



Universidad de San Andrés

Escuela de Administración y Negocios

Magister en Finanzas

Trabajo final de graduación

Mercado de futuros de soja: una aplicación del filtro de Kalman

Autor: Marco Nicolás Dibo

DNI: 31.769.914

Director: Gabriel Basaluzzo

Córdoba, Argentina, junio 2018

Abstract

En el presente trabajo se analiza la dinámica de precios de los futuros de soja que cotizan en Rosario y en Chicago. En primera instancia, se busca probar que existen períodos en los que se verifica una relación de equilibrio entre los mismos. El objetivo principal se centra en demostrar que, a pesar de que los activos muestren cointegración de precios, la estrategia basada en el filtro de Kalman permite obtener mejores resultados que aquellas basadas en niveles de equilibrio. La razón fundamental reside en la inestabilidad de dichos niveles de equilibrio y en la rapidez del filtro de Kalman para captar modificaciones en los mismos. Para probar que existe una relación de equilibrio entre los activos se utiliza el concepto de cointegración. Luego se calculan los resultados de dos estrategias que difieren en la forma de cálculo del ratio de cobertura: la primera basada en niveles de equilibrio utilizando mínimos cuadrados ordinarios, la segunda mediante el enfoque bayesiano del filtro de Kalman.

Palabras Claves: Filtro de Kalman, Cointegración, Arbitraje.



AGRADECIMIENTOS

A Flor, por acompañarme siempre de manera incondicional, por su cariño y por aguantarme durante toda la maestría. Siempre linda, siempre juntos.

A mis padres, Alejandra y Hugo por su eterno apoyo, confianza y amor.

A mi director, Gabriel Basaluzzo quien ha tenido una enorme influencia en mi carrera, por todos los aportes realizados a este trabajo, y sobre todo por su generosidad para impartir conocimiento.

A mis compañeros de la Maestría en Finanzas, por hacer de esta un experiencia muy especial.

A la Universidad de San Andrés, por el excelente programa y por el personal que la conforman. No solo tuve la posibilidad de acceder a un excelente grupo de docentes, con clases magistrales sino también por la oportunidad de realizar un programa de intercambio en ESCP París que sin duda marcará mi carrera para siempre.

Universidad de
SanAndrés



Universidad de
SanAndrés

Tabla de contenido

Introducción.....	5
Revisión de la literatura.....	6
Datos.....	8
Metodología	9
<i>Test de Raíces Unitarias</i>	10
<i>Estrategia</i>	13
<i>Filtro de Kalman</i>	16
<i>Bandas de Bollinger</i>	18
<i>Optimización y Resultados de la estrategia</i>	21
Resultados	22
<i>Test de raíces unitarias para cada una de las variables</i>	25
<i>Test de cointegración</i>	25
<i>Optimización Estrategia MCO</i>	26
<i>Optimización Estrategia Kalman</i>	27
Conclusiones.....	28
Bibliografía	36



Universidad de
SanAndrés

Introducción

Para un operador de arbitraje resulta indispensable conocer la relación entre los precios de los productos que opera como también poseer un conjunto de estrategias que le permitan explotar las ineficiencias de los mercados donde participa.

La hipótesis primaria de esta investigación es que existen períodos en los que la dinámica entre los precios de los futuros de soja de Chicago y Rosario muestra una relación de equilibrio. Es decir, siguiendo la metodología propuesta por Engle y Granger¹ (1987) mediante la utilización del test de Dickey-Fuller, existe evidencia empírica suficiente para demostrar que los residuos obtenidos de la regresión entre el precio de los futuros de ambos mercados (Chicago y Rosario) para algunos de los períodos comprendidos entre mayo de 2013 y mayo de 2017 son estacionarios. Según los criterios utilizados por Karbuz y Jumah² (1995), esto significa que existen momentos en los que los activos guardan una relación de equilibrio.

Como hipótesis secundaria se plantea que, entre las estrategias de arbitraje estadístico, la estrategia long-short basada en el filtro de Kalman es superior en términos de rendimiento a aquellas estrategias basadas en niveles de equilibrio como las propuestas por Tsay³ (2010), y Moore⁴ (2015). Esto se debe a que dichos niveles de equilibrio no son

¹ Engle, Robert F., and Clive WJ, Granger. 1987. "Co-integration and error correction: representation, estimation, and testing." *Econometrica: journal of the Econometric Society*. 251-276.

² Karbuz, Sohbet, and Adusei Jumah. 1995. "Cointegration and commodity arbitrage". *Agribusiness*. 11 (3): 235.

³ Tsay, Ruey S. 2010. *Analysis of financial time series*. New Jersey: Wiley.

⁴ Halls-Moore, Michael L. 2015. *Advanced Algorithmic Trading*; www.quantstart.com.



Universidad de
SanAndrés

estáticos en el corto plazo, sino que cambian permanentemente. El filtro de kalman es dinámico, y permite detectar rápidamente cambios en la relación entre los activos. Cuando se calcula un ratio de cobertura basado en niveles de equilibrio se deben considerar los últimos n períodos para su cálculo, no siendo así lo suficientemente flexible como para captar cambios en los niveles de equilibrio. La estrategia utilizada es del tipo long-short ya que se compran y se venden futuros de soja Rosario y Chicago de forma simultánea dependiendo del valor relativo de los mismos.

Revisión de la literatura

Siguiendo a Moore (2015) y Tsay (2010), la aplicación de cualquier estrategia de arbitraje estadístico implica conocer la dinámica entre los activos a operar y poder probar si existe o no una relación estable de largo plazo entre los mismos. El presente trabajo tiene como objetivo inicial, probar si existe cointegración entre el precio de los futuros de soja del mercado de Chicago y de Rosario. Karbuz y Jumah (1995) utilizan el concepto de cointegración para examinar relaciones de equilibrio de largo plazo entre los precios de los futuros (y contado) del Cacao y del Café en Nueva York y Londres. En el trabajo los autores obtienen resultados empíricos que muestran que en general los precios del Cacao y del Café tienden a moverse juntos en el largo plazo. El estudio muestra que la evidencia empírica puede utilizarse para apoyar el supuesto de que los precios de los commodities están perfectamente arbitrados en el largo plazo.

Utilizando la metodología planteada por los autores se busca comprobar si existe una relación estable entre los mercados de futuros



Universidad de
SanAndrés

de soja de Rosario y Chicago para cada uno de los períodos comprendidos entre los años 2013 y 2017. Demostrar que los precios de los activos están cointegrados permitiría al operador aplicar algún tipo de estrategia de arbitraje estadístico. Para ello se utiliza el test de raíces unitarias desarrollado por Dickey y Fuller⁵ (1979) utilizado por Engle y Granger⁶ (1987) para probar la existencia de cointegración.

Moore (2015) y Tsay (2010) plantean que, de existir cointegración entre los activos y, por ende, una relación de equilibrio, el operador estaría en condiciones de aplicar una estrategia basada en un ratio de cobertura que refleje la relación de equilibrio entre los mismos.

Según Chan⁷ (2013), cuando existen dos series de precios que no están cointegradas, pero las mismas muestran reversiones de corto plazo o estacionales, es posible utilizar el filtro de Kalman para encontrar un ratio de cobertura dinámico que permita detectar oportunidades de arbitraje.

El objetivo principal de este trabajo es demostrar que tanto con o sin la presencia de cointegración entre los precios de los futuros de soja en Chicago y en Rosario, la estrategia basada en el filtro de Kalman permite obtener mejores resultados que la estrategia basada en una

⁵ David A. Dickey and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". *Journal of the American Statistical Association*. Vol. 74, No. 366: 427-431.

