



UNIVERSIDAD CENTRAL DE VENEZUELA

FACULTAD DE CIENCIAS ECONÓMICAS Y SOCIALES
ESCUELA DE ESTADÍSTICA Y CIENCIAS ACTUARIALES

EVALUACIÓN DE LA EFICACIA DE UNA
ESTRATEGIA DE TRADING BASADA EN MODELO
LINEAL GENERALIZADO

Trabajo Final de Pregrado

PARA OPTAR POR EL TÍTULO DE:

Licenciado en Ciencias Actuariales

Rodrigo Alejandro Serrano Morales

TUTOR:

Prof. Jonattan Ramos & Prof. Eloy Eligon



Facultad de Ciencias
Económicas y Sociales

Caracas, Marzo 2019

*Dedicado a mis padres y a mi hermana, por ser los pilares de mi vida y ser mi motivo de
haber cumplido esta meta.*

Agradecimientos

Agradecimientos

Gracias

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	4
ÍNDICE DE CUADROS	5
Introduccion	6
1. El Problema	7
1.1. Justificación	7
1.2. Planteamiento del Problema	7
1.3. Objetivo General	7
1.4. Objetivos Específicos	7
2. Marco Teórico	8
2.1. Antecedentes	8
2.1.1. Trading de Cryptomonedas basado en Aprendizaje Automatico	8
2.1.2. Modelos predictivos para el mercado FOREX	8
2.1.3. Diseño e implementación de un sistema automatizado para operar en el mercado de divisas usando reglas de asociación	8
2.2. Bases Teóricas	8
2.2.1. Hipótesis del Mercado Eficiente	8
2.2.2. Análisis Técnico	9
2.2.3. Reglas de asociación	9
2.2.4. Introducción al aprendizaje automático	9
2.2.5. Aprendizaje Supervisado y No Supervisado	9
2.2.6. Métodos de Aprendizaje Supervisado Basado en GLM	9
2.2.7. Análisis de Componentes Principales como método de reducción de variables	9
2.2.8. Regresión Logística	9
2.2.9. Curva ROC	9
2.2.10. Función de Pérdida	9
2.3. Bases Legales	9
3. Marco Metódico	10
3.1. Análisis Exploratorio de los datos	10
3.1.1. Datos OHLC y Fuente de los datos	10
3.1.2. Series de Índices	10

3.2. Entrenamiento del Modelo	11
3.2.1. Indicadores Técnicos como variables predictoras	11
3.2.2. Variable dependiente	12
3.2.3. Validación Cruzada en Series de Tiempo	12
3.2.4. WalkForward Backtesting	12
3.2.5. Reducción de la dimensión con Análisis de Componentes Principales . . .	13
3.2.6. Evaluación del desempeño del modelo	13
4. Análisis de Resultados	15
4.1. Series de Rendimientos	15
Conclusiones y Recomendaciones	16
Lista de Referencias	17

ÍNDICE DE FIGURAS

3.1. Precios de Cierre de los índices en el período de estudio (26/10/2008 - 18/01/2019)	11
3.2. Metodología WalkForward	13
3.3. Matriz de Confusión	14

ÍNDICE DE CUADROS

Introducción

Las economías de países latinoamericanos son reconocidas históricamente por depender en gran magnitud del comercio de sus materias primas, lo cual hace que el comercio con dichos commodities resulte de gran impacto para el gasto fiscal y para la balanza de pagos, esto deja como consecuencia, la necesidad de los actores económicos de estudiar el riesgo a profundidad para poder evitar resultados que reduzcan sus retornos positivos.

CAPÍTULO 1

El Problema

1.1. Justificación

El mercado de commodities ha ido evolucionando paralelamente al de otros mercados financieros y actualmente es un mercado globalizado, como el que podemos encontrarnos en los mercados de divisas, de renta variable o de renta fija. Este desarrollo ha permitido la entrada de muchos participantes en el mercado y la implementación de diversos productos e instrumentos de inversión.

