

**APLICACIÓN DEL STATISTICAL ARBITRAGE A LA
SELECCIÓN DE UN PORTAFOLIO DE PARES BASADO EN
COINTEGRACIÓN PARA MERCADOS EMERGENTES**

Área de investigación: Finanzas

Daniel Ulises Urrutia Martínez

Escuela Superior de Economía
Instituto Politécnico Nacional
México

danielulises.urrutia@gmail.com

Agustín Cabrera Llanos

Unidad Profesional Interdisciplinaria de Biotecnología
Instituto Politécnico Nacional
México

aicllbuda@yahoo.com

Francisco Ortiz Arango

Escuela de Ciencias Económicas y Empresariales
Universidad Panamericana
México

fortizar@up.edu.mx

XVIII
CONGRESO
INTERNACIONAL
DE
CONTADURÍA
ADMINISTRACIÓN
E
INFORMÁTICA



Octubre 2, 3 y 4 de 2013 ♦ Ciudad Universitaria ♦ México, D.F.



ANFECA
Asociación Nacional de Facultades y
Escuelas de Contaduría y Administración

APLICACIÓN DEL STATISTICAL ARBITRAGE A LA SELECCIÓN DE UN PORTAFOLIO DE PARES BASADO EN COINTEGRACIÓN PARA MERCADOS EMERGENTES

Resumen

El trabajo plantea la aplicación del análisis de Cointegración basado en un modelo de *Statistical Arbitrage Trading* para la identificación de activos que puedan ser empleados en una estrategia de Trading por Pares para Mercados Emergentes, a fin de obtener un Spread con reversión a la media que permita obtener señales de entrada y salida válidas, basadas en intervalos de confianza. Se utilizan dos *ETFs* doblemente inversos (-2x), el *ProShares UltraShort FTSE China 25 (FXP)* el cual es un ETF doble inverso sobre el índice FTSE 25 de China, y el *ProShares UltraShort MSCI Emerging Markets (EEV)* que después de gastos y comisiones ofrece dos veces el inverso aditivo del rendimiento diario del *MSCI Emerging Markets Index*, además se emplea al *iShares MSCI Emerging Markets (EEM)* que replica al 100% el índice *MSCI Emerging Markets Index* como *Benchmark* de la estrategia planteada, considerando datos diarios de Dic-2008 a Abr-2013. Para periodos de 120 días se obtiene un desempeño acumulado de 14.2% con un índice de Sharpe de 3.22; en 252 días la estrategia muestra un desempeño acumulado de 11.2% con un Sharpe de 2.84, favorable con respecto a su *Benchmark* que obtuvo rendimientos acumulados de 0.28% y -1.79% en el mismo periodo respectivamente.

Palabras clave. Statistical Arbitrage, Trading por Pares, Cointegración



APLICACIÓN DEL STATISTICAL ARBITRAGE A LA SELECCIÓN DE UN PORTAFOLIO DE PARES BASADO EN COINTEGRACIÓN PARA MERCADOS EMERGENTES

1. Introducción

Pole (2007) plantea que el término *Statistical Arbitrage* involucra una gran variedad de estrategias las cuales comparten algunas características particulares. Todas ellas, están basadas en i) algún modelo de decisión estadístico que permita generar rendimientos superiores al mercado, ii) buscan tener su libro de *Trades* Neutral al Mercado¹ es decir busca no correlacionarse con el mercado y iii) las señales de trading arrojadas son sistemáticas. Thorp (2008) define al *Statistical Arbitrage* como un “Par” de posiciones que se compensan la una a la otra de forma que logran obtener ganancias seguras al momento de efectuar el *trading*. Dentro de las múltiples estrategias que involucra el *Statistical Arbitrage*, se encuentra el Trading por Pares o *Pair Trading*, la cual según Gatev, Goetzmann y Rouwenhorst (2006), Vidyamurthy (2004) y Pole (2007)² es una de las estrategias de especulación a corto plazo más populares en Wall Street, y se ha convertido en una de las herramientas de trading más empleada por los *Hedge Funds*³ y Bancos de Inversión⁴.