⁶ Engle, Robert F., and Clive WJ, Granger. 1987. "Co-integration and error correction: representation, estimation, and testing." *Econometrica: journal of the Econometric Society*. 251-276.

⁷ Chan, Ernest P. 2013. *Algorithmic trading, winning strategies and their rationale*. New Jersey: Wiley.



Universidad de
SanAndrés

relación de equilibrio de largo plazo. La explicación surge de la misma dinámica de precios: Los niveles de equilibrio de corto plazo no son estáticos, sino que cambian permanentemente. La estrategia basada en el filtro de Kalman permite captar rápidamente los cambios en la relación entre los activos mientras que aquella basada en niveles de equilibrio tiene un mayor rezago para identificar dichos cambios.

Datos

Los precios de los futuros de soja del mercado de Chicago y Rosario para el período mayo de 2013 - mayo de 2017 provienen del sitio web de Rosario Futures Exchange⁸ (ROFEX).

Para el caso de soja Rosario se utilizan los precios del contrato Índice Soja Rosafe. Para el caso de soja Chicago se utiliza el contrato que cotiza en Rosario que tiene como activo subyacente soja de grado n°2 (el mismo activo subyacente del contrato que cotiza en Chicago). Esto se debe a cuestiones de disponibilidad de datos, ya que sólo se cuenta con precios de cierre y los horarios de cierre de los contratos que cotizan en Rosario y en Chicago no coinciden. El contrato de soja Chicago que cotiza en Rosario comenzó a operar en el año 2013, y por esta razón el análisis se realiza desde ese momento. Se descartó del análisis el contrato soja Chicago con vencimiento octubre de 2015 por no haber cotización en el contrato índice de soja Rosario noviembre de 2015.

⁸ www.rofex.com.ar



Metodología

Con el objetivo de conocer si existe una relación de equilibrio entre los precios de la soja en los mercados mencionados se utiliza la metodología propuesta por Karbuz y Jumah (1995): probar si existe cointegración.

En su trabajo los autores definen cointegración como una metodología estadística utilizada para testear relaciones de largo plazo o estacionarias⁹ de equilibrio. Por su parte, Engle y Granger (1987) al definir la cointegración de dos variables establecen que las series de tiempo x_1 y x_2 no estacionarias son cointegradas si existe al menos una combinación lineal de las mismas que es estacionaria.

Según Chan (2013), la combinación más común es la de dos series de precios conocida como “pair trading”: utilizando los precios de los futuros de soja de Rosario y de Chicago, se compran contratos futuros en un mercado y al mismo tiempo se venden contratos en el otro, con una asignación de capital adecuada para cada mercado. Para el autor, “la descripción matemática de una serie de precios estacionaria es que la varianza del logaritmo de los precios aumenta más lentamente que lo que lo hace un camino aleatorio geométrico. No significa que los

⁹ Una serie de precios es estacionaria en el sentido amplio cuando media y varianza no cambian en el tiempo. Para ello se deben cumplir las siguientes condiciones:

- 1) Tanto el valor esperado de la serie como la varianza deben ser constantes y finitos:

$$E(y) = \mu$$
$$|u| < \infty$$

- 2) La covarianza de la serie contra un momento s de esta hacia el pasado o el futuro debe ser constante y finita en todo momento:

$$Cov(y_t, y_{t-s}) = \lambda_s$$
$$|\lambda_s| < \infty$$



Universidad de
SanAndrés

precios se encuentren necesariamente limitados dentro de un rango, con una varianza independiente del tiempo. Solo significa que la varianza aumenta más lentamente que una difusión normal”.

De acuerdo con el trabajo de Engle y Granger (1987), las series x_{1t} y x_{2t} son cointegradas de orden d, b , donde $d \geq b \geq 0$,

$$x_{1t}, x_{2t} \sim CI(d, b)$$

Solo si:

- 1) Ambas series son integradas de orden d .
- 2) Existe una combinación lineal de estas variables que es integrada de orden $d - b$.

Una serie es x_t es integrada de orden d $I(d)$ cuando es necesario diferenciarla d veces para que sea estacionaria. Para conocer si una serie es integrada de orden 1 (posee una raíz unitaria) es necesario realizar el test de Dickey Fuller (DF).

Test de Raíces Unitarias

El test está basado en la siguiente regresión:

$$\Delta x_{1t} = \rho x_{1t-1} + \varepsilon_t \quad (1),$$

donde $\rho > -2$

x_{1t} representa la serie de precios soja Chicago. Para el caso de soja Rosario x_{2t} es necesario realizar el mismo test.

La hipótesis por testear es:



Universidad de
SanAndrés

$H_0: \rho = 0, x_{1t}$ presenta una raíz unitaria

$H_1: \rho < 0, x_{1t}$ es estacionaria.

El estadístico del test es igual al coeficiente de la regresión ρ dividido por el error estándar de la regresión $SE(\rho)$:

$$\text{estadístico} = \frac{\rho}{SE(\rho)}$$

Si el estadístico no resulta menor al valor crítico correspondiente a un nivel de significación del 5%, entonces no es posible rechazar la hipótesis nula, y la serie de precios es integrada de orden 1, $I(1)$.

Para el test de raíces unitarias se utiliza la librería `STATSMODELS`¹⁰ disponible para Python.

Una vez que comprobamos que ambas series presentan raíces unitarias, el siguiente paso es realizar una regresión de mínimos cuadrados ordinarios para encontrar el ratio de cobertura o “hedge ratio”. Este ratio representa el número de unidades que se deben comprar o vender en cada mercado (enfoque utilizado por Chan (2013), Tsay (2010), Moore (2015) y Aldridge¹¹ (2013)).

En el mercado de futuros de Rosario (Rosario Futures Exchange o ROFEX) es posible operar tanto futuros del Índice Soja Rosafe (en adelante futuros de soja Rosario) como también los contratos de futuros de soja Chicago, que tienen como subyacente el precio del futuro de soja listado en el Chicago Mercantile Exchange (CME).

¹⁰ <https://www.statsmodels.org/stable/index.html>

¹¹ Aldridge, Irene. 2013. High-Frequency Trading: A practical guide to algorithmic trading strategies and trading systems. New Jersey: Wiley.



Universidad de
SanAndrés

Es importante destacar que los contratos de soja Rosario y Chicago (ambos listados en ROFEX) están expresados en las mismas unidades, con un multiplicador equivalente a 5 toneladas. Si los contratos a considerar estuviesen expresados en diferentes unidades sería necesario corregir los precios multiplicando los precios de cada contrato por su respectivo multiplicador para que de esta manera el ratio de cobertura refleje correctamente el número de contratos.

La regresión de mínimos cuadrados ordinarios para estimar el ratio de cobertura es la siguiente:

$$x_{1t} = \alpha + \beta * x_{2t} + u_t \quad (2)$$

Donde x_1 y x_2 son los precios de los futuros de soja en Chicago y en Rosario respectivamente, β es el ratio de cobertura y u los residuos de la regresión.

A partir de los mismos se realiza un nuevo test de DF para comprobar si los residuos son estacionarios (enfoque de Engle y Granger utilizado en el trabajo de Karbuz y Jumah).

El test está basado en la siguiente regresión:

$$\Delta y_t = \lambda y_{t-1} + \varepsilon_t \quad (3)$$

Donde y es el portfolio estimado en (2) asumiendo un desvío igual a cero:

$$y = x_{1t} - \beta x_{2t} \quad (4)$$

La hipótesis por testear es:

$H_0: \lambda = 0, y_t$ no es estacionaria y las series no son cointegradas,

$H_1: \lambda < 0, y_t$ es estacionaria y las series son cointegradas.