1.2. Planteamiento del Problema

1.3. Objetivo General

Evaluar la eficacia de una estrategia de trading basada en técnicas de aprendizaje automático en diferentes instrumentos financieros

1.4. Objetivos Específicos

- Definir los indicadores técnicos a utilizar como variables predictoras
- Utilizar Análisis de Componentes Principales como técnica de reducción de la dimensión de variables predictoras
- Aplicar modelo de Regresión Logística con las componentes arrojadas por el ACP
- Definir la mejor combinación de parametros (Take Profit, Stop Loss y Horizonte) para cada instrumento de inversión según la probabilidad positiva arrojada por el modelo
- Aplicar el modelo en la data de Validación y analizar resultados
- Aplicar Montecarlo a la variable ganancia

CAPÍTULO 2

Marco Teórico

2.1. Antecedentes

2.1.1. Trading de Cryptomonedas basado en Aprendizaje Automatico

2.1.2. Modelos predictivos para el mercado FOREX

2.1.3. Diseño e implementación de un sistema automatizado para operar en el mercado de divisas usando reglas de asociación

2.2. Bases Teóricas

2.2.1. Hipótesis del Mercado Eficiente

La Hipótesis del Mercado eficiente fue desarrollada por Eugene Fama en los años 60, en la misma argumenta que los precios de los activos reflejan toda la información disponible, es decir que siempre son transados a un valor adecuado para su riesgo, haciendo imposible para los inversores obtener retornos más elevados que los del mercado en general.

Fama sugiere tres suposiciones. Primero, el mercado eficiente requiere un gran número de competidores buscando maximizar ganancias. Segundo, la información que afecta al activo llega al mercado de manera aleatoria y cada anuncio es independiente de los demás. Tercero, todos los competidores intentarán ajustar sus posiciones lo más rapido posible conocida la información del mercado. Existen tres variantes de la hipótesis:

Eficienciadbil, en esta variante, los precios del pasado no sirven para predecir el precio futuro, es decir cualquier movimiento del activo es determinado por información no contenida en la serie de precios. *Eficienciamedia*, en esta forma se asume que los precios se ajustan instantáneamente a la información pública, por lo que rechaza cualquier tipo de arbitraje intentando aprovechar nueva información. *Eficienciafuerte*, esta última forma de la hipótesis plantea que los precios reflejan tanto información pública como privada, por lo cual incluso obteniendo información no conocida por todos los competidores, no se pueden obtener retornos anormales a los de los mercados.

Aunque esta hipótesis es la piedra angular de la teoría financiera moderna, es controversial entre la comunidad financiera y disputada frecuentemente. Algunas de las razones que exponen sus detractores son:

Los inversores ven la información de manera distinta, por lo que generarán diferentes valuaciones de un mismo activo. Los precios del activo pueden ser afectados por el error humano y decisiones emocionales. Se ha evidenciado en la historia, casos de inversores que han logrado vencer el mercado por largos períodos de tiempo, como Warren Buffet, lo cual por definición de la hipótesis es imposible.

Es evidente la postura que se asume en la presente investigación con respecto a la hipótesis de mercado eficiente. Los avances tecnológicos y la capacidad de procesamiento de las computadoras en la actualidad hacen pensar que cualquier anomalía presente en el mercado por muy pequeña que sea puede ser aprovechada por sofisticados softwares automatizados.

2.2.2. Análisis Técnico

Los inversionistas que rechazan la hipótesis del mercado eficiente buscan interpretar la situación del mercado, bien a través de noticias que afecten al activo o estudiando su movimiento intentando extraer patrones de conducta. A la primera técnica se le llama *AnlisisFundamental* y el segundo *AnlisisTcnico*. El Análisis Fundamental está mas asociado a estrategias de inversión pasivas a largo plazo aunque en la actualidad se han desarrollado algoritmos de compra y venta que buscan predecir la dirección del precio en función de noticas utilizando programación Minería de texto.

El análisis técnico es aquel que busca patrones y tendencias de comportamiento en la cotización de los activos financieros, basándose en la serie de tiempo del activo, con esto intenta predecir el movimiento futuro buscando maximizar las ganancias.