El Trading Por Pares no es nuevo, es una estrategia que data desde 1949 y que además se encuentra estrechamente involucrada con la creación del primer *Hedge Fund*. Según Wolfinger (2005) Alfred Winslow Jones fue el creador del primer *Hedge Fund*, cuya estrategia de inversión se caracterizaba por el empleo intensivo del apalancamiento a través de emplear posiciones accionarias largas y cortas en exactamente los mismos montos, consiguiendo reducir de forma significativa los riesgos de mercado y además obteniendo ganancias substanciales y más o menos constantes del *trade*, siendo el antecedente más antiguo del *Pair Trading*. Thorp (2008) plantea que en 1979, se avocó al diseño de un indicador que se conformara por las acciones de peor y mejor desempeño histórico en un horizonte determinado. La idea central fue ordenar las acciones en función de su rendimiento promedio de dos semanas, con ello se obtuvo que aquellas acciones con mejor desempeño tendían a bajar su rendimiento en las siguientes semanas y aquellas con peor rendimiento tendían a mejorarlo. De esta forma construyó una estrategia denominada *Most Up, Most Down*, que consistía en tomar posiciones largas en aquellas acciones con la mejor perspectiva en su desempeño, y ponerse corto en aquellas con el peor desempeño, obteniendo con ello una estrategia neutral al mercado.

¹ Un Libro de Trades Neutral al Mercado busca obtener beneficios de cualquier tendencia y condición del mercado, tendencia alcista, bajista o movimientos laterales. De forma que cuando exista un movimiento lo suficientemente brusco la estrategia permitirá que los beneficios se coloquen cercanos a cero y de esta forma se otorgue una cobertura al inversionista.

² *Id.*

³ McCrary (2002) y Wolfinger (2005), define a un *Hedge Fund* (Fondo de Cobertura) como un vehículo de inversión colectiva similar en su operatividad a una Sociedad de Inversión Común, pero con la característica de que constituido de forma privada, generalmente manejado por Inversionistas Institucionales, con estrategias caracterizadas por su carácter cuantitativo y flexibilidad para seleccionar estrategias, activos y emplear herramientas que no tienen disponibles los Administradores de Portafolio tradicionales.

⁴ Se refiere a un segmento de la Banca que se orienta a la atención de clientes individuales, corporativos y gubernamentales para la emisión y venta de valores en el Mercado, es decir funge como agente colocador, otorga consultoría en materia de fusiones y adquisiciones y actúa como formador de mercado en operaciones con derivados, deuda, divisas, acciones y commodities.



Gatev, Goetzmann y Rouwenhorst (2006)⁵ plantean que el *Pair Trading* se refinó y popularizó a mediados de los años 80's, cuando el Quant Nunzio Tartaglia conforma 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 trading automatizado que incorporaba la habilidad e intuición de los *traders* 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 *Pair Trading* y de implementarlo a través de la modelación cuantitativa y la ejecución computacional basada en reglas de decisión⁶.

Thomaidis, Kondakis y Dounias (2006), Gatev, Goetzmann y Rouwenhorst (2006)⁷, Vidyamurthy (2004)⁸, Pole (2007)⁹, Alexander y Dimitriu (2002), Alexander, Giblin y Weddington (2002) y Avellaneda y Lee (2008) coinciden en definir al *Pair Trading* como una estrategia que consiste en identificar dos activos con una trayectoria de precios muy similar. Si los precios de los activos se mueven conjuntamente, se asume que el Spread entre estos, su distancia relativa, presenta reversión a la media. Cuando el precio de uno de los activos se incrementa de relativamente con respecto al precio del "par", la estrategia sugiere ponerse corto en el activo sobrevaluado y ponerse largo en el activo subvaluado de forma simultánea. De esta forma el portafolio construido por "pares" mostrará un comportamiento muy distante con respecto al desempeño del mercado, o al menos con respecto a estrategias de inversión tradicionales que consisten en tomar posiciones largas sobre las "ganadoras", por tanto el Trading por Pares generalmente mostrará un desempeño favorable mientras el mercado en general experimenta grandes pérdidas.

El Trading por Pares ha despertado un gran interés en la literatura en recientes años, con lo cual se han generado un sinnúmero de metodologías para la identificación de los pares y la construcción de reglas de decisión del *trading day*. Sin embargo, el presente trabajo plantea la aplicación del análisis de Cointegración basado en un modelo de *Statistical Arbitrage Trading* para la identificación de activos que puedan ser empleados en una estrategia de Trading por Pares para Mercados Emergentes, a fin de obtener un Spread con reversión a la media que proporcione señales de entrada y salida válidas, basadas en intervalos de confianza. Se utilizan dos *ETFs* doblemente inversos (-2x), el *ProShares UltraShort FTSE China 25 (FXP)* el cual es un ETF doble inverso sobre el índice FTSE 25 de China, y el *ProShares UltraShort MSCI Emerging Markets (EEV)*¹⁰ que después de gastos y comisiones ofrece dos veces el inverso aditivo del rendimiento diario del *MSCI Emerging Markets Index*, además se emplea al EEM que replica al 100% del índice *MSCI Emerging*

⁵ *Id.*

⁶ Para ahondar en información adicional sobre las características e historia del *Pair Trading* se sugiere consultar <http://www.pairtradefinder.com/pairTrading.pdf>

⁷ *Loc.Cit.*

⁸ *Loc.Cit.*

⁹ *Loc.Cit.*

¹⁰ El FXP y EEV son gestionados por la empresa ProShares quien se especializa en el diseño de ETF apalancados e inversos y comunes como instrumentos estratégicos de trading para estrategias de inversión long-short para una mayor información sobre sus productos se puede consultar <http://www.proshares.com/>



Markets Index como *Benchmark* de la estrategia planteada, considerando datos diarios de Dic-2008 a Abr-2013. El diseño de la estrategia se encuentra programado en MATLAB y las pruebas econométricas realizadas para la selección de pares se efectuaron con la ayuda de E-Views.