El estadístico del test es igual al coeficiente de la regresión λ dividido por el error estándar de la regresión $SE(\lambda)$:

$$estadístico = \frac{\lambda}{SE(\lambda)}$$

Si el estadístico resulta menor al valor crítico correspondiente a un nivel de significación del 5%, entonces se rechaza la hipótesis nula y existe cointegración entre las series de precios.

Estrategia

Una estrategia de arbitraje consiste en definir un modelo para aprovechar ineficiencias en los mercados. Este modelo debe detectar las oportunidades que poseen la mayor probabilidad de éxito como también definir las reglas de entrada y salida.

Para definir la estrategia de arbitraje de futuros soja se deben responder las siguientes preguntas:

- 1) ¿Qué vencimientos operar?
- 2) ¿Cuántos contratos se debe comprar o vender de cada futuro cuando se detecta una oportunidad?
- 3) ¿Cuándo es el momento de realizar una operación?

Para responder la primera pregunta es necesario conocer qué vencimientos se operan en cada mercado y la liquidez con la que cuentan. En el caso de los futuros de soja Chicago, prácticamente todos los contratos con vencimientos cercanos (un año o menos) poseen liquidez. Para los futuros de soja Rosario no sucede lo mismo, y sólo

se encuentra liquidez en aquellos contratos con vencimientos en mayo y noviembre, por lo tanto, se analiza el arbitraje para los meses en cuestión. Vale la pena recordar que cuando se hace mención de los contratos de futuros de soja Chicago nos referimos a los contratos listados en Rosario con ajuste en base al contrato operado en CME y no a los operados directamente en este último. La liquidación de los contratos soja Chicago de ROFEX se realiza tomando el precio de ajuste de la primera posición abierta del contrato de futuros de soja del Chicago Mercantile Exchange (CME). Es decir, el contrato soja Chicago abril que cotiza en ROFEX ajusta contra el precio de ajuste del contrato de futuros soja mayo que cotiza en CME.

Los pares de vencimientos a utilizar son: abril Chicago contra mayo Rosario y octubre Chicago contra noviembre Rosario.

La segunda pregunta tiene relación con el concepto de cointegración definido previamente. El ratio de cobertura que se obtiene de la regresión utilizado para comprobar si las series de precios estaban cointegradas permite a su vez tener una idea de la cantidad de contratos que se debe comprar o vender en cada mercado. Este ratio se calcula mediante un regresión lineal utilizando el período completo en que se operan los diferentes vencimientos. Es decir, para calcular el ratio de cobertura entre los futuros de soja noviembre 2014 Rosario y soja octubre 2014 Chicago, se utilizan todos los datos desde el 14-4-2014 hasta el 15-10-2014.

Para el análisis de cointegración se utiliza el período completo, aunque esto no sería correcto si quisiéramos calcular el ratio de



Universidad de
SanAndrés

cobertura a aplicar en la estrategia. Esto se debe fundamentalmente a dos razones:

- 1) Tendríamos que esperar al vencimiento de ambos contratos para disponer con la serie completa de precios y esto no nos serviría al momento de operar.
- 2) Estaríamos incurriendo en el sesgo retrospectivo o “look-ahead bias”. Incurrimos en este sesgo cuando utilizamos datos que todavía no disponemos para el cálculo de un indicador. Para calcular el ratio de cobertura en un momento determinado no es posible utilizar datos de fechas posteriores, solo datos de los que se dispone en ese momento.

Utilizaremos un ratio de cobertura móvil, utilizando los últimos n días para cada período, permitiendo así que el mismo se vaya actualizando con los últimos datos y descartando los anteriores. El valor de n surge de la optimización de las estrategias que se discutirá más adelante. Como se observa en el anexo n°1 el ratio de cobertura varía durante el plazo en que se opera cada contrato como también entre los diferentes vencimientos. De esta manera no sería correcto utilizar un único ratio para determinar las cantidades que se deben comprar o vender ya que la relación entre los activos varía constantemente. Llamaremos “estrategia MCO” (mínimos cuadrados ordinarios) a la estrategia que utiliza un ratio de cobertura estimado de esta manera.



Universidad de
SanAndrés

Filtro de Kalman

Como alternativa al cálculo de un ratio móvil por regresión se utiliza el filtro de Kalman. El mismo permite obtener un ratio de cobertura dinámico, que incorpora los cambios en el tiempo y no depende del período contemplado para su cálculo.

El filtro de Kalman es básicamente un conjunto de ecuaciones matemáticas que implementan un estimador del tipo predictor corrector que resulta óptimo en el sentido que minimiza el error cuadrático medio de las variables estimadas. Chan (2013) lo define como un algoritmo lineal óptimo que actualiza el valor esperado de una variable oculta basado en el último valor de una variable observable. Es lineal dado que asume que la variable observable es una función lineal de la variable oculta con ruido. También asume que la variable oculta en el momento t es una función lineal de sí misma en el momento $t-1$ con ruido, y que los ruidos presentes poseen una distribución gaussiana.

Siguiendo la explicación utilizada por Chan (2013), para aplicar el filtro de Kalman es necesario definir las siguientes variables y matrices:

- Variable observable (vector)
- Variable oculta (vector)
- Modelo de transición de estado (matriz)
- Modelo de observación (matriz)



Universidad de
SanAndrés

La **variable observable** del modelo es una de las series de precios (y), en este caso, los precios de los futuros de soja en Chicago. La **variable oculta** es el ratio de cobertura (β). La función lineal que relaciona estas variables se conoce como “ecuación de medición”:

$$y(t) = x(t)\beta(t) + \epsilon(t), \quad (\text{“Ecuación de medición”}) \quad (5)$$

Donde x es la serie de precios de los futuros de soja en Rosario, y ϵ es el ruido gaussiano de varianza $V(\epsilon)$.

El vector β es de la forma 2×1 ya que incorpora el cálculo de una media no negativa del spread entre los precios. Por esta razón, el vector $x(t)$ incorpora una columna de unos con la forma $N \times 2$, siendo N el número de observaciones. El vector x constituye el “**modelo de observación**”.

El autor realiza el supuesto de que la variable oculta en el momento t es igual al valor de la misma en el momento $t - 1$ más ruido:

$$\beta(t) = \beta(t - 1) + \omega(t - 1), \quad (\text{“Estado de transición”}) \quad (6)$$

Donde ω representa ruido gaussiano con covarianza V_ω . Aquí el “**modelo de transición de estado**” está representado por la matriz identidad.

Una vez definidas las variables y matrices anteriores, es posible comenzar la iteración y utilizar el filtro de Kalman para obtener los valores de la variable oculta, en este caso, el ratio de cobertura. Llamaremos a esta estrategia “estrategia Kalman”. El proceso iterativo del algoritmo puede encontrarse en el anexo N°2.



Universidad de
SanAndrés

Uno de los beneficios de utilizar el filtro de Kalman planteado por Chan, es que no sólo se consigue una medida dinámica del ratio de cobertura, sino que a la vez se obtiene una media móvil del spread y la desviación estándar de la misma. Esto resulta de gran importancia a la hora de responder la tercer pregunta: ¿cuándo es el momento de realizar una operación?

Bandas de Bollinger

Se utiliza el concepto de bandas de Bollinger desarrollado por John Bollinger y utilizado por Tsay (2013) en su estrategia de “pair trading” para responder esta pregunta. Las bandas constituyen un indicador técnico que contempla la volatilidad del mercado y se obtienen de la siguiente manera:

Banda superior: se suman z desviaciones estándar a la media del precio.

Banda inferior: se restan z desviaciones estándar a la media del precio.

Si $z = \pm 2$, esto significa que si la serie de tiempo (en este caso el spread) tuviese una distribución normal, el 95.4% de los spreads se encontrarían dentro de estas bandas.

