las medias de  $spread_h$  y  $spread_l$ , respectivamente. El spread con reversión a la media de los precios ajustados es tal cual como en la ecuación (2.12). Para usos de las dos estrategias de inversión que vamos a utilizar, es necesario calcular también el spread suavizado de  $s$  como en el Apéndice B, es decir,  $s_s$  y también calcular el umbral  $\Delta$  como en el Apéndice A.

### 2.3.1. Primera estrategia de inversión

Se define un autómata determinista  $M1 = (Q, \Sigma, \varphi, s, F)$ , donde

- $Q = \{0, 1, 2\}$ , los elementos de  $Q$  son llamados estados;
- $\Sigma = \{a, b, c, d\}$ , es el alfabeto de entrada para el autómata, que en nuestro caso tiene el siguiente significado:

$a : sh[i] \geq s_s[i] + \Delta$ , lo cual significa que, en el día  $i$ , el spread alto, debe estar como mínimo una cantidad  $\Delta$  más arriba del spread suavizado.

$b : sl[i] \leq s_s[i] - \Delta$ , es decir, en el día  $i$ , el spread bajo, debe estar como mínimo una cantidad  $\Delta$  más abajo del spread suavizado.

$c : sl[i] \leq s_s[i]$ , en el día  $i$ , el spread bajo esta por debajo del spread suavizado.

$d : sh[i] \geq s_s[i]$ , se quiere que en el día  $i$ , el spread alto este por encima del spread suavizado..

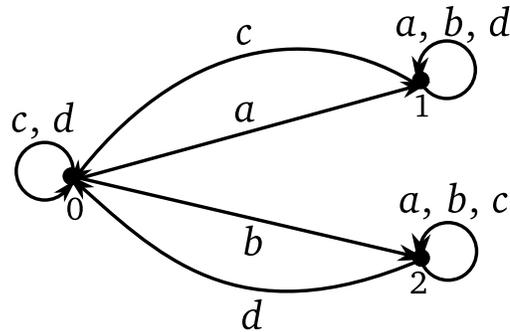
- $\varphi : Q \times \Sigma \rightarrow Q$  es la función de transición. Intuitivamente,  $\varphi$  es una función que indica a qué estado se debe mover en respuesta a una entrada, se comporta de la siguiente manera:

$\varphi$	$a$	$b$	$c$	$d$
0	1	2	0	0
1	1	1	0	1
2	2	2	2	0

Tabla 2.1: Función de transición de  $M1$

- $k = 0 \in Q$ , es el estado de inicio para el primer día, para los otros 251 días del año financiero, el estado inicial puede ser cualquier elemento de  $Q$ ;
- $F \subseteq Q$ , los elementos de  $F$  son llamados estados finales, no especificamos cuales son los estados finales, porque en nuestro autómata cualquier elemento en  $Q$  podrá ser estado final.

Gráficamente el autómata  $M1$  tiene este aspecto:

Figura 2.1: Diagrama de  $M1$ 

Ahora, vamos a entender como  $M1$  nos ayuda a encontrar las ganancias para un par cointegrante dado. Consideremos las siguientes variables  $g$ ,  $shigh$  y  $slow$  que en principio se toman como cero, pues son sólo contadores. Supongamos que estamos en un día arbitrario  $i$  del año, puede pasar lo siguiente:

Si estamos en el estado 0 y ocurre el evento  $a$ , se entra al mercado por arriba inicializando la variable  $shigh = sh[i]$ . Por otro lado, si ocurre el evento  $b$  se entra al mercado por debajo inicializando la variable  $slow = sl[i]$ .

Si estamos en el estado 1 y ocurre el evento  $c$ , nos salimos del mercado y se ejecuta el valor de la ganancia como:  $g = g + shigh - sl[i]$ .

Y finalmente si estando en el estado 2 ocurre  $d$ , nos salimos del mercado y se realiza el calculo de la ganancia así:  $g = g + sh[i] - slow$ .

Al final del año, es decir, cuando  $i = 252$  se calcula la ganancia neta como:

$$ganancia = 100g,$$

lo multiplicamos por 100 para que quede en términos de porcentajes.

### 2.3.2. Segunda estrategia de inversión

Se define un autómata determinista  $M2 = (Q, \Sigma, \varphi, s, F)$ , donde

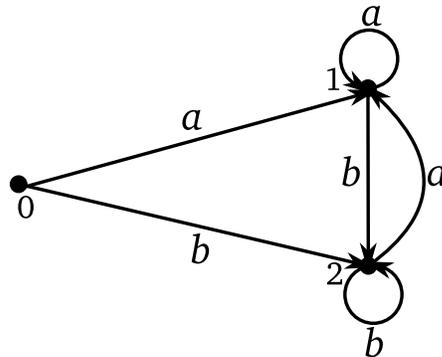
- $Q = \{0, 1, 2\}$

- $\Sigma = \{a, b\}$ , en este caso  $a$  y  $b$  se comportan igual que en el autómata  $M1$
- $\varphi : Q \times \Sigma \rightarrow Q$  es la función de transición, cuyos movimientos se muestran en la siguiente tabla:

$\varphi$	$a$	$b$
0	1	2
1	1	2
2	1	2

Tabla 2.2: Función de transición de  $M2$ 

- El estado inicial  $k$  y los estados finales  $F$  se definen tal cual como en  $M1$ .

Figura 2.2: Diagrama de  $M2$ 

Supongamos nuevamente que estamos en un día  $i$  arbitrario del año, y consideremos las siguientes variables  $g$ ,  $shigh$  y  $slow$ .

Igual que en la primera estrategia, si estamos en el estado 0 y ocurre el evento  $a$ , se entra al mercado por arriba inicializando la variable  $shigh = sh[i]$ . Por otro lado, si ocurre el evento  $b$  se entra al mercado por debajo inicializando la variable  $slow = sl[i]$ .

Ahora, si estamos en el estado 1 y ocurre el evento  $b$ , no nos salimos de mercado, se empieza otra transacción entrando por debajo del mercado y se ejecuta el valor de la ganancia como  $g = g + shigh - sl[i]$  con  $slow = sl[i]$ .

Y finalmente si estando en el estado 2 ocurre  $a$ , cambiamos la posición de negociación, es decir entramos por debajo, sin salirnos en ningún momento del mercado. El valor de la ganancia en este punto es  $g = g + sh[i] - slow$  donde  $shigh = sh[i]$ .

Por último, cuando  $i = 252$  se calcula la ganancia neta como:

$$ganancia = 100g.$$

## Capítulo 3

# Generalización del modelo inversión por pares

La idea que se propone en este capítulo, como su nombre lo indica, es generalizar el modelo de inversión por pares. Tomar sólo dos instrumentos parece un modelo restrictivo, sería interesante conformar portafolios sin límites en el número de instrumentos que lo constituyen. Es por esto que proponemos tomar no sólo dos instrumentos, sino un número  $n$  arbitrario, para luego seguir con las ideas de cointegración vistas en el Capítulo 2 y establecer las estrategias de inversión.

### 3.1. Identificación de clanes de activos.

Todas las pruebas de nuestros métodos se han hecho con el mercado de divisas, también llamado mercado Forex, sin embargo se puede utilizar cualquier otro mercado. Utilizamos 122 instrumentos que son los que distribuye la compañía Londinense *Oanda*. No sólo se tienen intercambio de monedas sino también metales preciosos y algunas materias primas. Nuestro objetivo principal es encontrar conjuntos de instrumentos, a los que en adelante llamaremos *clanes*, de tamaño arbitrario  $n$  tal que ellos cumplan en principio la propiedad de ser cointegrantes.

Se tendería a pensar que para hacer una búsqueda exhaustiva de todos los clanes cointegrantes, es posible generar un algoritmo que encuentre  $\binom{122}{n}$  clanes para luego hacer las pruebas de cointegración para cada uno de ellos. Aunque esto parece lógico, no es práctico ya que aún para valores de  $n$  pequeños se requeriría un tiempo de ejecución demasiado grande.

Es por esto que necesitamos establecer un método de filtrado para eliminar másivamente clanes con poca posibilidad de ser cointegrantes. Recordemos que en inversión por pares la forma de generar posibles pares cointegrates era observando que tan alineados estaban sus factores de exposición; intentaremos ir por esa dirección para encontrar clanes cointegrantes viables.

Si queremos clanes de tamaño  $n$  con  $n \geq 3$ , no es factible hablar de alineación entre factores de exposición de los instrumentos que conforman el clan en términos de la distancia definida en la Ecuación (2.10). Para esto hacemos uso del proceso de ortogonalización de Gram-Schmidt que permite construir de una lista arbitraria de vectores  $a_1, \dots, a_m$  una lista ortogonal  $b_1, \dots, b_m$  que genere al mismo subespacio.

Vamos a recordar un poco el proceso de Gram-Schmidt para comprender mejor la idea que queremos transmitir.