2. Metodología de la Estrategia.

La estrategia de trading planteada estará basada en la metodología propuesta por Avellaneda y Lee (2008) y Thomaidis, Kondakis y Dounias (2006), además de aplicar la técnica de cointegración sobre el spread de los dos activos seleccionados planteadas en los trabajos de Engle y Granger (1987), Burgess (2000), Vidyamurthy (2004) y Gatev, Goetzmann y Rouwenhorst (2006).

Para cada uno de los activos se construyen los Rendimientos Acumulados o base cero para periodos individuales de tiempo:

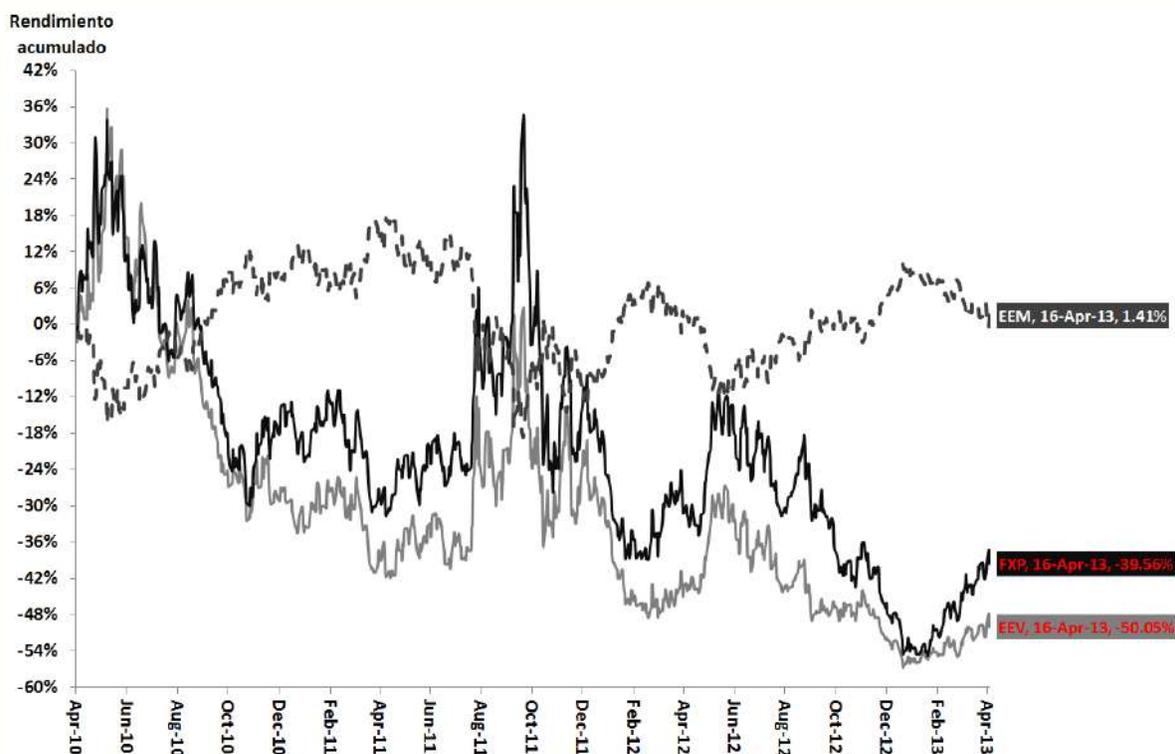
$$1 + R_t = \frac{S_t}{S_{t-1}}, \quad t=1,2,\dots,N \quad (1)$$

$$1 + R_t(k) = (1 + R_t)(1 + R_{t-1})\dots(1 + R_{t-k+1}), \quad t=1,2,\dots,N \quad (2)$$

$$1 + R_t(k) = \prod_{i=1}^{k+1} (1 + R_{t-i,t-k+1}), \quad t=1,2,\dots,N \quad (3)$$

Donde S_t es el precio del activo, R_t es el rendimiento del activo en el periodo t . La Gráfica 1 muestra el Rendimiento Acumulado base cero a tres años de los ETF EEM, FXP, EEV, de esta puede concluirse que de haberse invertido \$100,000 US el 16 de abril de 2010 en cual quiera de estos ETF, al 16 de abril de 2013 se tendrían \$101,410 US. De esta forma es posible identificar una posible relación entre los precios de los activos y de manera gráfica se observa el movimiento conjunto de estos.