Se utiliza el concepto conocido como puntuación estándar para definir los puntos de entrada y de salida de la estrategia. Para ello se debe, en primer lugar, definir los límites de compra y de venta del spread, es decir, el número de desviaciones estándar respecto de la media a las que se realiza una operación.



Universidad de
SanAndrés

En segundo lugar, es necesario calcular la puntuación estándar en cada momento del tiempo:

$$z_i = \frac{(s_i - \bar{s})}{std} \quad (\text{"Puntuación estándar"}) \quad (7)$$

Donde s_i representa al spread en el momento i , \bar{s} es la media móvil de spread y std es la desviación estándar de la media.

Tanto \bar{s} como std surgen del cálculo del filtro de Kalman evitando de esta manera introducir un parámetro libre: el período ventana para dicho cálculo. Mientras mayor sea el número de parámetros libres a optimizar en la estrategia, mayor es la probabilidad de cometer el error conocido como "data snooping bias". Este tipo de error es causado por tener demasiados parámetros libres ajustados a patrones aleatorios de mercado ocurridos en el pasado. De esta forma se comete el error de optimizar estos parámetros hasta obtener una estrategia con buen rendimiento histórico, cuando en realidad se trata de una serie de relaciones espurias. Para el caso de la estrategia MCO donde el ratio de cobertura móvil es calculado mediante una regresión, resulta necesario definir el período ventana para su cálculo como el de la media y desviación estándar del spread, introduciendo así un parámetro libre, n .

Las reglas de la estrategia son las siguientes:

Si $z > z^*$ (encima de la banda superior), se debe vender el spread.

Donde z^* representa el número de desviaciones estándar óptimo y el spread está definido por la ecuación (8):



Universidad de
SanAndrés

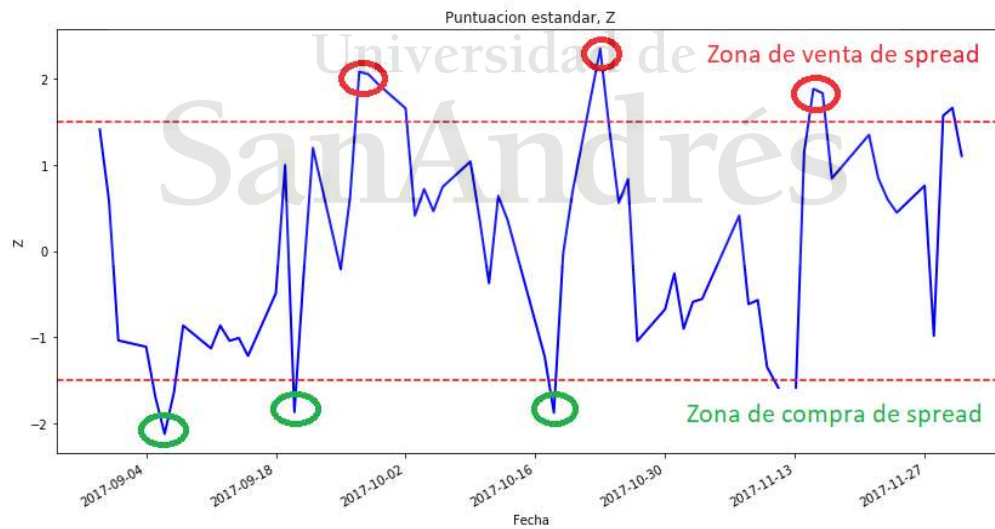
$s = x_{1t} - \beta x_{2t}$ (8), vender un contrato de x_1 (soja Chicago) y comprar β contratos de x_2 (soja Rosario).

Si $z < -z^*$ (debajo de la banda inferior), se debe comprar el spread: comprar un contrato de soja Chicago y vender β contratos de soja Rosario.

Si $z \leq z^*$ o $z \geq -z^*$ y existe una posición abierta, cerrar dicha posición mediante órdenes inversas.

En el siguiente gráfico es posible observar un ejemplo de cómo quedarían conformadas las zonas de venta y compra de spread con un $z^* = 1.5$:

Figura N°1: Puntuación Estándar



Los círculos verdes representan oportunidades de compra mientras que los rojos oportunidades de venta del spread.

Optimización y Resultados de la estrategia

Para evaluar la estrategia se calcula un conjunto de métricas estándar que le permiten conocer al operador los resultados esperados de la misma. Se utilizan los siguientes criterios o métricas:

- Rentabilidad porcentual anualizada.
- El ratio de Sharpe. Este representa una medida de exceso de rendimiento ajustado por riesgo. Para calcularlo debemos obtener el retorno promedio de la estrategia o activo menos el costo de fondeo y dividirlo por la desviación estándar de dichos retornos.
- Cantidad de operaciones

Dado que se trata de una estrategia long-short, se supone que el costo de fondeo es 0 ya que los márgenes utilizados en ambas posiciones se colocan a un interés similar al costo de oportunidad de los fondos.

En la optimización se busca obtener el mayor retorno anualizado para cada uno de los períodos contemplados, con el mayor ratio de sharpe posible. A su vez mientras mayor sea el número de operaciones mayor será la consistencia de los resultados obtenidos.

Para el caso de la estrategia MCO los parámetros a optimizar serán los siguientes:

- a) El número de desviaciones estándar, z^* .
- b) La cantidad de observaciones para el cálculo del ratio de cobertura, n . Tanto la media móvil del spread como la desviación estándar del mismo serán calculados utilizando el



Universidad de
SanAndrés

mismo número de períodos que se hayan considerado para el ratio. Es importante recordar que para la estrategia de Kalman tanto la media móvil del spread como la desviación estándar del mismo surgen del algoritmo de cálculo del ratio de cobertura, evitando así la introducción de un parámetro libre.

En la estrategia de Kalman solo se busca optimizar el parámetro z^* .