**Teorema 3.1** (Proyección ortogonal de un vector sobre el subespacio generado por una lista ortogonal). Sean  $V$  un espacio vectorial real o complejo con producto interno,  $b_1, \dots, b_j$  algunos vectores ortogonales no nulos,  $v \in V$  y  $l(b_1, \dots, b_j)$  el espacio generado por  $b_1, \dots, b_j$ . Definimos los vectores  $u, w \in V$  de la siguiente manera:

$$u = \sum_{k=1}^m \frac{\langle b_k, v \rangle}{\langle b_k, b_k \rangle} b_k, \quad w = v - u \quad (3.1)$$

Entonces  $w \perp l(b_1, \dots, b_j)$ .

**Proceso de ortogonalización de Gram-Schmidt.** Sea  $V$  un espacio vectorial real o complejo con producto interno y sean  $a_1, \dots, a_m \in V$ . Se quiere construir vectores ortogonales  $b_1, \dots, b_m \in V$  de tal manera que para todo  $j \in \{1, \dots, m\}$ , el espacio generado por los vectores  $b_1, \dots, b_j$  sea igual al espacio generado por

los vectores  $a_1, \dots, a_j$ , es decir,

$$l(b_1, \dots, b_j) = l(a_1, \dots, a_j).$$

La idea del proceso consiste en que en el  $j$ -ésimo paso se define el vector  $b_j$  como  $a_j$  menos la proyección ortogonal del vector  $a_j$  al subespacio generado por los vectores  $b_1, \dots, b_{j-1}$ . Así, en el  $j$ -ésimo paso, suponemos que los vectores  $b_1, \dots, b_{j-1}$  ya están construidos y son ortogonales entre si. Buscamos  $b_j$  de la forma

$$b_j = a_j - \sum_{k=1}^{j-1} \lambda_{j,k} b_k,$$

donde

$$\lambda_{j,k} := \begin{cases} \frac{\langle b_k, a_j \rangle}{\|b_k\|^2} & , b_k \neq 0 \\ 0 & , b_k = 0 \end{cases}$$

Obteniendo así que para todo  $j \in 1, \dots, m$  los vectores  $a_1, \dots, a_j$  generan al mismo subespacio que los vectores  $b_1, \dots, b_j$ . Ahora, daremos un par de corolarios que son clave para nuestro ejercicio de aplicación,

**Corolario 3.1** (Ortogonalización de Gram–Schmidt y dependencias lineales). *Sean  $\{a_1, \dots, a_m\}$  un conjunto de vectores en un espacio vectorial  $V$  y  $\{b_1, \dots, b_m\}$  el conjunto obtenido de  $\{a_1, \dots, a_m\}$  al aplicar el proceso de ortogonalización de Gram-Schmidt. Entonces para todo  $j \in \{1, \dots, m\}$  las siguientes condiciones son equivalentes:*

1.  $a_j \in l(a_1, \dots, a_{j-1})$ .
2.  $b_j = \mathbf{0}$ .

**Corolario 3.2.** *Partiendo de los mismos elementos introducidos en el enunciado del Corolario 3.1, las siguientes dos condiciones son equivalentes:*

1. Los vectores  $a_1, \dots, a_m$  son linealmente independientes.
2. Todos los vectores  $b_1, \dots, b_m$  son no nulos.

En muchos libros el proceso de ortogonalización de Gram–Schmidt se estudia solamente en el caso si los vectores originales  $a_1, \dots, a_m$  son linealmente independientes, pero preferimos trabajar en el caso general para tener presentes las consecuencias que esto tiene y porque se acopla más a lo que queremos hacer. Para aterrizar la idea a nuestro modelo, supongamos que tomamos el clan XAU\_NZD, CHF\_HKD y XAU\_SGD del mercado Forex, queremos ver si el clan es cointegrante, para esto tomamos las series de tiempo de los rendimientos para estos instrumentos en el año 2015, luego se calculan los factores de exposición de cada uno de los rendimientos, de tal modo que se obtienen tres vectores de factores de exposición, digamos  $a_1, a_2$  y  $a_3$ , ya con estos vectores se procede a efectuar el proceso de ortogonalización de Gram–Schmidt a  $a_1, a_2$  y  $a_3$ . Si  $b_1, b_2$  y  $b_3$  son los vectores que se obtienen después de aplicar el proceso de ortogonalización de Gram–Schmidt y se da que  $b_3 = 0$  entonces por Corolario 3.1,  $a_3$  es combinación lineal de  $a_1$  y  $a_2$ , es decir, hay una relación de equilibrio a largo plazo entre ellos, dando indicios de que el clan podría ser cointegrante, decimos “podría” porque realmente falta hacer pruebas exhaustivas de cointegración.

Bajo lo anterior, lo que hicimos fue generar un algoritmo que tome al azar  $n$  instrumentos de los 122 que están en el mercado Forex, formando un clan, a estos  $n$  vectores le aplicamos el proceso de ortogonalización de Gram–Schmidt, y nos cercioramos que al final del algoritmo  $b_n$  sea tal que  $\|b_n\| = 0$ , hay que tener en cuenta que es muy difícil en la práctica que  $\|b_n\| = 0$ , es por esto que seremos más laxos, es decir, sólo pediremos  $\|b_n\| \approx 0$ , para esto creamos el siguiente algoritmo:

---

**Algoritmo 3.1** Vector de Normas
 

---

```

1: procedure NORMVECTOR( $f$ )
2:    $n$  = El número de instrumentos que tiene  $f$ 
3:    $normVector$  = []
4:   for  $j \in \{1, \dots, n\}$  do
5:      $x$  = Vector de factores de exposición del  $j$ -ésimo instrumentos
6:      $normVector[j]$  =  $\|x\|$ 
7:   return  $NormVector$ 
8: end procedure

```

---

De esta forma generamos el vector  $normVector$  de tamaño 122 que tiene en cada entrada la norma de cada uno de los instrumentos o series de tiempo del

mercado Forex, esto fue para conocer en que rangos se mueve la norma de cada vector, en particular su media, de esto modo podría decirse que un vector de norma casi cero es una fracción muy pequeña de la media de `normVector`, por tanto un criterio para capturar un clan es que alguno de sus vectores en el proceso de Gram-Schmidt, digamos  $v$ , sea tal que, para  $x \in [0, 100]$  se cumpla:

$$v \leq \text{norma}(\text{normVector}) \frac{x}{100}.$$

### 3.2. Spread para clanes

Supongamos que con los métodos de la sección anterior se encontró un clan  $I_1, I_2, \dots, I_n$ , siguiendo las ideas de la Sección 2.2, vamos a aplicar una regresión multilínea a los instrumentos del clan para encontrar los coeficientes que los relacionan, es decir, se encontrará el vector de cointegración  $(\gamma_1, \gamma_2, \dots, \gamma_n)$ , donde  $\gamma_1 = 1$ , de tal forma que:

$$r_{I_1} = \gamma_2 r_{I_2} + \gamma_3 r_{I_3} + \dots + \gamma_n r_{I_n} + \mu,$$

donde cada  $r_{I_i} = \log(p_t^{I_i}) - \log(p_{t-1}^{I_i})$  para  $i \in \{1, 2, \dots, n\}$ , son los rendimientos de los instrumentos  $I_1, I_2, \dots, I_n$ , respectivamente, y  $\mu$  es el error en la regresión multilínea, cabe mencionar que  $\gamma_1 = 1$ , pues es el coeficiente correspondiente a la variable dependiente en la regresión.

Para estar en condiciones de aplicar el Algoritmo 2.1, tenemos que dar una definición de spread que sea consistente con lo que vimos para cointegración por pares. Sean  $A$  y  $B$  dos portafolios definidos de la siguiente manera:

$$A := \sum_{\gamma_i > 0} \gamma_i \log(p_t^{I_i}),$$

$$B := \sum_{\gamma_i < 0} \gamma_i \log(p_t^{I_i}),$$

siendo  $p_t^{I_i}$  el precio del instrumento  $I_i$  en el tiempo  $t$ . Note que en  $A$  se han concentrado los instrumentos con constante de regresión positiva, y en  $B$  los demás

instrumentos. El spread para el clan se define como:

$$spread := \frac{A+B}{\sum_{i=1}^n |\gamma_i|}. \quad (3.2)$$

Ya que si se quiere invertir un capital  $C$ , para preservar la proporción se debe invertir  $\frac{\gamma_i C}{\sum_{i=1}^n |\gamma_i|}$  en el instrumento  $I_i$ , siendo así, el rendimiento del portafolio para un periodo de tiempo, esta dado como:

$$\begin{aligned} r_t^{t+j} &= spread_{t+j} - spread_t \\ &= \frac{\sum_{i=1}^n \gamma_i \log(p_{t+j}^{I_i})}{|\gamma_1| + |\gamma_2| + \dots + |\gamma_n|} - \frac{\sum_{i=1}^n \gamma_i \log(p_t^{I_i})}{|\gamma_1| + |\gamma_2| + \dots + |\gamma_n|} \\ &= \frac{\sum_{i=1}^n \gamma_i [\log(p_{t+j}^{I_i}) - \log(p_t^{I_i})]}{|\gamma_1| + |\gamma_2| + \dots + |\gamma_n|} \\ &= \sum_{i=1}^n \frac{\gamma_i}{|\gamma_1| + |\gamma_2| + \dots + |\gamma_n|} [\log(p_{t+j}^{I_i}) - \log(p_t^{I_i})] \\ &= \frac{\sum_{i=1}^n \frac{C \gamma_i}{|\gamma_1| + |\gamma_2| + \dots + |\gamma_n|} [\log(p_{t+j}^{I_i}) - \log(p_t^{I_i})]}{C} \\ &= \frac{\text{Ganancia total}}{\text{Inversión}} \end{aligned}$$