Fuente: Elaboración propia con base en información de Yahoo Finance y MSCI

Gráfica 1: Rendimiento Acumulado base cero de tres años de EEV, FXP, EEM (Abr-10 a Abr-2013)

Posteriormente se obtienen los rendimientos logarítmicos conocidos como rendimientos compuestos continuamente si $P_t = \ln(S_t)$ es el precio logarítmico de un ETF y t es pequeña, entonces los Rendimientos Continamente Compuestos son aproximadamente iguales a los Rendimientos Efectivos $\ln(1 + R_t) \approx R_t$.

$$r_t = \ln(1 + R_t) = \ln\left(\frac{S_t}{S_{t-1}}\right) = P_t - P_{t-1}, t=1,2,\dots,N \quad (4)$$

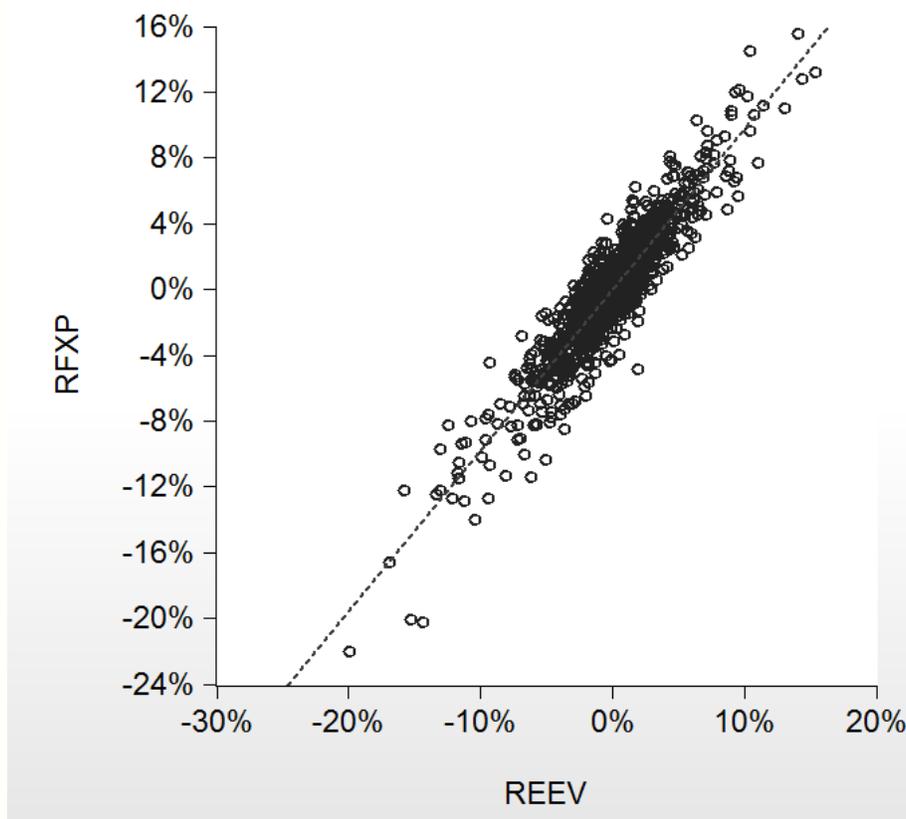
Si se tiene que P_t es la serie de precios del FXP y Q_t es la serie de precios para el EEV, se comienza el análisis planteando el siguiente modelo para intentar encontrar la correlación entre el rendimiento de los activos:

$$\ln\left(\frac{P_t}{P_{t-1}}\right) = \alpha + \beta \ln\left(\frac{Q_t}{Q_{t-1}}\right) + X_t \quad (5)$$



$$\ln\left(\frac{P_t}{P_{t-1}}\right) = -0.0005437 + 0.868583 \ln\left(\frac{Q_t}{Q_{t-1}}\right) + X_t \quad (6)$$

De esta forma es posible apreciar que existe una correlación positiva entre los rendimientos logarítmicos del FXP y el EEV significativa a un 95% de confianza, además esta relación es estrechamente fuerte pues se obtuvo un R^2 de 84.88%, lo cual nos indica en una primera instancia que existe un movimiento conjunto entre el par seleccionado. Como lo muestra la Gráfica 2 es posible apreciar pequeñas desviaciones positivas y negativas con respecto al rendimiento de equilibrio del par.