Resultados

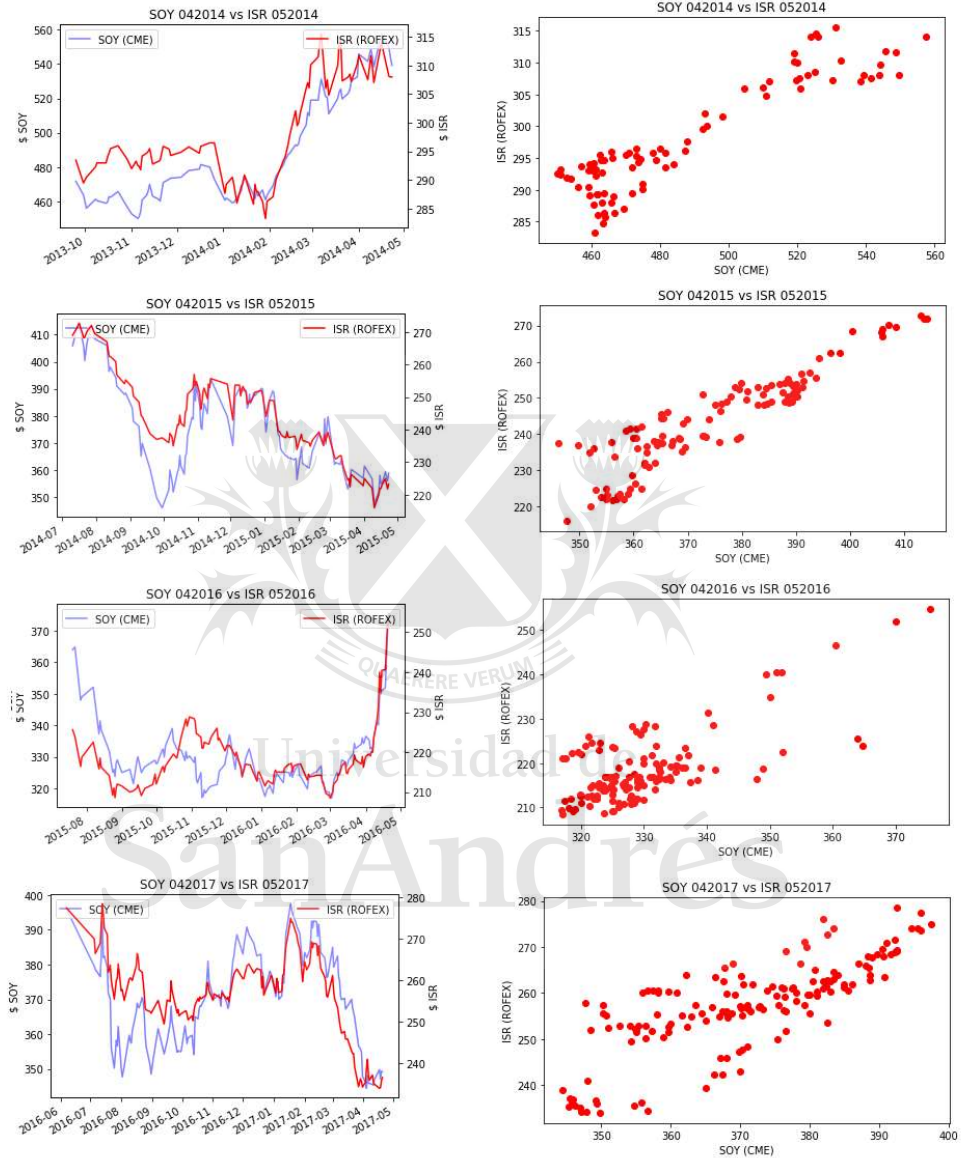
En las figuras n°2 y n°3 se observan los precios tanto de los contratos soja Chicago como los de soja Rosario para los pares abril-mayo y octubre-noviembre respectivamente.

Los gráficos de dispersión reflejan una clara correlación positiva entre los precios de ambos mercados para la mayoría de los pares. Esto nos hace suponer que los activos pueden tener una relación de equilibrio en el largo plazo, y que es posible desarrollar una estrategia de arbitraje estadístico.



Universidad de
San Andrés

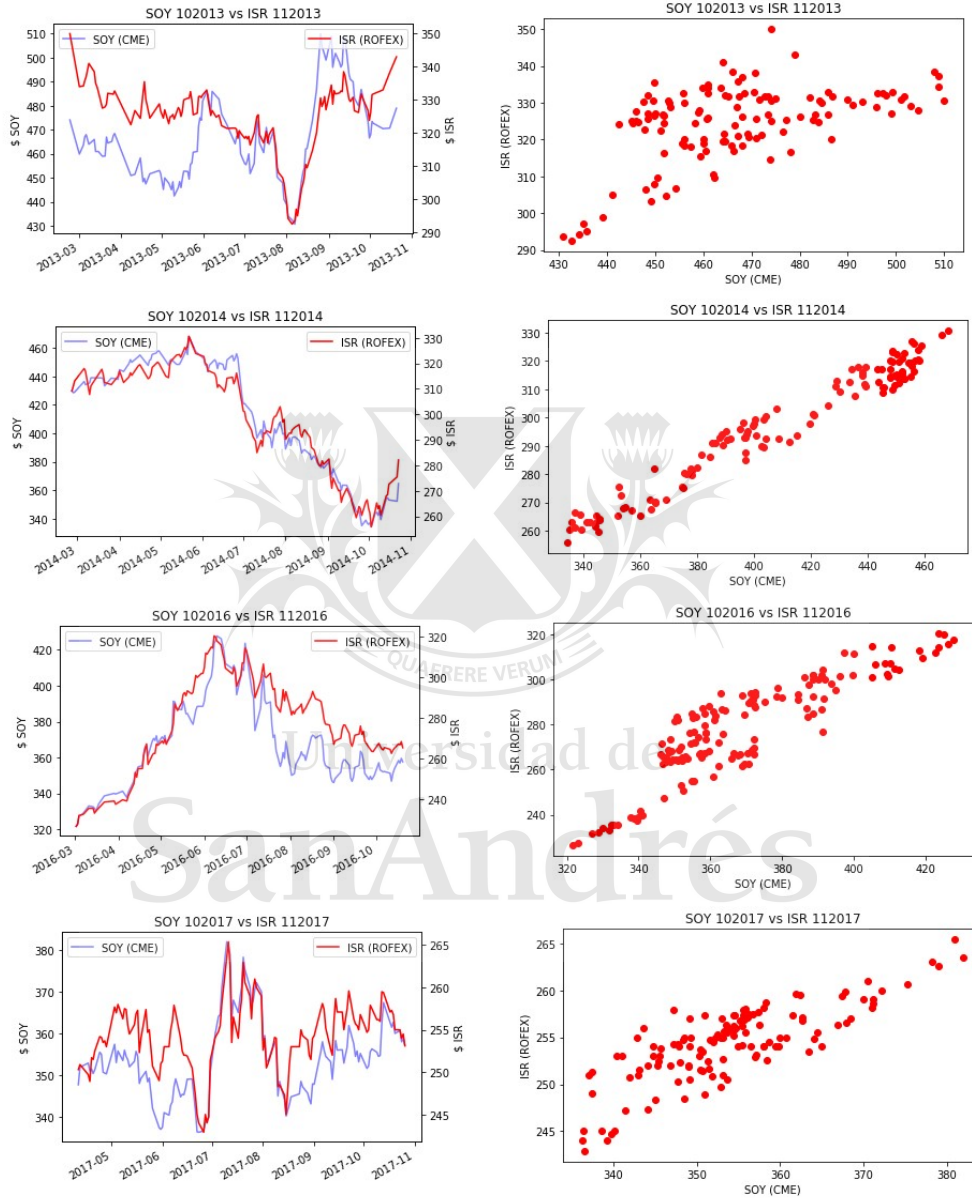
Figura N°2: Soja CME abril vs Índice Soja Rosafe ROFEX mayo





Universidad de
San Andrés

Figura N°3: Soja CME octubre vs Índice Soja Rosafe ROFEX noviembre



Test de raíces unitarias para cada una de las variables

Los resultados del test de DF para cada una de las series de precios se presentan en la siguiente tabla:

Tabla N°1: Test de DF para comprobar si las variables son $I(1)$.

Variable	Año	Observaciones	Valores críticos ($\alpha = 5\%$)	Estadístico	Se rechaza Ho?
soy_octubre	2013	124	-2.88	-1.83	NO
isr_noviembre	2013	124	-2.88	-2.39	NO
soy_octubre	2014	128	-2.88	-0.43	NO
isr_noviembre	2014	128	-2.88	-0.83	NO
soy_octubre	2016	144	-2.88	-2.06	NO
isr_noviembre	2016	144	-2.88	-2.3	NO
soy_abril	2014	86	-2.89	-0.02	NO
isr_mayo	2014	86	-2.89	-1.08	NO
soy_abril	2015	112	-2.88	-2.35	NO
isr_mayo	2015	112	-2.88	-1.51	NO
soy_abril	2016	131	-2.88	-1.16	NO
isr_mayo	2016	131	-2.88	1.33	NO
soy_abril	2017	154	-2.88	-2.17	NO
isr_mayo	2017	154	-2.88	-1.66	NO

De acuerdo con la tabla anterior no es posible rechazar la hipótesis nula para ninguna de las variables de precios. Es decir, todas las variables son integradas de orden 1 (poseen raíces unitarias) y podemos continuar con la prueba de cointegración.

