

INSTITUTO TECNOLÓGICO DE BUENOS AIRES – ITBA

ESCUELA DE INGENIERÍA Y GESTIÓN

Sistema automático para compraventa de activos financieros

AUTORES: Mounier, Agustin (Leg. N° 54037)

Suárez Bodean, Joaquín (Leg. N° 54361)

TUTOR: Parisi, Daniel Ricardo

TRABAJO FINAL PRESENTADO PARA LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN INFORMÁTICA

Lugar: Buenos Aires, Argentina
Fecha: 28/07/2018

Índice

1. Introducción al mercado de valores	3
1.1 Hipótesis del mercado eficiente	3
1.2 Operaciones y conceptos básicos del mercado accionario	4
1.3 Indicadores de rendimiento	6
1.3.a Sharpe ratio	6
1.3.b Beta	7
1.3.c Alpha	7
2. Comercio de alta frecuencia	9
2.1 Backtesting	9
2.1.a Survivorship bias	9
2.1.b Look Ahead Bias	10
2.1.c Data Snooping Bias	10
2.1.d Standard & Poor's Index	10
2.2 Mecanismo de recolección de datos	11
2.2.a Google Finance	11
2.2.b Yahoo Finance	11
2.2.c Alpha Vantage	12
2.2.d Intrinio	12
2.2.e Módulo automático de recolección de datos	12
2.3 Quantopian	13
2.3.a Limitaciones	14
3. Estrategias	15
3.1 Convergencia y Divergencia por Media Móvil (Moving Average Convergence Divergence o MACD)	15
3.2 ARMA	17
3.3 Máquina de soporte vectorial	19
3.3.a Datos y Modelo	21
3.3.b Vectores de input/output	21
4. Resultados	24
4.1 MACD	24
4.2 ARMA	26
4.3 Máquina de soporte vectorial	27
5. Análisis de los resultados	29
5.1 MACD	30
5.2 ARMA	32

5.3 Máquina de soporte vectorial	32
6. Conclusiones	36
7. Referencias	37

1. Introducción al mercado de valores

Un mercado de valores, es una organización privada que facilita a sus miembros la compra y venta de valores como: acciones de sociedades o compañías, bonos públicos o privados, certificados y una amplia variedad de instrumentos de inversión. Realiza esta actividad atendiendo los mandatos de sus clientes a la hora de introducir órdenes de compra o venta y realizando las negociaciones pertinentes.

La negociación de los valores en los mercados de valores, se hace tomando como base precios conocidos y fijados en tiempo real, en un entorno seguro para la actividad de los inversionistas, donde el mecanismo de las transacciones está totalmente regulado, lo que garantiza la legalidad y seguridad.

La negociación de valores en los mercados (o bolsas) se efectúa a través de sus miembros, conocidos usualmente con el nombre de corredores (o *brokers* en inglés) quienes realizan este trabajo a cambio de una comisión.

En este trabajo nos centraremos en el comercio de acciones, uno de los valores más conocidos y comunes con los que opera el mercado. Probablemente esté familiarizado con la definición popular de acción como: “Una acción es una parte de la empresa y al ser de su propiedad tiene derecho a reclamar parte de las ganancias de la misma.”. Desafortunadamente, esta definición es incorrecta en algunos aspectos.

Una acción brinda al tenedor el derecho de votar en reuniones de accionistas, recibir dividendos (que son las ganancias de la empresa) si es que son distribuidos, y vender sus acciones a otra persona. Si se poseen la mayor parte de las acciones el poder de voto incrementa lo que permite gobernar la compañía de forma indirecta al poder nombrar personas que integran la junta directiva. Esta actividad se ve reflejada cuando una compañía compra a otra, la compra de una compañía no significa comprar los edificios, computadoras y empleados que la integran, sino comprar la mayor cantidad de acciones. Al poseer la mayor cantidad de acciones y por ende, tener mayor participación en la junta directiva, se pueden designar profesionales (como CEO, CTO y otros altos directivos de la empresa) que optimicen o cambien el rumbo de la misma.

Para el resto de los accionistas y comerciantes del mercado accionario, no tener las riendas en lo que respecta a la dirección de la compañía no es un problema ya que lo que se busca es el aumento del valor de la acción y/o percibir de sus dividendos. En nuestro caso y para el desarrollo de este trabajo formaremos parte de este último grupo buscando predecir el valor futuro de diversas acciones.