Habiendo definido el spread, se ejecuta la prueba de estacionariedad tal como aparece en la Sección 2.2 y las estrategias de inversión como en la Sección 2.3 solamente definiendo el spread alto y bajo de la siguiente manera. Sea  $ph_t^{I_i}$  el precio alto del instrumento  $I_i$  en el tiempo  $t$ , y  $pl_t^{I_i}$  el precio bajo del instrumento  $I_i$  en el tiempo  $t$ , definimos los portafolios altos como:

$$Ah := \sum_{\gamma_i > 0} \gamma_i \log(ph_t^{I_i})$$

$$Bh := \sum_{\gamma_i < 0} \gamma_i \log(ph_t^{I_i})$$

y los portafolios bajos:

$$Al := \sum_{\gamma_i > 0} \gamma_i \log(pl_t^{I_i})$$

$$Bl := \sum_{\gamma_i < 0} \gamma_i \log(pl_t^{I_i})$$

Ahora definimos el spread alto del par  $A-B$  como la máxima diferencia, es decir,

$$spread_h = \frac{Ah + Bl}{\sum_{i=1}^n |\gamma_i|} \quad (3.3)$$

y el spread bajo del par  $A-B$  como la mínima diferencia:

$$spread_l = \frac{Al + Bh}{\sum_{i=1}^n |\gamma_i|} \quad (3.4)$$

que son los análogos de las Ecuaciones (2.13) y (2.14), respectivamente. El resto del proceso continua tal cual como en la Sección 2.3.



## Capítulo 4

# Otras implementaciones y resultados

Este capítulo se divide en dos partes, la primera contiene dos métodos que intentan encontrar subclanes cointegrantes de algún clan dado. En la segunda se dan las pautas necesarias para definir un algoritmo que ajusta los parámetros  $A, B, C, D$  y  $E$  del Algoritmo 2.1 con el objetivo de maximizar la correlación entre una nueva medida de estacionariedad  $M_s$ , y la prueba de estacionariedad descrita en la Sección 2.2 del Capítulo 2. La medida de estacionariedad  $M_s$  aún esta en desarrollo, pero se ha probado determina de manera precisa la estacionariedad de procesos ARIMA y caminatas aleatorias prefabricadas. También en esta misma sección se introducen las pruebas de estacionariedad usuales como la de Dickey-Fuller aumentada y la de Phillips-Perron, que en adelante las abreviaremos por ADF y PP, para igual con el mismo algoritmo encontrar su máxima correlación con la prueba de estacionariedad descrita en el Algoritmo 2.1.

## 4.1. Búsqueda de subclanes cointegrantes

El objetivo inicial de esta tesis, es encontrar clanes cointegrantes de tamaño  $n \geq 2$ . Por ejemplo, para  $n = 5$  nuestros métodos determinan que el clan

$$\mathcal{A} = \{EUR\_ZAR, NZD\_CAD, EUR\_PLN, USD\_TRY, FR40\_EUR\}$$

es cointegrante. La pregunta inmediata es, ¿existirá un clan más pequeño contenido en  $\mathcal{A}$  que también cointegre? Esto sólo puede ocurrir en algunos casos, y pueden haber varias circunstancias para que suceda. Retomando nuestro ejemplo, consideremos  $\mathcal{A}' = \{EUR\_ZAR, EUR\_PLN, FR40\_EUR\} \subseteq \mathcal{A}$  que resulta ser de nuevo un clan cointegrante. Esto pudo deberse a que los coeficientes en la regresión multilínea para  $BCO\_USD$  y  $XAG\_CAD$  con los que se estableció la correlación de los instrumentos en  $\mathcal{A}$  resultaron demasiado pequeños con respecto a los demás, de tal forma que su aporte a la combinación lineal es insignificante. Así, para encontrar sub-clanes cointegrantes en un clan cointegrante dado aplicamos dos métodos distintos:

### 4.1.1. Subclan con mejor coeficiente de determinación $R^2$ .

Desarrollamos un algoritmo para encontrar sub-clanes cointegrantes que toma como entrada un clan cointegrante de tamaño arbitrario  $n$ ,  $\mathcal{L} = \{I_1, I_2, \dots, I_n\}$  y con rendimientos  $\mathcal{R} = \{r_1, r_2, \dots, r_n\}$ , (para  $i \in \{1, 2, \dots, n\}$ ,  $r_i$  es el rendimiento de  $I_i$ ). Al conjunto  $\mathcal{R}$  se le aplica una regresión multilínea con el rendimiento  $r_i$  como dependiente, además se rescata de la regresión el valor  $R^2$ , es decir,

$$\alpha_{11}r_1 = \alpha_{12}r_2 + \alpha_{13}r_3 + \dots + \alpha_{1n}r_n + \varepsilon_1, \quad (4.1)$$

donde  $\alpha_{11} = 1$  y se guarda el valor  $R_{11}^2$ .

Permutar el orden de los términos independientes en la Ecuación (4.1) no cambia en absoluto el valor de  $R_{11}^2$ , pero si lo cambia permutar el término dependiente por alguno de los independientes, por lo tanto, el próximo paso es calcular

la regresión multilínea:

$$\alpha_{22}r_2 = \alpha_{21}r_1 + \alpha_{23}r_3 + \cdots + \alpha_{2n}r_n + \varepsilon_2, \quad (4.2)$$

donde  $\alpha_{22} = 1$  y se guarda el valor  $R_{22}^2$ .

Se sigue así sucesivamente hasta encontrar

$$\alpha_{nn}r_n = \alpha_{n1}r_1 + \alpha_{n2}r_2 + \cdots + \alpha_{n(n-1)}r_{n-1} + \varepsilon_n, \quad (4.3)$$

donde  $\alpha_{nn} = 1$  y se guarda el valor  $R_{nn}^2$ .

Después de realizar estas  $n$  regresiones, se analiza cuál de las  $n$  combinaciones entregó el mayor  $R^2$ , y cuál el menor  $R^2$ . Se realiza una gráfica en el plano euclidiano donde la abscisa representa el tamaño del clan, y la ordenada representa el valor del coeficiente de determinación mayor en todas las  $n$  combinaciones. Luego en la combinación donde este el menor  $R^2$ , se fija cual es la variable dependiente y la eliminamos de  $\mathcal{R}$ , y se vuelve a repetir el procedimiento con el nuevo  $\mathcal{R}$  de tamaño  $n - 1$ , hasta que el tamaño de  $\mathcal{R}$  sea 2.

Pretendíamos que de la gráfica final obtuviéramos el tamaño del sub-clan con mejor  $R^2$ , para tomar el coeficiente de determinación como filtro en el proceso de selección. Pero para todo  $n \in \{2, 3, \dots, 122\}$  obtuvimos gráficas como las siguientes:

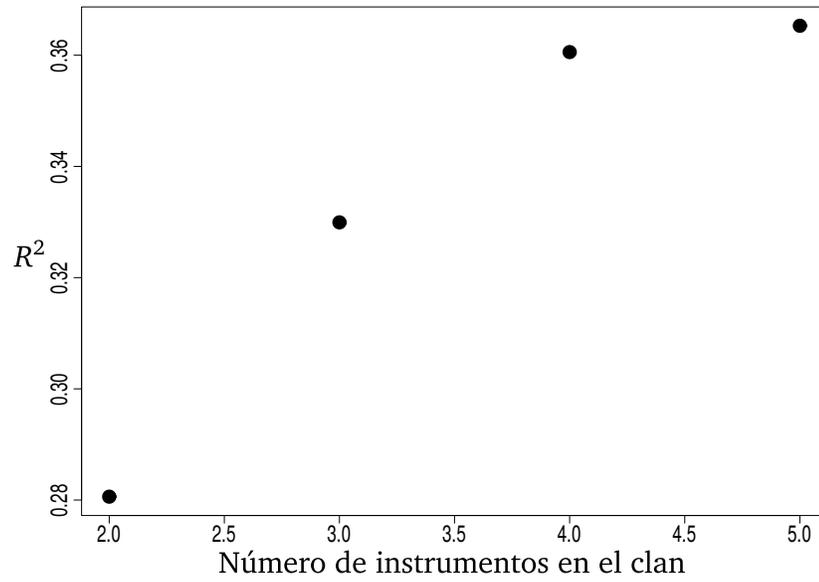


Figura 4.1: Clanes de tamaño menor o igual a 5.