Test de cointegración

Habiendo confirmado que las series de precios son integradas de orden 1 es posible realizar el test de DF para el portfolio y estimado en (4).

Los resultados de este se presentan en la tabla n°2:

Tabla N°2: Test de DF para comprobar cointegración

Contratos	Año	Observaciones	Valores críticos ($\alpha = 5\%$)	Estadístico	Cointegración?
soyoct13 vs sefnov13	2013	124	-2.88	-1.75	NO
soyabr14 vs sefnov14	2014	86	-2.89	-1.52	NO
soyabr14 vs sefnov14	2014	128	-2.88	-2.99	SI
soyabr15 vs sefnov15	2015	112	-2.88	-2.4	NO
soyabr15 vs sefnov15,	2015	na	na	na	na
soyabr16 vs sefnov16	2016	131	-2.88	-3.58	SI
soyabr16 vs sefnov16	2016	144	-2.88	-1.86	NO
soyabr17 vs sefnov17	2017	154	-2.88	-2.15	NO

Utilizando un nivel de significación del 5% en el test de raíces unitarias (con hipótesis nula de que los activos no están cointegrados) solamente es posible rechazar la hipótesis nula en dos ocasiones: octubre-noviembre 2014 y abril-mayo 2016. En el par abril-mayo 2015 el valor crítico está muy próximo al del estadístico, sin embargo, no lo suficientemente negativo como para rechazar la hipótesis nula. En el resto de los pares no se verifica la existencia de cointegración. A pesar de que no todos los pares estén cointegrados, esto permite confirmar que por períodos las series de precios guardan una relación de equilibrio. Esta relación no es estable, por lo que utilizar una relación de equilibrio estática no sería lo adecuado para desarrollar una estrategia de trading.

Optimización Estrategia MCO

El anexo n°3 contiene los resultados de las métricas para diferentes valores de **n** y de **z**.

Ninguna de las combinaciones permite obtener resultados consistentes en el tiempo en términos de rendimiento porcentual anualizado. Aquellas marcadas con doble asterisco son las que generaron el mejor retorno anualizado para ese par de contratos. La combinación **z = 1** y **n = 40** fue la



Universidad de
SanAndrés

que generó el mayor retorno porcentual anualizado promedio: 1.73%. En promedio, el coeficiente de Sharpe fue de 0.48 y se hicieron 4.42 operaciones para los siete períodos analizados.

Optimización Estrategia Kalman

En el anexo n°4 se encuentran los resultados de la estrategia de Kalman para diferentes valores de **z**.

Tanto para el caso del par Soja CME abril vs Soja ROFEX mayo como para el par Soja CME octubre vs Soja ROFEX noviembre, los mejores resultados se obtienen para un **z = 1.5**:

Tabla N°3: Resultados de la estrategia de Kalman para **z = 1.5**

	SOY 04 vs ISR 05 2014	SOY 04 vs ISR 05 2015	SOY 04 vs ISR 05 2016	SOY 04 vs ISR 05 2017
APR	4.25	32.02	10.65	16.32
Sharpe	1.06	3.91	1.58	2.36
N° of trades	17	21	22	25

	SOY 10 vs ISR 11 2013	SOY 10 vs ISR 11 2014	SOY 10 vs ISR 11 2016
APR	27.88	13.42	14.54
Sharpe	2.99	2.42	1.71
N° of trades	22	21	21

En promedio, el retorno anualizado fue de 17.01% con un coeficiente de Sharpe de 2.29 y 21.28 operaciones.

Conclusiones

Del análisis de las pruebas de cointegración se verifica que la dinámica de precios muestra una relación de equilibrio para algunos de los períodos analizados. Como se observó en el anexo n°1, la relación representada por el ratio de cobertura o “hedge ratio” es inestable, y una estrategia basada en un ratio de cobertura estático no permitiría reaccionar a los cambios en los niveles de equilibrio.

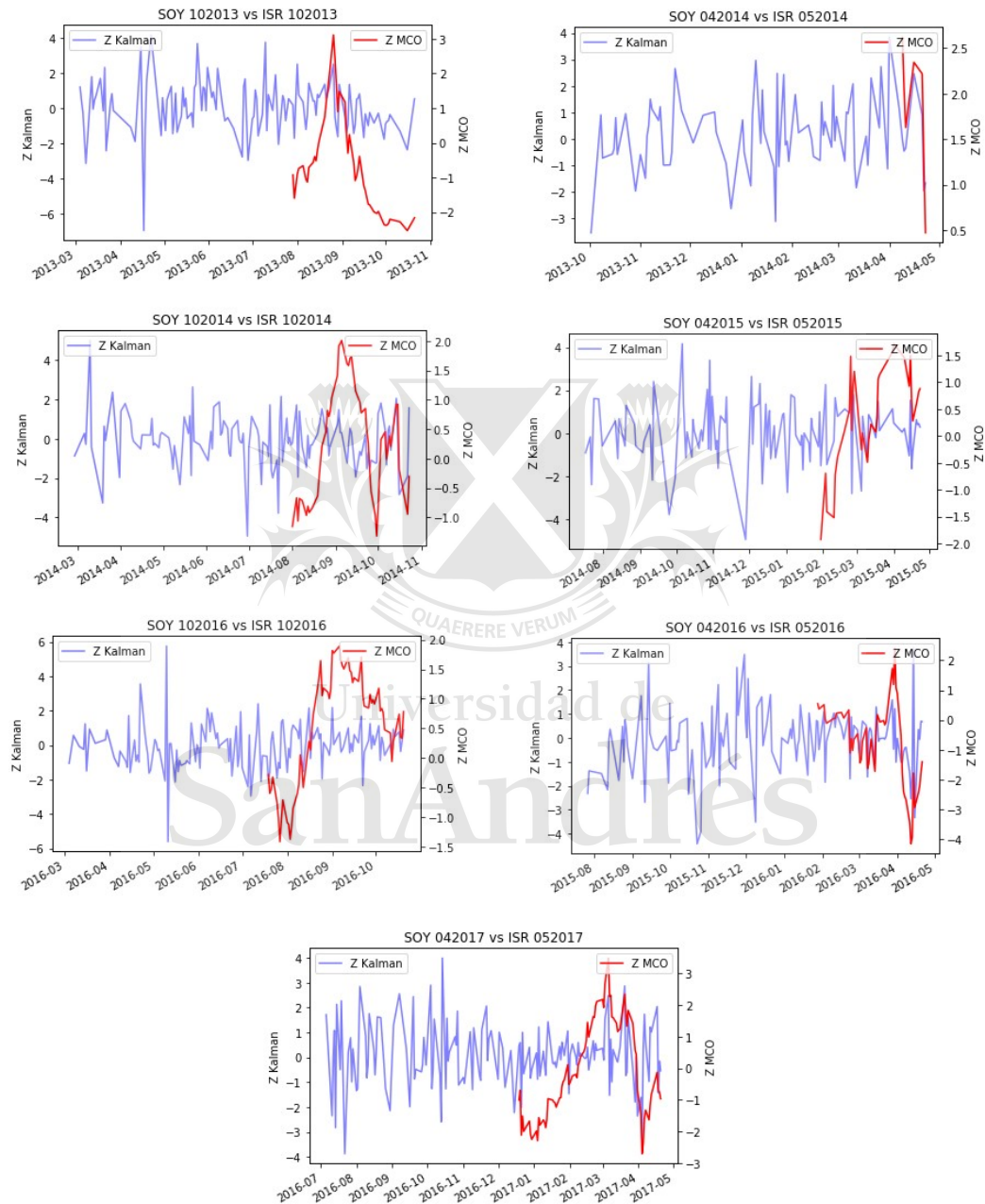
Se analizaron dos estrategias en las que se calculó un ratio de cobertura dinámico: la estrategia MCO y la estrategia de Kalman. Se demostró que la estrategia basada en el filtro de Kalman permite obtener resultados superiores a los de la estrategia MCO. La razón fundamental radica en que el filtro de Kalman no se basa en niveles de equilibrio sino en una relación dinámica entre precios y de esta forma permite captar rápidamente los cambios en dicha relación (anexo n°1). Esta estrategia no solo genera un mayor retorno anualizado y un mayor coeficiente de Sharpe promedio, sino que también permite realizar una mayor cantidad de operaciones que la estrategia MCO. Esto último se debe a dos razones:

- 1) La estrategia de MCO requiere de una cierta cantidad mínima n de datos para el cálculo del ratio de cobertura. Si $n = 40$, eso significa que con esta estrategia debemos esperar al menos 40 días para obtener el primer beta o ratio de cobertura. Dado que para el cálculo de la media y desviación estándar se utiliza la misma cantidad de períodos, entonces el tiempo se duplica ya que debemos contar con al menos 40 ratios de cobertura para calcular el z .
- 2) Una mayor sensibilidad en la puntuación estándar (z) de la estrategia de Kalman.



Universidad de
San Andrés

Figura N°4: Puntuación estándar.



En el gráfico anterior es posible observar cómo la puntuación estándar en la estrategia de Kalman varía con mayor velocidad que en el caso de la



Universidad de
SanAndrés

estrategia de MCO. Esto genera una mayor cantidad de señales de entrada y salida y, por ende, una mayor cantidad de operaciones.

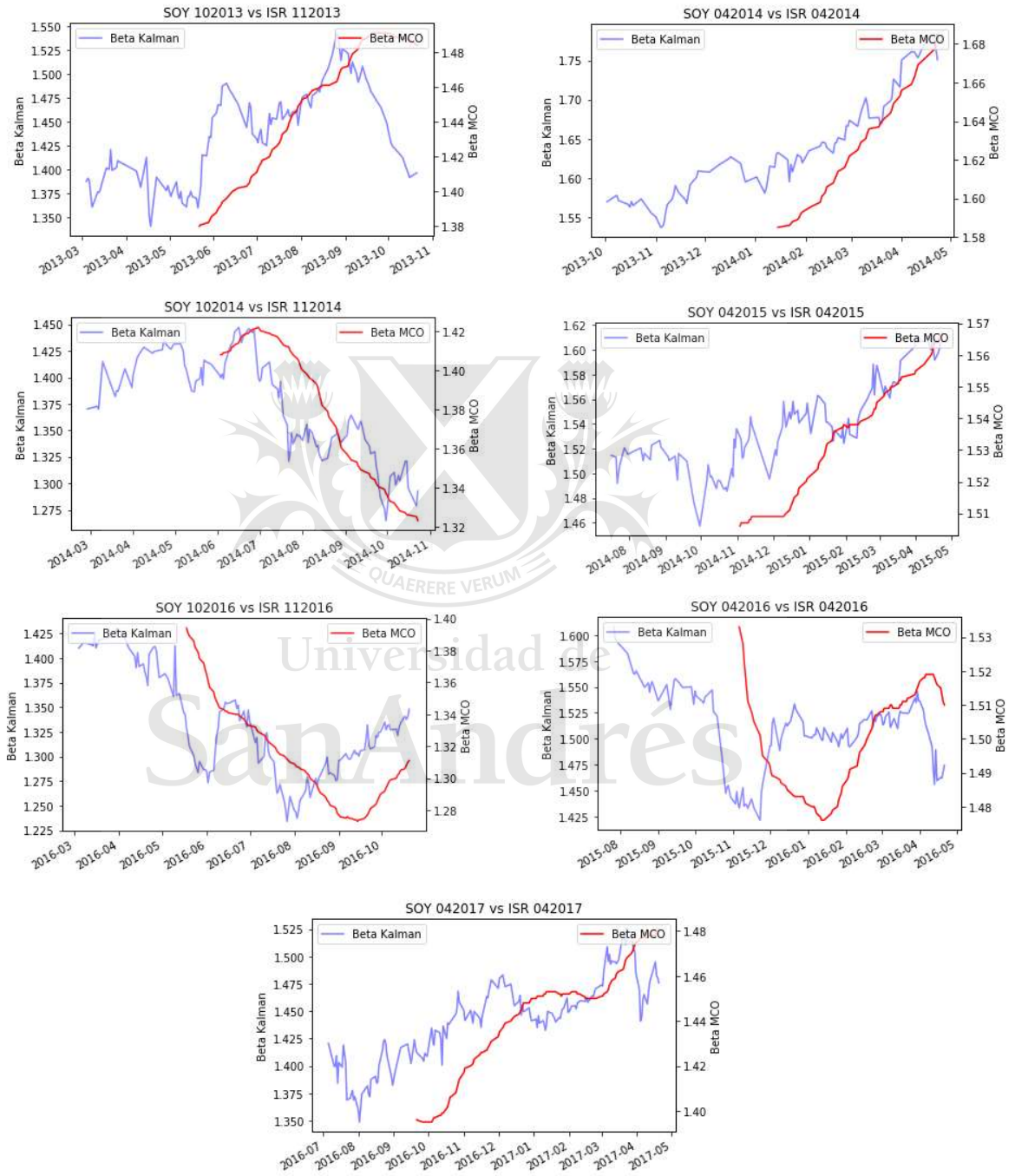


Universidad de
SanAndrés



Universidad de
San Andrés

Anexo N°1: Ratios de cobertura



Anexo N°2: Proceso iterativo para el cálculo del filtro de Kalman según Chan (2013).

$\hat{\beta}(t|t-1)$: valor esperado de β en el momento t dada la observación en el momento $t-1$.

$\hat{\beta}(t|t)$: valor esperado de β dada la observación en el momento t .

$\hat{y}(t|t-1)$: valor esperado de $y(t)$ dada la observación en $t-1$.

Dados $\hat{\beta}(t-1|t-1)$ y $R(t-1|t-1)$ en el momento $t-1$, podemos hacer las predicciones de un paso:

$$\hat{\beta}(t|t-1) = \hat{\beta}(t-1|t-1) \text{ (“Predicción de estado”)}$$

$$R(t|t-1) = R(t-1|t-1) + V_w \text{ (“Predicción de covarianza de estado”)}$$

$$\hat{y}(t) = x(t)\hat{\beta}(t|t-1) \text{ (“Predicción de medición”)}$$

$$Q(t) = x(t)'R(t|t-1)x(t) + V_e \text{ (“Predicción de varianza de medición”)}$$

Donde $R(t|t-1)$ es la covarianza del error de la variable oculta estimada $cov(\beta(t) - \hat{\beta}(t|t-1))$. Se trata de una covarianza ya que $\hat{\beta}$ posee dos componentes independientes.

De la misma manera $R(t|t)$ es $cov(\beta(t) - \hat{\beta}(t|t))$.

Recordando que la variable oculta consiste en ambas media del spread y ratio de cobertura, R es una matriz 2×2 .