1.1 Hipótesis del mercado eficiente

La hipótesis del mercado eficiente nos dice que es imposible “ganarle al mercado” ya que debido a su eficiencia los precios de los valores que comercia siempre incorporan y reflejan

toda la información posible. De acuerdo con esta hipótesis [1] las acciones siempre se comercian a su valor justo en los mercados, por lo que sería imposible que un inversor compre acciones infravaloradas o venda acciones sobrevaloradas. Esto hace imposible para un inversor obtener un rendimiento (constante en el tiempo) mejor que el que posee el mercado por más experiencia que posea. La única forma de obtener un rendimiento mejor es a partir de la obtención de información privilegiada o mera suerte. Es por esta razón que los precios de las acciones siguen una trayectoria incierta lo que resulta en que la mejor predicción posible sobre el valor futuro de una acción es el valor actual.

Como se podran imaginar, en el presente trabajo intentaremos refutar esta hipótesis y mostrar que es posible ganarle al mercado y predecir de forma constante los futuros movimientos de las acciones. Cabe destacar que esta hipótesis recibió múltiples críticas ya que asume, como su nombre lo indica, que el mercado y toda su información se comporta de forma totalmente eficiente negando la posibilidad de una lenta difusión de datos, o la presencia de mercados dominantes.

1.2 Operaciones y conceptos básicos del mercado accionario

A continuación se definen una serie de conceptos y operaciones que consideramos básicos y necesarios para el entendimiento del presente trabajo.

- **Precio:** el precio de una acción se determina en función a la negociación en el mercado. La lucha entre el valor que los compradores estén dispuestos a pagar y los vendedores a vender da origen al valor de cotización. En este trabajo denotaremos el precio de una acción en un instante t como p_t .
- **Retorno:** ganancia o pérdida de un valor dado en un período determinado. Si se toma el precio p_t y Δt como el período para calcular el retorno, se define el retorno para el período mencionado como:

$$r_{t+\Delta t} = \frac{p_{t+\Delta t} - p_t}{p_t}$$

- **Retorno logarítmico:** representa lo mismo que el retorno anterior agregando algunas ventajas aritméticas. Basándonos en la definición anterior definimos el retorno logarítmico como:

$$R_{t+\Delta t} = \log(1 + r_{t+\Delta t}) = \log\left(\frac{p_{t+\Delta t}}{p_t}\right)$$

Al calcular el retorno de esta manera poseemos tres ventajas fundamentales. La primera es la normalidad de los valores, si asumimos que los precios poseen una distribución logarítmica entonces $\log\left(\frac{p_{t+\Delta t}}{p_t}\right)$ posee una distribución normal, lo cual es útil para funciones estadísticas que asumen normalidad. En segundo lugar, sabemos que para valores pequeños de retornos (comunes en operaciones de corta duración) el retorno logarítmico será similar al retorno real ya que $\log(1 + r_i) \approx r_i$, $r_i \ll 1$. Por último, los retornos

logarítmicos son aditivos gracias a las propiedades de los logaritmos. Supongamos una serie de retornos consecutivos de n operaciones entonces:

$$\sum_{i=0}^n \log(1 + r_i) = \log(1 + r_0) + \log(1 + r_1) + \dots + \log(1 + r_n) = \log(p_n) - \log(p_0)$$

- **Volatilidad:** representa la dispersión de los precios de las acciones en un periodo de tiempo determinado. Se puede medir calculando el desvío estándar de los precios de la acción. Para un periodo de N instantes de tiempo se calcula como:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - \mu)^2}, \text{ con } \mu = \frac{1}{N} \sum_{i=1}^N p_i$$