Gráfica 2: Diagrama de dispersión entre rendimiento logarítmico EEV y FXP

Se efectuó la prueba de cointegración de Engel y Granger (1987) con constante de forma que se logre verificar que existe cointegración entre los pares seleccionados, obteniendo que efectivamente se cumple esta condición con base en ello es posible asumir que existe reversión a la media entre estos dos activos, es decir que en algún periodo del tiempo convergerán sus precios y sus rendimientos. Ver Tabla 1.

Dependiente	tau-statistic	Prob.*	z-statistic	Prob.*
REEV	-35.412	0.000	-1159.886	0.000
RFXP	-34.187	0.000	-1121.580	0.000

Tabla 1: Resultados Prueba de cointegración Engel y Granger (1987).

Una vez corroborado que el par funciona y presenta relaciones de largo plazo entre sus rendimientos acumulados, y que además es posible asumir la existencia de un proceso de reversión a la media, se procede a construir el Spread sobre el cual se fijarán las reglas de trading.

Se asume que existe una relación entre los vectores de precios P_t y Q_t la cual satisface la siguiente ecuación:

$$P_t = \alpha + \beta Q_t + X_t \quad (7)$$

Que en su forma diferencial satisface un proceso dinámico estocástico:

$$dP_t = \alpha dt + \beta dQ_t + dX_t \quad (8)$$

Donde α indica la parte sistemática de la ecuación, X_t son los residuos o la parte idiosincrática de la ecuación definidos como $X_t = P_t - P_t'$, siendo estacionarios y un proceso de reversión a la media por construcción, de forma que puede ser como un indicador que captura las desviaciones temporales del precio relativo entre P_t y Q_t con respecto a su precio de equilibrio, β es la razón de cobertura o *Hedge Ratio*, el cual indica la cantidad de acciones que se deben adquirir de Q_t para cubrir una posición de una unidad en P_t y cuyo signo “+” o “-” indica si se trata de una posición larga o una posición corta. Todos los parámetros de la ecuación fueron calculados a través de una regresión simple mediante Mínimos Cuadrados Ordinarios.

Se efectúa el cálculo de los parámetros para 120 y 252 días, se obtiene que para 252 días hay un *Hedge Ratio* $\beta=1.1683$, que indica que por cada acción que compre del activo P_t compre ese número de acciones de Q_t para cubrir la posición, teniendo posición larga en los dos activos al inicio de la estrategia. Para 120 días se obtiene un *Hedge Ratio* $\beta=1.2010$. Además de esta ecuación se extrae la serie de residuos X_t , que se empleará para formar las reglas de trading.

Además el modelo asume que la fluctuación de los precios del par, definido como un Spread entre el precio relativo de P_t y Q_t puede ser estimado fácilmente a través de un modelo paramétrico para X_t , a través de un proceso Ornstein- Uhlenbeck con la siguiente forma:

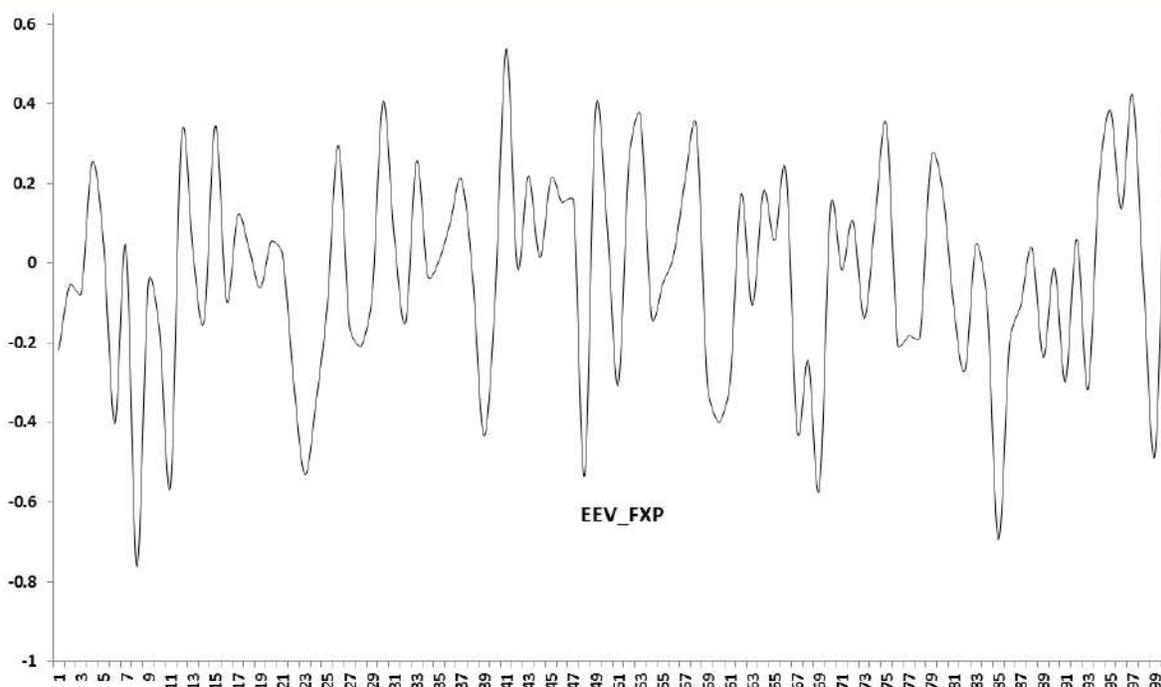
$$dX_t = \kappa(X_{t-1} - \mu)dt + dW_t, \kappa > 0 \quad (9)$$