$e(t) = y(t) - x(t)\hat{\beta}(t|t-1)$ es el error estimado para $y(t)$ dada la observación en $t-1$ y $Q(t)$ es la varianza del error estimado $var(e(t))$

Luego de observar la medición en el momento t , las ecuaciones de Kalman de actualización de estado y de covarianza son:

$$\hat{\beta}(t|t) = \hat{\beta}(t|t-1) + K(t)*e(t) \text{ (“Actualización de estado”)}$$

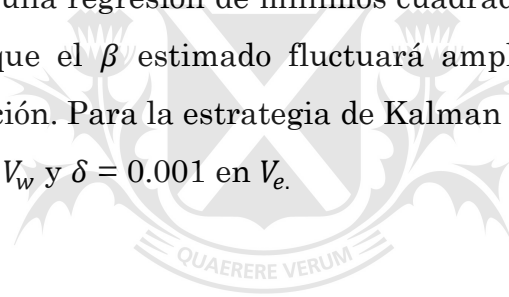


Universidad de
SanAndrés

$R(t|t) = R(t|t - 1) - K(t) * x(t) * R(t|t - 1)$ (“Actualización de covarianza de estado”), donde $K(t)$ es la llamada “Ganancia de Kalman” y está dada por:
 $K(t) = R(t|t - 1) * x(t) / Q(t)$

Para comenzar el proceso recursivo asumimos que $\hat{\beta}(1|0) = 0$ y $R(0|0) = 0$.

Chan utiliza la metodología propuesta por Montana (2009) y asume que $V_w = \frac{\delta}{1-\delta} I$, donde δ es un parámetro entre 0 y 1 e I es una matriz identidad 2×2 . Si $\delta = 0$, eso significa que $\beta(t) = \beta(t - 1)$ lo que lleva a que el filtro de Kalman sea una regresión de mínimos cuadrados ordinarios. Si $\delta = 1$, esto significa que el β estimado fluctuará ampliamente basado en la última observación. Para la estrategia de Kalman se utilizó un $\delta = 0.0001$ en el cálculo de V_w y $\delta = 0.001$ en V_e .



Universidad de
SanAndrés

Anexo N°3: Optimización de la estrategia MCO.

SOY 04 vs ISR 04 2014	Z								
	Z = 0.5			Z = 1			Z = 1.5		
n	20	30	40	20	30	40	20	30	40
APR	-0.47	-8.12	1.75	16.34*	-6.76	1.75	5.26	1.75	-0.79
Sharpe	-0.02	-1.21	0.52	2.82	-1.2	0.52	1.49	0.52	-0.24
N° of trades	8	3	0	9	3	0	7	0	0

SOY 04 vs ISR 04 2015	Z								
	Z = 0.5			Z = 1			Z = 1.5		
n	20	30	40	20	30	40	20	30	40
APR	25.75*	10.28	8.57	16.68	13.78	13.51	10.66	12.31	6.66
Sharpe	2.72	1.53	1.64	2.44	2.42	2.82	2.01	2.6	2.29
N° of trades	15	7	4	9	7	5	6	6	3

SOY 04 vs ISR 04 2016	Z								
	Z = 0.5			Z = 1			Z = 1.5		
n	20	30	40	20	30	40	20	30	40
APR	-5.46	0.62*	-0.14	-5.51	-5.31	-1.53	-4.64	-1.15	-3.23
Sharpe	-0.57	0.12	0.01	-0.73	-0.82	-0.24	-0.65	-0.21	-0.57
N° of trades	9	9	6	13	3	7	8	5	5

SOY 04 vs ISR 04 2017	Z								
	Z = 0.5			Z = 1			Z = 1.5		
n	20	30	40	20	30	40	20	30	40
APR	10.5*	0.81	-5.04	4.9	7.11	-2.57	7.09	3.4	2.86
Sharpe	1.26	0.15	-0.78	0.73	1.19	-0.42	1.23	0.71	0.71
N° of trades	20	8	4	18	12	5	12	12	6

SOY 10 vs ISR 11 2013	Z								
	Z = 0.5			Z = 1			Z = 1.5		
n	20	30	40	20	30	40	20	30	40
APR	-26.38	-11.09	-9.19	-27.19	-6.56*	-7.44	-18.63	-17.85	-8.43
Sharpe	-3.29	-1.47	-1.56	-3.76	-0.99	-1.4	-3.41	-4.36	-1.77
N° of trades	8	5	4	8	8	6	10	2	3

SOY 10 vs ISR 11 2014	Z								
	Z = 0.5			Z = 1			Z = 1.5		
n	20	30	40	20	30	40	20	30	40
APR	-10.83	-1.12	10.17*	-5.24	-4.32	6.47	2.58	-5.12	2.54
Sharpe	-1.15	-0.09	1.6	-0.64	-0.61	1.73	0.53	-0.88	1.12
N° of trades	6	5	4	10	4	3	9	7	1

SOY 10 vs ISR 11 2016	Z								
	Z = 0.5			Z = 1			Z = 1.5		
n	20	30	40	20	30	40	20	30	40
APR	-14.01	2	-3.08	-14.35	2.79	1.92	-2.3	4.03	4.32*
Sharpe	-1.67	0.3	-0.5	-2	0.47	0.39	-0.42	0.89	1.41
N° of trades	14	8	5	9	7	5	8	4	3

*Representan la mejor combinación en términos de retorno anualizado.



Universidad de
SanAndrés

Anexo N°4: Optimización de la estrategia de Kalman

	SOY 10 vs ISR 11 2013			SOY 10 vs ISR 11 2014			SOY 10 vs ISR 11 2016		
	Z = 0.5	Z = 1	Z = 1.5	Z = 0.5	Z = 1	Z = 1.5	Z = 0.5	Z = 1	Z = 1.5
APR	-6.92	-0.31	27.88*	-11.02	1.07	13.42*	19.3*	18.89	14.54
Sharpe	-0.51	0.03	2.99	-1.21	0.18	2.42	1.52	1.83	1.71
N° of trades	55	41	22	46	36	21	63	41	21

	SOY 04 vs ISR 05 2014			SOY 04 vs ISR 05 2015			SOY 04 vs ISR 05 2016			SOY 04 vs ISR 05 2017		
	Z = 0.5	Z = 1	Z = 1.5	Z = 0.5	Z = 1	Z = 1.5	Z = 0.5	Z = 1	Z = 1.5	Z = 0.5	Z = 1	Z = 1.5
APR	-5.38	8.31*	4.25	39.61*	31.88	32.02	14.1*	12.69	10.65	19.42*	8.6	16.32
Sharpe	-0.54	1.47	1.06	2.9	3.35	3.91	1.31	1.41	1.58	1.77	1	2.36
N° of trades	36	23	17	59	38	21	53	33	22	69	40	25

*Representan la mejor combinación en términos de retorno anualizado.



Universidad de
SanAndrés



Universidad de
SanAndrés

Bibliografía

- Aldridge, I., 2013. High-Frequency Trading: A practical guide to algorithmic trading strategies and trading systems. *Wiley*.
- Chan, E., 2013. Algorithmic trading, winning strategies and their rationale. *Wiley*.
- Karbuz, S. & Jumah, A., 1995. Cointegration and commodity arbitrage. *Agribusiness*, 11(3), pp.235–243.
- Montana, Triantafyllopoulos & Tsagaris, 2009. Flexible least squares for temporal data mining and statistical arbitrage. *Expert Systems With Applications*, 36(2), pp.2819–2830.
- Engle, R.F. & Granger, C.W.J., 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), pp.251–276.
- Dickey, D.A. & Fuller, W.A., 1979. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), pp.427–431.
- Markowitz, H., 1952. PORTFOLIO SELECTION*. *Journal of Finance*, 7(1), pp.77–91.
- Tsay, R.S., 2010. Analysis of financial time series. *Wiley*.
- Halls-Moore, M., 2015. Advanced Algorithmic Trading. www.quantstart.com
- Elton, Gruber & Gruber, Martin Jay, 1995. *Modern portfolio theory and investment analysis* 5th ed., New York: John Wiley.