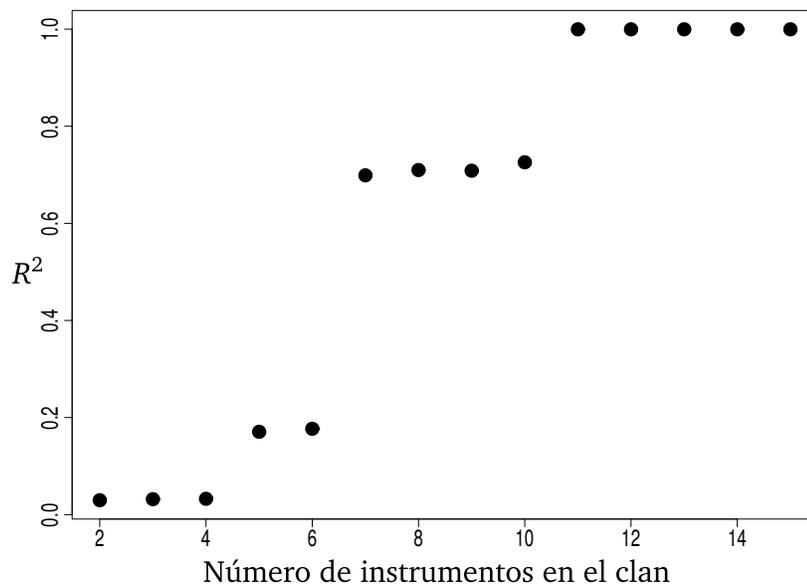


Figura 4.2: Clanes de tamaño menor o igual a 15.

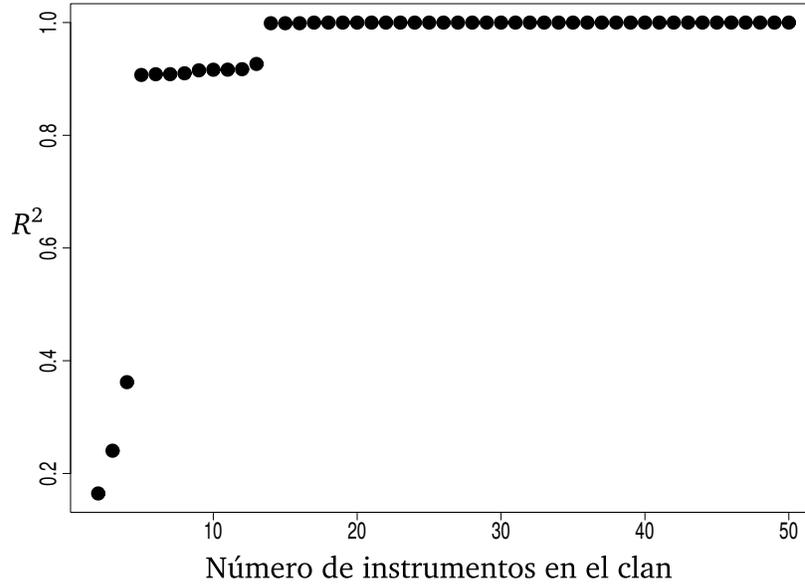


Figura 4.3: Clanes de tamaño menor o igual a 50.

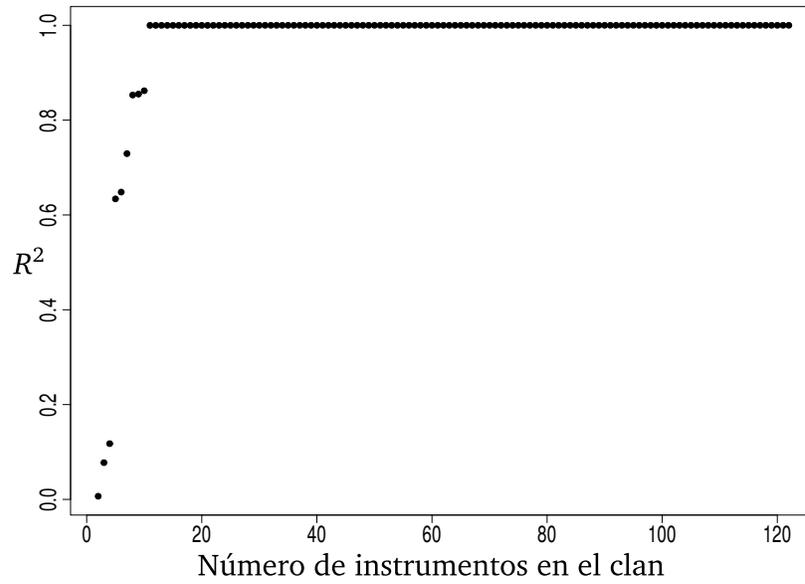


Figura 4.4: Clanes de tamaño menor o igual a 122.

Lo cual no sirve para buscar el tamaño del sub-clan óptimo. Así, el coeficiente de determinación  $R^2$  no es un buen filtro para este propósito, de hecho ahondando un poco más en la literatura, se tiene que  $R_{j-1}^2 \leq R_j^2$  para todo  $j \geq 2$ , donde  $R_j^2$  representa el coeficiente de determinación de una regresión multilínea de  $j$  variables; y esto fue constatado con nuestro algoritmo.

#### 4.1.2. Búsqueda exhaustiva de subclanes

Aquí no sólo se resuelve la pregunta que se planteó al inicio de la sección, sino que vamos un poco más allá, supongamos que tenemos el clan

$$\mathcal{C} = \{GBP\_CHF, CHF\_HKD, AUD\_CHF, BCO\_USD, XAG\_CAD\}$$

donde el orden de sus instrumentos es muy importante, en este caso con nuestros métodos se tiene que  $\mathcal{C}$  no es cointegrante. Estamos interesados en conocer si existe alguna combinación entre los instrumentos de  $\mathcal{C}$  que sea un subclan cointegrante. Habiendo definido el problema a solucionar, lo que se hace es una búsqueda exhaustiva de todos los subconjuntos de  $\mathcal{C}$ , para luego examinarlos con el Algoritmo 2.1, y así conocer cuales sub-clanes sobreviven a la prueba de estacionariedad.

Para esto se desarrolló el Algoritmo 4.1. A grandes rasgos hace lo siguiente: partiendo de un clan (*clan* en el algoritmo) dado de tamaño arbitrario  $n$ , se encuentran todas las  $\binom{n}{i}$  combinaciones para  $i \in \{2, 3, \dots, n\}$ . Luego, a cada combinación (que son subclanes de tamaño  $i$ ), se le permuta el primer instrumento con el  $j$ -ésimo instrumento, donde  $j \in \{1, 2, \dots, i\}$ , a este subclan de *clan* en el algoritmo lo denotamos por  $x$ . Posteriormente, se encuentran los rendimientos de  $x$  y se le realiza la regresión multilínea a estos, la regresión nos entrega los coeficientes de las variables (rendimientos) independientes, que son los que se usaran para definir el spread. Note que en la línea 8 del Algoritmo 4.1, en cada combinación del conjunto *clan*, sólo se permuta la variable dependiente con cada una de las variables independientes, pero no se permutan las variables independientes entre sí, esto es porque como está definido el spread en la Ecuación 3.2, el orden de las variables independientes no altera su valor, e igualmente al hacer

las regresiones multilineales de los rendimientos tampoco se altera el valor de los coeficientes de regresión. Una vez calculado el spread, se calculan el spread suavizado y el spread ajustado, que son las variables que requiere el Algoritmo 2.1 para ejecutar la prueba de estacionariedad. Finalmente si el subclan  $x$  logra sobrevivir al Algoritmo 2.1, se crea una tabla donde se imprime el subclan cointegrante  $x$  con los respectivos coeficientes en la regresión multilineal de sus rendimientos.

---

**Algoritmo 4.1** Búsqueda de subclanes cointegrantes de un clan dado.

---

```

1: procedure SUBCLANES(clan)
2:    $n :=$  Numero de instrumentos en clan
3:    $T :=$  Tabla para guardar los subclanes de clan cointegrantes.
4:   for  $i \in \{2, \dots, n\}$  do
5:      $X :=$  Combinaciones de tamaño  $i$  de los instrumentos de clan.
6:     for  $y \in X$  do
7:       for  $j \in \{1, \dots, i\}$  do
8:          $x :=$  Clan que se obtiene de permutar el instrumento 1
           con el  $j$ -ésimo instrumentos de  $y$ 
9:          $c :=$  Coeficientes de las variables independientes
           en la regresión multilineal de los
           rendimientos de  $x$ 
10:         $s :=$  Spread de  $x$  (Usando la Ecuación 3.2)
11:         $s_s :=$  Spread suavizado (tal cual como en la Sección 2.2)
12:         $s_{adj} :=$  Spread ajustado (tal cual como en la Sección 2.2)
13:         $nst :=$  STATIONARITYTEST( $s, s_s, s_{adj}$ )
14:        if  $nst = TRUE$  then
15:          Se le inserta a  $T$  el clan  $x$  y los coeficientes  $c$ 
16:        return  $T$ 
17: end procedure

```

---

Para ejemplificar mejor la situación, vamos a generar todos los subclanes cointegrantes de  $\mathcal{C}$ , los cuales presentamos en las siguientes tablas, para mayor comprensión de ellas, se denota  $r_i$  como el rendimiento del  $i$ -ésimo instrumento en el subclan, y  $c_i$  denota el coeficiente de la regresión multilineal del rendimiento  $r_i$ , claro está que el coeficiente  $r_1 = 1$  para todo subclan, pues es la variable dependiente en la regresión.