2.2.3. Reglas de asociación

2.2.4. Introducción al aprendizaje automático

Aprendizaje automático refiere a un amplio abanico de herramientas para el entendimiento de los datos. Estas herramientas pueden ser clasificadas en supervisadas o No Supervisadas

2.2.5. Aprendizaje Supervisado y No Supervisado

2.2.6. Métodos de Aprendizaje Supervisado Basado en GLM

2.2.7. Análisis de Componentes Principales como método de reducción de variables

2.2.8. Regresión Logística

2.2.9. Curva ROC

2.2.10. Función de Pérdida

2.3. Bases Legales

CAPÍTULO 3

Marco Metódico

3.1. Análisis Exploratorio de los datos

3.1.1. Datos OHLC y Fuente de los datos

La estructura de los datos utilizados en el trabajo es de tipo OHLC por sus siglas en inglés Open, High, Low, Close. La misma, resume en 4 registros el comportamiento del precio del activo (Apertura, Cierre, Mínimo y Máximo) en un intervalo de tiempo. En el caso de la presente investigación, de un día. Este tipo de dato provee la información necesaria para cubrir las exigencia del modelo, tanto para la creación de la variable dependiente como para el cálculo de los indicadores técnicos.

Los datos fueron extraídos del portal www.investing.com, uno de los portales financieros con mayor prestigio en el mundo. Fue fundado en 2007 y es conocido por su prestigioso calendario económico y directorio de brokers.

3.1.2. Series de Índices

El universo de estudio está representado por los índices bursátiles de los mercados financieros existentes entre el período 26/10/2008 - 18/01/2019. Un índice bursátil es un promedio de los precios de los activos que representan un mercado o sector determinado. Los mismos sirven como 'benchmark' o referencia de la economía de un país, sector financiero, etc. En el ámbito de los 'hedge funds' son una referencia para medir la rentabilidad de una estrategia de inversión y el riesgo del mercado.

En la presente investigación se utilizan los índices como reflejo del comportamiento de varios activos, de esta manera, se mide la estrategia en un sector y no en un instrumento en específico. Otras de las ventajas de utilizar los índices es que al representar un promedio de varios activos, sus variaciones son menos drásticas. La muestra está constituida por 5 índices bursátiles que representan distintos mercados del mundo: NASDAQ, NIKKEI, FTSE 100, BOVESPA y SP500.

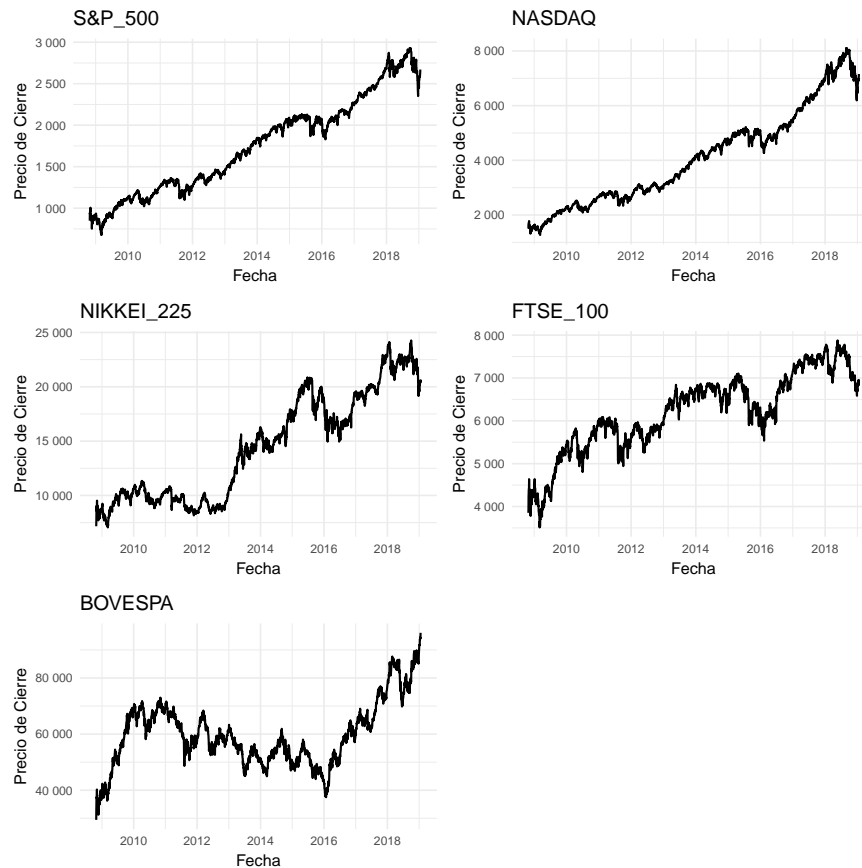


Figura 3.1: Precios de Cierre de los índices en el período de estudio (26/10/2008 - 18/01/2019)

3.2. Entrenamiento del Modelo

3.2.1. Indicadores Técnicos como variables predictoras

Los indicadores a utilizar fueron seleccionados buscando recoger la mayor información posible sobre el precio del activo que se pueden resumir en tres categorías: tendencia, momentum y volatilidad.