En donde $dX_t = X_t - X_{t-1}$, y el proceso resulta estacionario y autorregresivo de orden 1, es decir es calibrado a través de un modelo AR-1 y μ es obtenido a través de la media aritmética de X_t :

$$\mu = \frac{1}{N} \sum_{t=1}^N X_{t,t=1,2,\dots,N} \quad (10)$$

En particular $dX_t = X_t - X_{t-1}$ posee una media no condicional de cero y una media condicional igual a $E\{dX_t/X_t\} = \kappa(X_{t-1} - \mu)dt$, y donde los parámetros de la ecuación diferencial estocástica varían en función de los incrementos en el término de perturbación estocástico dW_t . Todos los parámetros fueron calculados mediante Mínimos Cuadrados Ordinarios.



Gráfica 3: Simulación a 100 pasos de Proceso O-U FXP v.s. EEV

A través del proceso de reversión a la media es posible el cálculo de un indicador medio de rebalanceo del portafolio, que estima el promedio de días en que el *Spread* X_t retorna a su valor de equilibrio y es necesario cambiar el signo de las posiciones, para ello se emplea la fórmula $c = \frac{-\log(2)}{\kappa}$, obteniendo que para la ventana de 252 días, el portafolio de pares presenta reversión a la media cada 31 días en promedio, mientras que para la ventana de 120 día el *Spread* presenta reversión a la media cada 20 días.

A fin de corroborar que X_t sea un proceso estacionario que sigue un proceso Ornstein Uhlenbeck como el modelado anteriormente, se emplea la prueba Dickey-Fuller

Aumentada, la cual permite conocer cuando un proceso estocástico es estacionario en función de las raíces de la ecuación característica la cual se denota como:

$$\Delta X_t = \sum_{i=1}^p \rho \Delta X_{t-i} + u_t \quad (11)$$

En donde la prueba de hipótesis que se plantea es :

$$\begin{aligned} H_0: \rho &= 0 \\ H_1: \rho &< 0 \end{aligned} \quad (12)$$

Si $\rho = 0$ entonces la serie es no estacionaria y presenta raíces unitarias y por lo tanto no seguiría un proceso de reversión a la media, en consecuencia el *Spread* formado no sería apto para el *Trading* y se deberían cambiar los activos que lo formaron, efectuada la prueba se obtiene que el *Spread* formado no presenta raíz unitaria y por lo tanto es estacionario. Ver Tabla 2.

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.888075	0.0000
Test critical values:	1% level	-2.567145	
	5% level	-1.941122	
	10% level	-1.616498	
	level		

Tabla 2: Prueba Dickey-Fuller Aumentada para el *Spread*

Enfocados en el proceso X_t es posible establecer intervalos de confianza que permitan la construcción de la reglas de trading que se incorporarán al sistema, el intervalo superior está definido como $Z_t = \mu + n\sigma$ y el intervalo inferior está definido como $Z_t = \mu - n\sigma$, donde n es un intervalo entre 1 y 0 y σ es la desviación típica de los residuos denotada como:

$$\sigma = \sqrt{\frac{1}{N} \sum_{t=1}^N (X_t - \mu)^2} \quad (13)$$



De esta forma las reglas de trading quedan definidas como :