$r_1$	$r_2$	$c_2$
<i>XAG_CAD</i>	<i>GBP_CHF</i>	-0.06285
<i>CHF_HKD</i>	<i>BCO_USD</i>	-0.02911
<i>XAG_CAD</i>	<i>CHF_HKD</i>	0.08406
<i>AUD_CHF</i>	<i>BCO_USD</i>	0.11540
<i>AUD_CHF</i>	<i>XAG_CAD</i>	-0.01591
<i>XAG_CAD</i>	<i>AUD_CHF</i>	-0.01672
<i>XAG_CAD</i>	<i>BCO_USD</i>	0.03312

Tabla 4.1: Subclanes cointegrantes de  $\mathcal{C}$  de tamaño dos.

$r_1$	$r_2$	$r_3$	$c_2$	$c_3$
<i>XAG_CAD</i>	<i>CHF_HKD</i>	<i>GBP_CHF</i>	0.25746	0.21520
<i>AUD_CHF</i>	<i>GBP_CHF</i>	<i>BCO_USD</i>	0.89777	0.06043
<i>AUD_CHF</i>	<i>GBP_CHF</i>	<i>XAG_CAD</i>	0.91308	0.03487
<i>XAG_CAD</i>	<i>AUD_CHF</i>	<i>GBP_CHF</i>	0.16017	-0.20875
<i>XAG_CAD</i>	<i>BCO_USD</i>	<i>GBP_CHF</i>	0.03747	-0.07099
<i>AUD_CHF</i>	<i>CHF_HKD</i>	<i>BCO_USD</i>	-0.76730	0.09307
<i>XAG_CAD</i>	<i>AUD_CHF</i>	<i>CHF_HKD</i>	0.25562	0.28202
<i>XAG_CAD</i>	<i>BCO_USD</i>	<i>CHF_HKD</i>	0.03565	0.08680
<i>AUD_CHF</i>	<i>BCO_USD</i>	<i>XAG_CAD</i>	0.11633	-0.02799
<i>XAG_CAD</i>	<i>BCO_USD</i>	<i>AUD_CHF</i>	0.03666	-0.03068

Tabla 4.2: Subclanes cointegrantes de  $\mathcal{C}$  de tamaño tres.

$r_1$	$r_2$	$r_3$	$r_4$	$c_2$	$c_3$	$c_4$
<i>XAG_CAD</i>	<i>CHF_HKD</i>	<i>AUD_CHF</i>	<i>GBP_CHF</i>	0.33024	0.23327	0.08132
<i>XAG_CAD</i>	<i>CHF_HKD</i>	<i>BCO_USD</i>	<i>GBP_CHF</i>	0.23586	0.02862	0.18566
<i>AUD_CHF</i>	<i>GBP_CHF</i>	<i>BCO_USD</i>	<i>XAG_CAD</i>	0.89976	0.05938	0.02796
<i>XAG_CAD</i>	<i>AUD_CHF</i>	<i>BCO_USD</i>	<i>GBP_CHF</i>	0.13490	0.02932	-0.19210
<i>XAG_CAD</i>	<i>AUD_CHF</i>	<i>BCO_USD</i>	<i>CHF_HKD</i>	0.23938	0.01337	0.27048

Tabla 4.3: Subclanes cointegrantes de  $\mathcal{C}$  de tamaño cuatro.

Finalmente se obtuvo con el Algoritmo 4.1, que la permutación

$$\mathcal{C}' = \{XAG\_CAD, CHF\_HKD, AUD\_CHF, BCO\_USD, GBP\_CHF\}$$

es un subclan cointegrante de  $\mathcal{C}$ , note que el orden es muy importante, pues

aunque  $\mathcal{C}$  no es cointegrante,  $\mathcal{C}'$  si lo es. Los coeficientes de cointegración para los rendimientos de  $\mathcal{C}'$  son 0.31670, 0.21966, 0.01231 y 0.07642 para cada una de las respectivas variables independientes.

A modo de conclusión, se puede observar en las Tablas 4.2 y 4.3 que algunas constantes de cointegración son muy cercanas a cero. Podría interpretarse que, a menos que se invierta mucho dinero, los portafolios casi no van a tener los instrumentos asociados, esto es lo que se supone, de todos modos hay que pensar un poco en las consecuencias que este hecho conlleva.

## 4.2. Correlaciones entre pruebas de estacionariedad

Como ya mencionamos al inicio del capítulo, vamos a realizar un contraste entre varias pruebas de estacionariedad ya existentes y la prueba de estacionariedad descrita en el Algoritmo 2.1. Mediante un algoritmo genético se ajustaran los parámetros  $A, B, C, D$  y  $E$  del Algoritmo 2.1, para que haya una mejor correlación con la prueba de estacionariedad de Dickey-Fuller y la prueba de Phillips-Perron. Finalmente, se encuentra la mejor correlación con la medida de estacionariedad  $M_s$ . Para entender mejor lo que vamos hacer, primer daremos un breve recorrido por los conceptos de índice de correlación, las pruebas de estacionariedad  $ADF$ ,  $PP$  y  $M_s$ .

### 4.2.1. Índice de Correlación

La correlación estadística determina la relación o dependencia que existe entre dos variables que intervienen en una distribución bidimensional. Es decir, determinar si los cambios en una de las variables influyen en los cambios de la otra. En caso de que suceda, diremos que las variables están correlacionadas o que hay correlación entre ellas.

En caso de que se estén estudiando dos variables aleatorias  $X$  y  $Y$  sobre una

población; la correlación se simboliza con la letra  $r$ , y su expresión esta dada por:

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

donde  $\sigma_{XY}$  es la covarianza de  $(X, Y)$ ,  $\sigma_X$  y  $\sigma_Y$  son las desviaciones estándar de  $X$  y  $Y$ , respectivamente.

El valor del índice de correlación  $r$  varía en el intervalo  $[-1, 1]$ :

- Si  $r = 1$ , existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si  $0 < r < 1$ , existe una correlación positiva, entre más cerca este  $r$  de uno hay una fuerte relación lineal entre las variables  $X$  y  $Y$ ; por el contrario si  $r$  esta muy próximo a cero, indica que las variables no están muy relacionadas.
- Si  $r = 0$ , no existe relación lineal.
- Si  $-1 < r < 0$ , existe una correlación negativa.
- Si  $r = -1$ , existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante.

#### 4.2.2. Pruebas de estacionariedad ADF y PP

ADF y PP son ambas pruebas de raíces unitarias, con la diferencia que PP no es paramétrica. Estas pruebas se usan para determinar si una serie de tiempo es estacionaria o no, en esta tesis no se ahondará en la teoría de ADF y PP, los detalles pueden consultarse en [MK98]. Para todos los algoritmos e implementaciones que se han realizado, hemos utilizado el lenguaje de programación R, el cual es un entorno y lenguaje de programación con un enfoque al análisis estadístico, que proporciona muchas herramientas para el análisis de series financieras. En particular, R tiene el paquete *tSeries*, el cual ofrece las pruebas ADF y PP de

estacionariedad para series de tiempo basadas en pruebas de hipótesis. Para la prueba de estacionariedad ADF, `tSeries` tiene la función `adf.test`, y para la prueba de estacionariedad PP, `tSeries` tiene la función `pp.test`, donde la hipótesis nula y alternativa son, respectivamente:

$$\begin{aligned} H_0 &= \text{La serie es no estacionaria,} \\ H_a &= \text{La serie es estacionaria.} \end{aligned} \tag{4.4}$$

Tanto la función `adf.test`, como `pp.test`, dada una serie de tiempo arbitraria, proporcionan su respectivo *p-valor*. El *p-valor* es una probabilidad, por lo que oscila entre 0 y 1. Se suele decir que valores altos de *p* no rechazan la hipótesis nula o, dicho de forma correcta, no permiten rechazar la  $H_0$ . De igual manera, valores bajos de *p* rechazan la  $H_0$ . Es importante recalcar que el valor de *p* no permite aceptar una hipótesis; simplemente la rechaza o no la rechaza, es decir que la tacha de verosímil (lo que no significa obligatoriamente que sea cierta, simplemente que es más probable de serlo) o no. A modo de ejemplo, si tenemos la serie de tiempo  $s$ , y le aplicamos la prueba ADF, es decir, `adf.test(s)` y esta función nos entrega como *p-valor* el número 0.2, esto quiere decir que la serie  $s$  tiene una probabilidad del 20% de ser no estacionaria.