No es de interés en la presente investigación describir como funciona cada indicador para la toma de decisiones en el trading basado en fundamentos técnicos. Cada indicador puede utilizarse de distintas maneras, calcularse con distintos parámetros y asociarse a discreción del trader, lo que conlleva a un sin fin de reglas de asociación.

Lo que busca la investigación es utilizar la relación entre estos indicadores como variables independientes que ayuden al modelo a predecir oportunidades de entradas. En este sentido se asume la existencia de una dinámica local del mercado que puede ser predecida con ayuda de estos indicadores.

3.2.2. Variable dependiente

Las decisiones de entrada en el trading pueden ser producto de muchos factores, en la presente investigación se analiza el enfoque donde se define un porcentaje objetivo de ganancia y se intenta predecir si dicho objetivo se materializará en un futuro cercano, sin que se haya concretado una venta por Stop Loss. Este enfoque reduce la toma de decisión en una variable tal que:

$$P_X(x) = \begin{array}{ll} p ; & x = c \\ 1 - p ; & x = -d \end{array}$$

Dado los datos OHLC del activo es posible identificar los períodos en donde se materializa la variable dependiente. la identificación se realiza, comparando el precio de cierre con los precios máximos y mínimos de las siguientes h observaciones, donde h es el número de períodos, en este caso días en los cuales se desea evaluar la condición.

En la práctica se identifica los registros que cumplen con esta condición añadiendo una columna a la data donde incluimos 'buy' para identificar los registros donde se da la señal y 'stay' en caso de que no haya ocurrido o hubiese ocurrido primero el retroceso del precio.

3.2.3. Validación Cruzada en Series de Tiempo

La validación Cruzada es un método de validación y prueba que consiste en dividir los registros aleatoriamente en grupos de similar tamaño. El primer grupo es utilizado como validación del modelo que ha sido entrenado en el resto de los datos, este proceso se realiza k veces, y el resultado final es el promedio arrojado por cada una de las k validaciones.

Ahora bien este método asume que no existe relación entre las observaciones, es decir que son independientes. Esto no es verdad en el caso de las series de tiempo debido a la condición de autoregresión. Por lo tanto al dividir la data se debe respetar el orden temporal de cada observación.

3.2.4. WalkForward Backtesting

Al principio de la investigación se implementó el método de entrenamiento, validación y prueba comúnmente utilizado, en donde la mayor parte de la data es destinada a entrenamiento del modelo, otra sección es destinada a validación, para elegir los parámetros óptimos, y finalmente se testeaba el modelo en la data de prueba. Sin embargo este tipo de metodología en opinión del investigador no es el más óptimo para desarrollar el presente modelo, dado el dinamismo de los mercados bursátiles la estrategia no puede permanecer estática en el tiempo.

Para contrarrestar esta situación se optó por el método de *backtestingWalkforward*, el cual consiste en entrenar el modelo en un período base de data, en este caso los primeros 4 años de estudio, posteriormente se aplica la estrategia en el año siguiente y se obtiene los primeros resultados. Luego este año de aplicación es incluido en la data de entrenamiento -es decir, la data de entrenamiento pasa a ser de 5 años- y se evalúa el modelo en el siguiente año. De esta manera, contemplamos el dinamismo del mercado permitiéndole al modelo -y por ende a

la estrategia- utilizar el período mas reciente con respecto al cual será implementado. En la presente figura se ilustra la metodología implementada.