Comprar el Spread si: $X_t < Z_t$

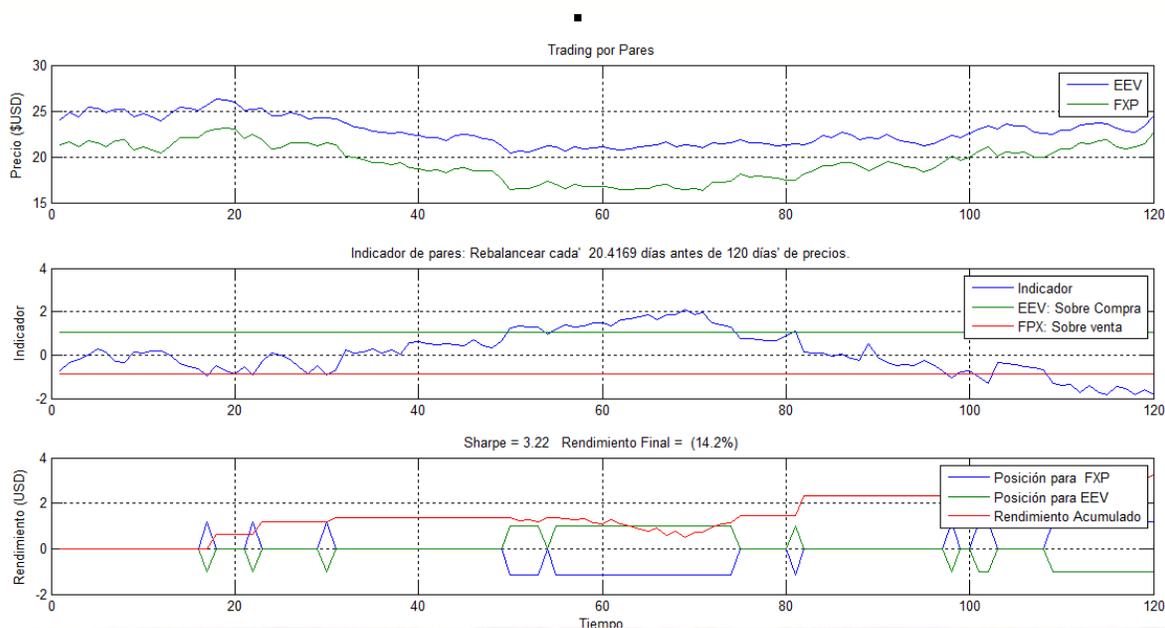
Vender el Spread si: $X_t > Z_t$

Este es el criterio estándar para reglas de *trade* adoptado por la mayoría de los estudios, el intervalo permanece constante dentro de la ventana de *trade* que sea escogida.

3. Resultados de la Estrategia.

Se plantea que en la ventana de 120 días el portafolio de pares está formado por $Z_t = P_t - 1.2010Q_t - X_t$, es decir inicialmente compro el *Spread* tres veces, me encuentro largo en FXP y corto en EEV por 1.20 acciones, aproximadamente cada 20 días existe un rebalanceo de las posiciones, para el día 40 se aprecia que mi posición pasó a ser neutral por las condiciones de mercado y el fondo sólo obtuvo Cash en su interior, para el día 60 existió un rebalanceo y ahora me encuentro largo 1.20 acciones en EEV y corto 1 acción en FXP, para el día 100 se vuelve a la posición inicial.

La estrategia planteada en 120 días obtuvo un desempeño acumulado del 14.2% neto, quitando costos de transacción por trade del 0.05%, y un índice de Sharpe de 3.22 lo cual indica que la estrategia es redituable en el corto plazo. Por otro lado el rendimiento acumulado del Benchmark para 120 días fue de 0.28% .Ver Gráfica 4.



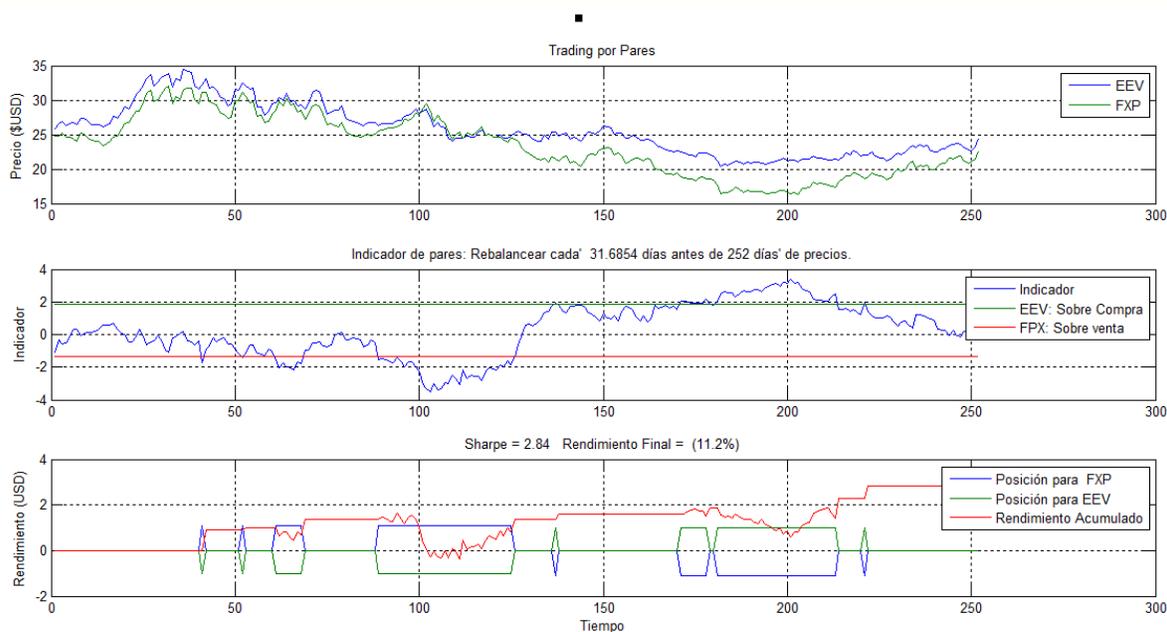
Gráfica 4: Estrategia Pair Trading 120 días