### 4.2.3. Medida de estacionariedad $M_s$

Como ya vimos, el reconocimiento de cointegración de series de tiempo puede realizarse estableciendo la estacionariedad del spread, construido simplemente como la regresión lineal de las series de tiempo participantes. Hay varias pruebas estadísticas para investigar esto. Tradicionalmente, en inversión por pares se utilizan las pruebas ADF y PP, pero sólo funcionan bien cuando los spreads originales son caminatas aleatorias. Al menos, con series de tiempo financieras no es una suposición razonable.

Se desarrolla una función denominada medida de estacionariedad de un spread  $s$ , la cual denotaremos por  $M_s := M(s)$ . Donde se hace una prueba extensa sobre caminatas aleatorias  $w(\alpha)$  de la forma  $w_n = \alpha \times w_{n-1} + \epsilon_n$ ,  $w_0 = \epsilon_0$  donde  $\alpha$  es cierto número positivo real y  $\epsilon_n$  es un ruido gaussiano aleatorio. Es muy conocido

que este tipo de caminatas aleatorias son estacionarios si y sólo si  $|\alpha| < 1$ . Después de varios intentos, la medida de estacionariedad se desarrolló de tal manera que el promedio  $\hat{M}(\alpha)$  de  $M_w(\alpha)$  sobre una gran cantidad de procesos  $w(\alpha)$  con un  $\alpha$  fijo cumple que  $\hat{M}(\alpha)$  es creciente en  $\alpha$  y  $\hat{M}(1) = 1$ . La idea es aplicar la medida de estacionariedad a un spread arbitrario  $s$ , deseando un resultado menor a 1 si y sólo si  $s$  es una serie de tiempo estacionaria.

La función de clasificación  $M_s$  funcionó bien para caminatas aleatorias cuando se toma el promedio de miles de realizaciones de una misma caminata, pero no cuando tomamos exactamente una realización de la misma. Por lo tanto, construimos varios miles de caminatas aleatorias y los usamos para determinar que si  $M_s$  es inferior a 0.35, entonces con una certeza del 95 % podríamos asegurar que la caminata es estacionaria. No sabemos realmente qué tanto se alejen los spreads de ser caminatas aleatorias pero consideramos que este umbral es adecuado para los propósitos de localizar clanes cointegrantes. Un cálculo más preciso del umbral debería ser tomado mediante un proceso de simulación que lo optimice en términos de las ganancias totales. Seguramente este umbral será dependiente del mercado que estemos considerando.

Ahora, veamos explícitamente como se define la medida de estacionariedad  $M_s$  para un spread específico  $s$ . Sea  $\{s_t\}$  un proceso estocástico que tiene como realización  $\{s_t\}$ . En lo que sigue se asumirá que  $\tau$  es fijo. Se define el siguiente par de series de tiempo a partir de la realización  $\{s_t\}$ .

$$e_t^\tau \equiv \frac{1}{\tau + 1} \sum_{i=1}^{t+\tau} s_i, \quad (4.5)$$

$$\sigma_t^\tau \equiv \sqrt{\frac{1}{\tau + 1} \sum_{i=1}^{t+\tau} (s_i - e_t^\tau)^2}, \quad (4.6)$$

de donde,

$$M_s = \|s_t\|^2 = \frac{Cn^2}{(n - \tau)\sigma^2} [\text{var}(e_t^\tau) + \text{var}(\sigma_t^\tau)], \quad (4.7)$$

el término  $\text{var}(e_t^\tau)$  mide la varianza de los promedios en ventanas de tamaño  $\tau$  de los valores de la serie. En una serie estacionaria se espera que la varianza sea muy pequeña. El segundo término, es decir,  $\text{var}(\sigma_t^\tau)$  corresponde a la variación del

promedio de la desviación estándar. La constante  $\frac{n^2}{(n-\tau)\sigma^2}$  sirve para obtener una medida que no dependa ni de  $n$  ni de  $\tau$  y finalmente el valor de  $C$  es una constante que vale 1,14758597, para que la caminata aleatoria tenga  $E(\|x_t\|) = 1$ .

Los detalles de la medida de estacionariedad  $M_s$  se han omitido, porque esta medida es el resultado de la tesis de doctorado de Andrés Felipe Tellez Crespo. La medida  $M_s$  aún esta en desarrollo y se tuvo conocimiento de su existencia una vez que ya teníamos implementado el Algoritmo 2.1, sin embargo resultó útil para afinar los parámetros  $A, B, C, D$  y  $E$  de la prueba de estacionariedad descrita en la Sección 2.2, como veremos a continuación.

#### 4.2.4. Optimización de la correlación

Ya habiendo definido el marco teórico necesario para entender cómo vamos a ajustar los parámetros  $A, B, C, D$  y  $E$  del Algoritmo 2.1, procedemos a explicar cómo logramos valores óptimos de estos parámetros para que la prueba 2.1 tenga una mayor correlación con ADE, PP y por último con  $M_s$ .

En principio se define de qué tamaño queremos el clan, digamos  $n$ , por tanto de los 122 instrumentos que conforman el mercado con el que estamos trabajando, elegimos  $n$  instrumentos al azar, sin importar si cointegran o no, tampoco nos va a importar tener una buena ganancia; esta va ser la forma de tener una muestra aleatoria de clanes. Para que los resultados que se obtengan sean más verídicos, se aconseja tomar una muestra lo suficientemente grande, en nuestro caso elegimos una muestra de 75,000 clanes. Luego de esto a cada clan se le encuentra su spread como ya aprendimos en capítulos anteriores. Cada spread como ya sabemos es en sí una serie de tiempo, digamos  $s$ , de la cual buscamos una serie de información que se va guardando en la Tabla 4.4 como mostramos a continuación:

Columna	Información
$C_1$	Instrumento 1
$C_2$	Instrumento 2
$\vdots$	$\vdots$
$C_n$	Instrumento $n$
$C_{n+1}$	Coficiente 2
$C_{n+2}$	Coficiente 3
$\vdots$	$\vdots$
$C_{2n-1}$	Coficiente $n$
$C_{2n}$	lengthSpread
$C_{2n+1}$	diffSpread
$C_{2n+2}$	rata
$C_{2n+3}$	lapso
$C_{2n+4}$	booting
$C_{2n+5}$	$M_s$
$C_{2n+6}$	El p-valor de $adf.test(s)$
$C_{2n+7}$	El p-valor de $pp.test(s)$

Tabla 4.4: Muestra para optimización

La Tabla 4.4 realmente es una matriz de 75,000 filas y  $(2n+7)$  columnas, donde cada fila contiene la información de un clan específico. En las columnas  $C_{2n}$  a  $C_{2n+4}$  se encuentran los valores de las variables *lengthSpread*, *diffSpread*, *rata*, *lapso* y *booting* para cada spread, la forma de calcular estos valores es tal como se indica en el Algoritmo 2.1. Considérese la  $i$ -ésima fila de la Tabla 4.4, además de tener los valores  $C_{2n+5}[i]$ ,  $C_{2n+6}[i]$  y  $C_{2n+7}[i]$ ; se puede gracias a ella hacer rápidamente la prueba de cointegración descrita en la Sección 2.2 bajo los parámetros  $A, B, C, D$ , y  $E$  de la siguiente manera:

$$\begin{aligned}
 NST[i] = 1 \text{ sí y sólo sí } & C_{2n}[i] \geq A \wedge C_{2n+1}[i] \leq B \wedge \\
 & C_{2n+2}[i] \geq C \wedge C_{2n+3}[i] \leq D \wedge \\
 & C_{2n+4}[i] \leq E,
 \end{aligned}$$

donde el rango de  $NST$  es  $\{0, 1\}$ . La prueba de estacionariedad descrita en el Algoritmo 2.1 y  $NST$  son exactamente lo mismo, solo que  $NST$  es la forma abreviada, por tanto si el spread de un clan específico es aceptado por el Algoritmo

2.1, entonces  $NST = 1$ . Tenga en cuenta que vamos a estar buscando los mejores parámetros  $A, B, C, D$  y  $E$  en  $NST$  para que exista una mejor correlación entre esta prueba y las demás pruebas usuales de estacionariedad.