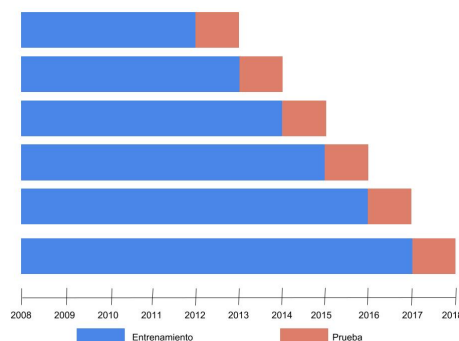


Figura 3.2: Metodología WalkForward

Otra de las características de la metodología que se modificó fue la elección de los parámetros óptimos. Previamente se utilizaba la data de validación para buscar la combinación de parámetros óptima. Ahora bien en la metodología de Walkforward se utilizan los mismo parámetros. A opinión del investigador al buscar los mejores parámetros se estaría incurriendo en un posible sesgo de sobreoptimización. El hecho de que en un año determinado unas configuraciones óptimas den los mejores resultados no asegura que se replique en el siguiente año.

La selección de los parámetros debería ser un estudio previo de la serie financiera a testear. En caso de la estrategia actual toma un horizonte de 20 períodos ya que este valor representa un umbral para la ocurrencia del objetivo. Evidentemente al seleccionar un período de tiempo mas corto se obtendrán menos observaciones que cumplan con el patrón por lo tanto estaremos en presencia de un problema de data imbalanceada que debe tener un tratamiento distinto. Por otro lado el utilizar un horizonte mayor no representa gran cambio en el número de ocurrencias, pero sí en el caso de que la transacción quede abierta -hay recordar que h representa una condición de salida para la estrategia, en caso de no ocurrir el target ni el stop loss-. Por lo que se decidió 20 como el número de períodos óptimos a utilizar.

Por su lado el porcentaje de profit se decidió en base a los retornos

3.2.5. Reducción de la dimensión con Análisis de Componentes Principales

3.2.6. Evaluación del desempeño del modelo

En los problemas de clasificación se utiliza la matriz de confusión para evaluar el desempeño del modelo. La misma es una tabla que categoriza las predicciones realizadas por el modelo de acuerdo a la coincidencia con los valores reales.

La estrategia solo toma la señal cuando el modelo predice un incremento en el precio, la venta por el contrario no depende del modelo, sino de los parámetros predefinidos (porcentaje de Stop Loss y Horizonte de tiempo). Esta característica implica que el valor a maximizar es la predicción de los verdaderos positivos, conocido como *Precisin*.

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 3.3: Matriz de Confusión

$$Precisin = \frac{VP}{VP + FN}$$

CAPÍTULO 4

Análisis de Resultados

En el presente capítulo se realiza la descripción de los resultados obtenidos despues de la aplicación del método propuesto para la estrategia. De igual modo, se presentan los resultados arrojados por las pruebas de Backtesting simulando las entradas y salidas.

4.1. Series de Rendimientos

A partir de las series de precios correspondientes a los commodities Petróleo, Oro, Cacao, Harina de Soja y Aluminio; comprendidas entre el periodo (10/2010–10/2018) se calcularon las series de rendimientos logaritmicos para cada uno de los activos. Cabe destacar que se disponen de 1679 observaciones de rendimientos diarios para cada commodity.

```
> list_serie[[1]] %>% head()

# A tibble: 6 x 5
  timestamp close open high low
  <date>      <dbl> <dbl> <dbl> <dbl>
1 2008-10-27  849.  874.  894.  847.
2 2008-10-28  941.  849.  941.  845.
3 2008-10-29  930.  940.  970.  922.
4 2008-10-30  954.  939.  963.  928.
5 2008-10-31  969.  953.  984.  945.
6 2008-11-03  966.  969.  976.  959.
```

Conclusiones y Recomendaciones

El estudio presentado se centró en confirmar alternativas eficaces para el cálculo del VaR en mercados de materias primas. Los hechos estilizados de las series de tiempo financieras fueron considerados a lo largo del trabajo. Y frente a estos planteamientos, se desarrolló una investigación que permitiera detectar una aproximación con el método GARCH-EVT-COPULAS que tuviese la capacidad de estimar el riesgo del mercado de commodities.

Lista de Referencias

- Alexander, N. Y Zafer, D. (2016). *Gestión del riesgo con el uso de commodities energéticos utilizando Valor en riesgo y Teoría del Valor Extremo*. Suecia, Universidad de Lund.