Se plantea que en la ventana de 252 días el portafolio de pares está formado por $Z_t = P_t - 1.1683Q_t - X_t$, es decir inicialmente compro mi *Spread* durante 120 días al



menos, me encuentro largo en FXP y corto en EEV por 1.16 acciones, aproximadamente cada 31 días existe un rebalanceo de las posiciones, para el día 130 se aprecia que mi posición pasó a ser neutral por las condiciones de mercado y el fondo sólo obtuvo Cash en su interior, para el día 160 existió un rebalanceo y ahora me encuentro largo 1.16 acciones en EEV y corto 1 acción en FXP, para el día 220 la posición se vuelve neutral.

La estrategia planteada en 252 días obtuvo un desempeño acumulado del 11.2% neto, quitando costos de transacción por *trade* del 0.05%, y un índice de Sharpe de 2.84 lo cual indica que la estrategia es redituable en un año de *trade*. Mientras que el rendimiento del EEM propuesto como *Benchmark* para ese periodo fue de -1.79%. Ver Gráfica 5.



Gráfica 5: Estrategia de Trading 252 días.

4. Conclusiones.

La estrategia propuesta resulta redituable en el corto plazo, y permite ser implementada como una estrategia de gestión activa no correlacionada con el mercado. Para futuros trabajos se plantea la inclusión de Inteligencia Artificial para la ejecución de las señales de Trading.

La elaboración de un intervalo de Trading dinámico, así como la estimación de un *Hedge Ratio* dinámico a través de Filtros de Kalman,

Finalmente se considera el empleo de datos de alta frecuencia y un parámetro de estimación de efectivo en el portafolio y los picos de efectivo que va dejando el *Hedge Ratio* con cada operación.



5. Bibliografía

Alexander, C & Dimitriu, A. (2002). Cointegration-based trading strategies: A new approach to enhanced index tracking and statistical arbitrage. *Discussion Paper 2002-08, ISMA Centre Discussion Papers in Finance Series.*

Alexander, C, Giblin, I, & Weddington, W.(2002). Cointegration and asset allocation: A new active hedge fund strategy. *Discussion Paper 2003-08, ISMA Centre Discussion Papers in Finance Series.*

Avellaneda, M., & Lee, J., (2008). Statistical Arbitrage in the U.S. Equities Market. *Quantitative Finance*, 10: 61-782.

Burgess, N. (2000). Statistical arbitrage models of the FTSE 100. Abu-Mostafa, Y., LeBaron, B., Lo, A. W., and Weigend, A. S., (eds.). *Computational Finance* (1999). The MIT Press. 297–312.

Engle, R. F. & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*. 55:251–276.

Gatev, E., Goetzmann, W., & Rouwenhorst, K. (2006). Pairs trading: performance of a relative-value arbitrage rule. *The Review of Financial Studies*, 19(3):797–827.

Thomaidis, N. S., Kondakis, N., & Dounias, G. (2006). An intelligent statistical arbitrage trading system. *Lecture Notes in Artificial Intelligence*. 3955:596–599.

Pole, A. (2007). *Statistical arbitrage: algorithmic trading insights and techniques*. New Jersey: Wiley.

Thorp, E., (1967). *Beat the Market: A scientific Stock Market System*. Random House. New York: Random House.

Thorp, E., (2008). Statistical Arbitrage – Part I. *Wilmott Magazine*. http://www.wilmott.com/pdfs/080617_thorp.pdf

Vidyamurthy, G. (2004). *Pairs trading: quantitative methods and analysis*. New Jersey: Wiley.

Wolfinger, D. (2005). *Create Your Own Hedge Fund: Increase Profits and Reduce Risks with ETFs and Options*. New Jersey: Wiley.