Para hacer la optimización de la correlación mediante el ajuste de los parámetros ya mencionados vamos a usar un algoritmo genético, donde la función de adaptación se muestra en el Algoritmo 4.2. Si se quiere encontrar la mejor correlación entre  $NST$  y  $M_s$ ;  $X$  debe tomar el valor de  $C_{2n+5}$ , o si queremos hallar la mejor correlación entre  $NST$  y ADF,  $X$  debe ser reemplazado por  $C_{2n+6}$ ; y por último, si queremos ver la mejor correlación entre  $NST$  y PP, simplemente  $X$  toma el valor de  $C_{2n+7}$ . La variable  $\alpha$  en la función  $CORRELATION(A, B, C, D, E, X, \alpha)$ , en el caso de  $M_s$  es el mayor valor que le permitiremos a esta variable para concluir estacionariedad, y en el caso de ADF y PP, es el máximo p-valor que aceptaremos. Una vez establecida la función de adaptación, se define el dominio de los parámetros  $A, B, C, D$ , y  $E$  del Algoritmo 4.2, es decir,

$$(A, B, C, D, E) \in [a_1, a_2] \times [b_1, b_2] \times [c_1, c_2] \times [d_1, d_2] \times [e_1, e_2], \quad (4.8)$$

donde  $a_i, b_i, c_i, d_i, e_i \in \mathbf{R}$ , con  $i = 1, 2$ .

---

**Algoritmo 4.2** Función de adaptación
 

---

```

1: procedure CORRELATION( $A, B, C, D, E, X, \alpha$ )
2:    $Y :=$  Es un vector de 0's y 1's, tal que  $Y[i] = 1$  sii  $X[i] < \alpha$ 
3:    $NST :=$  Es un vector de 0's y 1's, tal que  $NST[i] = 1$  sii  $(C_{2n}[i] \geq A) \wedge$ 
       $(C_{2n+1}[i] \leq B) \wedge (C_{2n+2}[i] \geq C) \wedge (C_{2n+3}[i] \leq D) \wedge$ 
       $(C_{2n+4}[i] \leq E)$ 
4:   return  $|cor(Y, NST)|$ 
5: end procedure

```

---

Ya definidos los intervalos, en principio se elige al azar una población de tamaño arbitrario  $m$ , por ejemplo si  $m = 6$ , tendríamos seis quintuplas al azar de 4.8. Digamos que cada elemento de la población inicial lo denotamos como  $\mathcal{P}_j$  para  $j \in \{1, 2, \dots, m\}$ , donde  $|\mathcal{P}_j| = 5$ . Para cada  $j$ , con los parámetros que hay en  $\mathcal{P}_j$  se ejecuta la función de cointegración  $NST$ , resultando un vector de 0's y 1's, de dimensión 75,000; luego se encuentra la correlación entre  $Y$  y  $NST$  como en el Algoritmo 4.2. De aquí la población sucesora o nueva generación, se obtiene

de la población actual (padres) por medio de cruce y mutación entre los que tuvieron mejor correlación. El proceso se sigue repitiendo sucesivamente hasta que la correlación permanezca inmóvil por 100 generaciones o que ya hayan pasado 1000 generaciones con continuos cambios en la correlación.

Para obtener resultados reales de nuestra muestra, nos apoyamos en el paquete *Genetic Algorithms* o GA de R, que se encarga por si solo de los procesos de selección al azar de los parámetros  $A, B, C, D$  y  $E$  de la función de adaptación y de crear todas las generaciones hasta obtener los parámetros adecuados que maximicen la correlación entre  $NST$  y  $Y$ . A la función GA se le debe entregar la función de adaptación  $CORRELATION(A, B, C, D, E, X, \alpha)$ , donde tomamos el tamaño de la población como  $m = 5$  para los siguientes ejemplos.

**Ejemplo 4.1.** Si estamos interesados en encontrar la máxima correlación entre el Algoritmo 2.1 y  $M_s$ , entonces a GA debemos decirle que  $X = C_{2n+5}$  y supongamos que a los parámetros que queremos ajustar les damos la siguiente libertad de movimiento:  $A \in [70, 200]$ ,  $B \in [0.01, 0.1]$ ,  $C \in [0.001, 0.1]$ ,  $D \in [5, 70]$ ,  $E \in [20, 40]$  y por último sea  $\alpha = 0.34$ , siendo así se tiene que  $\max(\text{cor}(M_s, NST)) = 0.4413323$ , donde los parámetros adecuados para que esto se de son:

$$(A, B, C, D, E) = (133.2401, 0.04612005, 0.05329019, 38.93517, 32.9424)$$

Por otro lado, si se desea ajustar también el valor  $\alpha$ , digamos  $\alpha \in [0, 0.4]$ , entonces resulta que  $\max(\text{cor}(M_s, NST)) = 0.4494876$ , con  $\alpha = 0.3978093$  y

$$(A, B, C, D, E) = (128.4455, 0.04558207, 0.07801047, 36.2405, 35.74052)$$

en cualquiera de los dos casos, da una excelente correlación.

**Ejemplo 4.2.** Ahora si deseamos encontrar la máxima correlación entre el Algoritmo 2.1 y  $PP$ , entonces  $X$  en GA debe ser  $C_{2n+7}$ . Por otro lado, supongamos que  $\alpha = 0.2$ , es decir, un clan se acepta como cointegrante si su  $p$ -valor es menor que 0.2 (existe un porcentaje del 20% de no estacionariedad) y además sea  $(A, B, C, D, E) \in \mathcal{P}$ , con

$$\mathcal{P} = [125, 185] \times [0.03, 0.06] \times [0.03, 0.05] \times [30, 50] \times [25, 35]$$

Con esta información el algoritmo nos dice que la  $\max(\text{cor}(PP, NST)) = 0.1872722$

con  $(A, B, C, D, E) = (156.9009, 0.04498385, 0.04203171, 49.576633, 34.29693)$

Ahora supongamos que  $\alpha \in [0, 0.08]$ , es decir, sólo se aceptan un clan como cointegrantes si su  $p$ -valor es menor o igual a 0.08, veamos hasta que punto se mejora la correlación y que valor debe tomar  $\alpha$  para que esto se de. En este caso, se tiene que  $\max(\text{cor}(PP, NST)) = 0.1678163$  con  $\alpha = 0.01078138$  y

$$(A, B, C, D, E) = (147, 2438, 0.05695511, 0.04062347, 48.08664, 34.90715).$$

**Ejemplo 4.3.** Finalmente, si queremos encontrar la máxima correlación entre el Algoritmo 2.1 y ADF, entonces sea  $X = C_{2n+6}$ . Supongamos que  $\alpha = 0.05$  e igual como en el ejemplo anterior sea  $(A, B, C, D, E) \in \mathcal{P}$ , entonces, GA nos entrega el siguiente resultado:  $\max(\text{cor}(ADF, NST)) = 0.1714448$  con

$$(A, B, C, D, E) = (140.1805, 0.05925457, 0.03871177, 48.39294, 34.98984)$$

Si ajustamos también el parámetro  $\alpha$ , digamos  $\alpha \in [0, 0.1]$ , entonces la correlación se mejora un poco:  $\max(\text{cor}(ADF, NST)) = 0.1870343$  con un excelente valor de  $\alpha$  de 0.03724387 y los siguientes nuevos parámetros en el Algoritmo 2.1:  $(A, B, C, D, E) = (146.5592, 0.05984317, 0.03634506, 48.0684, 34.84543)$ .



## Apéndice A

### Calculo del umbral $\Delta$

#' Módulo para determinar el parámetro la desviación Delta a  
#' la media del spread que resulte óptima. En esta parte  
#' seguimos el método descrito en la Sección "Nonparametric  
#' approach" del libro de Ganapathy Vidyamurthy.

#' Suavizado de la función de ganancias a partir del umbral  
#' Delta

#'

#' Cálculo del valor del parámetro  $\lambda$  que aparece en  
#' la ecuación 8.1 del texto de referencia. Seguimos la  
#' técnica que se explica que se explica en la Sección  
#' "Regularization" del libro. Lo único adicional que se  
#' hizo fue hallar la derivada de la función de costo 8.1 y  
#' resolver el sistema que resulta de igualar todas las  
#' derivadas parciales a cero.

#'

#' @param y - Serie de tiempo con la función de  
#' probabilidad que el spread esté por encima de  
#' Delta. Aquí Delta es la variable  
#' independiente y se expresa como una fracción

```

#'           de la desviación estándar. Ver la Figura
#'           8.2A del texto
#' @param l   - Es el valor lambda que aparece en la función
#'           de costo (ec. 8.1)
#' @return    - Una lista. que contiene
#'           z           - Es el vector z que minimiza la
#'           función de costo 8.1
#'           cost        - Costo de la solución z
#'           costFit     - Costo asociado al primer término
#'           de 8.1 (Costo de ajuste).
#'           costSmooth - Costo asociado al segundo
#'           término de 8.1 (Suavidad del
#'           ajuste)
smoothSolution <- function(y, l) {
  n <- length(y)

  # Matriz asociada al sistema para localizar el óptimo
  a <- matrix(rep(0,n*n), nrow = n, ncol = n)
  a[1,1] <- 1+l
  a[1,2] <- -l
  a[n,n] <- 1+l
  a[n,n-1] <- -l
  for (i in 2:(n-1)) {
    a[i,i] <- 1+2*l
    a[i,i-1] <- -l
    a[i,i+1] <- -l
  }
  # Solución del sistema
  sol <- solve(a,y)

  # Cálculo de los costos
  costSmooth <- 0.0
  for (i in 1:(n-1))

```

---

```
    costSmooth <- costSmooth + 1*(sol[i] - sol[i+1])**2
costFit      <- sum((y - sol)**2)
return (list("z" = sol, "cost" = costFit+costSmooth,
            "costFit" = costFit,
            "costSmooth" = costSmooth))
}
require(graphics)
library(zoo)

#' Cálculo de Delta
#'
#' Calcula la distancia a la media óptima para comprar el
#' spread de dos pares cointegrantes.
#' Se utiliza el método descrito en la Sección
#' "Regularization" del libro de texto.
#'
#' @param spread      - Spread al que se le quiere
#'                    calcular el delta óptimo.
#' @param demoMode    - Si se requieren o no
#'                    explicaciones y gráficas del
#'                    proceso.
#' @param lambdaNBuckets - Número de divisiones del umbral
#'                    delta para suavizar la función de
#'                    costo. Es el parámetro n de la
#'                    ecuación 8.1 en el libro de texto.
#'                    Sigue una escala logar'itmica.
deltaComputation <- function(spread, demoMode,
                            lambdaNBuckets = 80) {
  # Cálculo de la desviación
  standardDeviation = sd(spread)
  if (demoMode)
    plot.zoo(as.zoo(spread), main=NA, ylab = "Spread")
```

```

# Conteo de frecuencias por encima de los umbrales
rango      <- max(abs(spread))/standardDeviation
stepDelta  <- rango/lambdaNBuckets
xScale     <- seq(0, rango, stepDelta)
rawCount   <- rep(0, lambdaNBuckets)
for (i in 1:length(spread))
  rawCount[round(lambdaNBuckets*abs(spread[i])/
                 (rango*standardDeviation))] <-
    rawCount[round(lambdaNBuckets*abs(spread[i])/
                 (rango*standardDeviation))] + 1
if (demoMode)
  plot.zoo(as.zoo(rawCount), main=NA, x=xScale, type="l",
          ylab = "Conteo de valores por encima del umbral",
          xlab = "Umbral lambda (como veces la desviación
                 estándar)")

# Probabilidades de caer por encima de cada umbral
for (i in (length(rawCount)-1):1) {
  rawCount[i]   <- rawCount[i]+rawCount[i+1]
  rawCount[i+1] <- rawCount[i+1]/(2*length(spread))
}
rawCount[1] <- rawCount[1]/(2*length(spread))
if (demoMode)
  plot.zoo(as.zoo(rawCount), main=NA, x=xScale, type="l",
          ylab = "Probabilidad",
          xlab = "Umbral lambda (como veces la desviación
                 estándar)")

# Conversión de la función de conteo a una forma
# estrictamente decreciente
for (i in 2:length(rawCount)) {
  if (rawCount[i] == rawCount[i-1]) {
    j = i

```

```
while (j < length(rawCount) &&
      rawCount[j] == rawCount[i-1])
  j <- j+1
deltax <- (rawCount[j] - rawCount[i-1])/(j-i+1)
for (k in i:(j-1))
  rawCount[k] <- rawCount[k-1] + deltax
}
}
if (demoMode)
  plot.zoo(as.zoo(rawCount), main=NA, x=xScale,
          type="l", ylab = "Probabilidad",
          xlab = "Umbral lambda (como veces la desviación
                estándar)")

# Cálculo de la función de ganancias
profit = (rawCount * xScale[1:length(xScale)-1]) *
  standardDeviation
if (demoMode)
  plot.zoo(as.zoo(profit), main=NA, x=xScale, type="l",
          ylab = "Ganancias",
          xlab = "Umbral lambda (como veces la desviación
                estándar)")

# Cálculo del valor óptimo de lambda en la ecuación 8.1 del
# texto Ganapathy Vidyamurthy
ldom <- seq(-4,12,0.25)
costl <- seq(-4,12,0.25)
for (i in 1:length(ldom)) {
  l <- exp(ldom[i])
  result <- smoothSolution(rawCount, l)
  costl[i] <- result$cost
}
if (demoMode)
```

```
plot(cost1, x=lom, type="l", ylab = "Costo",
      xlab="log(lambda)")

equilibriumCost <- (max(cost1)-min(cost1))/30
equilibriumEntry <- findInterval(equilibriumCost, cost1)
loptimal <- exp(lom[equilibriumEntry])
result <- smoothSolution(rawCount, loptimal)
if (demoMode)
  plot.zoo(as.zoo(result$z), main=NA)

revenue <- (result$z*xScale[1:length(xScale)-1]) *
  standardDeviation
if (demoMode)
  plot.zoo(as.zoo(revenue), main=NA, x=xScale, type="l",
          ylab = "Ganancias",
          xlab = "Umbral lambda (como veces la desviación
                estándar)")
maxUmbral = xScale[which.max(revenue)]
return (maxUmbral)
}
# Quitando el siguiente comentario se ejecuta el ejemplo de
# cálculo de delta que aparece en el texto
# cat("Delta = ", deltaComputation(rnorm(200, 0, 2), TRUE),
#     " veces la desviación estándar")
```

## Apéndice B

# Promedio móvil del spread

```
#' Función principal para suavizar
#' @param spread - Spread al que se le quiere calcular el
#'               suavizado.
suavizar = function (serie) {
  #' Función para ignorar los valores muy altos de AIC
  #' @param xx      - Seire de tiempo
  #' @param minOrden - Límite inferior de parámetros
  #' @param maxOrden - Límite superior de parámetros
  FiltroAIC = function (xx, minOrden=c(0,0),
                        maxOrden=c(5,5)){
    mejorAic = 1e9 #valor inicial de AIC
    len = NROW(xx)
    a=c()
    for (p in minOrden[1]:maxOrden[1])
      for (q in minOrden[2]:maxOrden[2]) {
        if (p == 0 && q == 0)
          next
        fit1 <- tryCatch(arima(xx, c(p, 0, q)),
                        error=function(err) FALSE,
                        warning=function(warn) FALSE)
```

```

    if (!is.logical(fit1)) {
      fit1Aic = fit1$aic
      a=c(a,fit1Aic)
    }
  }
  return (min(a)+sd(a)*qnorm(0.6)) # selecciona el 10%
                                  # menor de los datos
}

#' @param xx          - Serie de tiempo
#' @param minOrden - Límite inferior de parámetros
#' @param maxOrden - Límite superior de parámetros
#' @param umbral     - Umbral para seleccionar AIC
BusquedaArima = function (xx, minOrden=c(0,0),
                          maxOrden=c(5,5), trace=FALSE,
                          umbral=FiltroAIC(xx)) {

  mejorAic = umbral
  mejorpq = 25
  len = NROW(xx)
  for (p in minOrden[1]:maxOrden[1])
    for (q in minOrden[2]:maxOrden[2]) {
      if (p == 0 && q == 0)
        next
      fit1=FALSE
      if ((p+q) < mejorpq)
        fit1 <- tryCatch(arima(xx, c(p, 0, q)),
                        error=function(err) FALSE,
                        warning=function(warn) FALSE)

      if (!is.logical(fit1)) {
        fit1Aic = fit1$aic
        # me quedo con el mejor AIC
        if (fit1Aic < mejorAic) {
          mejorAic = fit1Aic

```

---

```
        mejorFit    = fit1
        mejorModel = c(p, q)
        mejorpq     = p+q
    }

    if (trace) {
        ss = paste(sep="", "(", p, ",", q, "): AIC = ",
                    fit1Aic)
        print(ss)
    }
}
else {
    if(trace)
        ss = paste(sep="", "(", p, ",", q, "): None")
    print(ss)
}
}
return(list(aic=mejorAic, fit=mejorFit,
            model=mejorModel))
}
# Modelo con mejor AIC para la serie de tiempo
ARIMA_MODEL=BusquedaArima(serie)
# Modelo para generar parámetros
fit3 <- arima(serie, c(ARIMA_MODEL$model[1], 0,
                      ARIMA_MODEL$model[2]),
              include.mean =
                  ARIMA_MODEL$fit$coef["intercept"])
# Cambio el ruido de la observación
fit3$model$h=50
# Suavizado
sua=KalmanSmooth(serie, fit3$model, nit = 0L)
return (as.xts(sua$smooth, order.by = index(serie)))
}
```



# Bibliografía

- [DH97] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1997.
- [Ehr06] Douglas S. Ehrman. *The handbook of pairs trading: strategies using equities, options, and futures*. Wiley trading. Wiley, 2006.
- [MK98] G. S. Maddala and In-Moo Kim. *Unit roots, cointegration and structural change*. Themes in Modern Econometrics. Cambridge University Press, Cambridge, 1998.
- [Pol07] Andrew Pole. *Statistical arbitrage: algorithmic trading insights and techniques*. Wiley Finance. Wiley, 2007.
- [Vid04] Ganapathy Vidyamurthy. *Pairs trading: quantitative methods and analysis*. Wiley Finance. Wiley, 2004.
- [Whi04] Mark Whistler. *Trading pairs: capturing profits and hedging risk with statistical arbitrage strategies*. Wiley, 2004.