- **Short order:** transacción en la que se vende una acción esperando poder comprarlo posteriormente a un precio inferior. Es posible realizar esta operación sin poseer acciones de la empresa en cuestión. De ser éste el caso el broker actuara como prestamista vendiendo acciones que él posee y entregando el dinero a su cliente. Aquí se genera un contrato donde el cliente se compromete a comprarle al broker la misma cantidad de acciones que éste vendió independientemente de si la acción aumentó o bajo en valor. El broker por otro lado, cobra intereses sobre la suma de dinero entregada.
- **Long order:** transacción en la que se decide comprar una acción esperando poder venderla a un precio superior en el futuro.
- **Slippage:** diferencia del precio esperado de una transacción con el precio efectivo de la transacción cuando se ejecuta.
- **Limit orders:** orden que se ejecuta cuando el precio de la acción alcanza o supera cierto valor. También conocida como *Take-profit order*. Puede no ser ejecutada.
- **Stop order:** orden que vende una determinada cantidad de acciones cuando el precio de la acción cae por debajo de cierto valor si la orden inicial fue *long*. En caso de que la orden inicial fuese *short*, compra una determinada cantidad de acciones si esta sobrepasa cierto valor. También conocida como *Stop loss order*. Puede no ser ejecutada.
- **Bid-ask spread:** diferencia entre el precio solicitado por el vendedor y el precio ofrecido por el comprador para un determinado valor. Denota la liquidez de un valor.
- **Max drawdown:** mayor pérdida desde un máximo local hasta un mínimo local. Se utiliza como indicador de riesgo sobre un periodo determinado.
- **Apalancamiento:** uso de instrumentos financieros o capital prestado con el fin de incrementar el retorno potencial de una inversión.

- *Liquidez*: describe la velocidad con el que se puede comprar/vender un activo en el mercado, sin modificar el valor del mismo. Es común encontrar activos con altos valores de liquidez en mercados “profundos”.
- *Equidad*: diferencia entre el valor de un activo y la deuda que se adquirió al conseguirlo. Se puede definir matemáticamente como:

$$E = V - d$$

Donde V es el valor del activo y d la deuda.

- *Portfolio*: grupo de instrumentos financieros como acciones, bonos, moneda y equivalentes en dinero destinados a inversión.
- *Hit Rate*: porcentaje de predicciones acertadas durante la ejecución de una estrategia independientemente de si luego se realiza una operación o no.
- *Lookback*: cantidad de valores de precios de acciones pasados utilizados para el funcionamiento de la estrategia registrados por minuto.

1.3 Indicadores de rendimiento

A la hora de comerciar en el mercado de valores existen ciertos indicadores que nos indican si nuestro *portfolio* se encuentra operando de forma correcta y acorde al riesgo deseado por el inversor. Son de gran utilidad a la hora de comparar *portfolios* teniendo en cuenta no solo sus ganancias sino también su estabilidad, riesgo y comportamiento acorde al mercado.

1.3.a Sharpe ratio

El *Sharpe ratio* es una unidad de medida para la comparación de *portfolios* de inversión la cual toma en cuenta el riesgo de dicho *portfolio* y no solamente su retorno.

Se calcula mediante la siguiente fórmula:

$$Sr = \frac{\bar{r}_p - r_f}{\sigma_p}$$

Donde \bar{r}_p = Retorno promedio del *portfolio*

r_f = Retorno libre de riesgo

σ_p = Desvío estándar del retorno del *portfolio*

Esta unidad es de alta utilidad a la hora de comparar *portfolios* ya que ante dos *portfolios* que poseen el mismo retorno nos indica cual de los dos es el menos riesgoso. El riesgo del *portfolio* es proporcional al desvío estándar de su retorno. Cuanto mayor es su desvío estándar mayor es la diferencia entre las ganancias (o pérdidas) y por ende mayor es su riesgo. Un valor

aceptable de *sharpe ratio* para un *portfolio* es cuando su numerador y denominador son iguales (el promedio del retorno es igual a su riesgo) lo que nos da un *sharpe ratio* de uno. Cuanto mayor sea el *sharpe ratio* mejor es el *portfolio* para invertir.

En la ecuación del *sharpe ratio* se resta al promedio del retorno el *retorno libre de riesgo*. Este retorno es el que un inversor esperaría de una inversión con muy bajo riesgo (usualmente bonos del estado). Dado que toda inversión se corresponde con un riesgo, al restar el *retorno libre de riesgo* estamos “llevando al cero” el retorno del *portfolio*.

1.3.b Beta

El coeficiente *Beta* es una unidad de medida que representa la tendencia de los retornos de una acción o *portfolio* ante movimientos del mercado. Se calcula realizando el cociente de la covarianza de los retornos de la acción y el mercado y la varianza de los retornos del mercado.

$$\beta = \frac{COV(r_p, r_m)}{V(r_m)}$$

Donde r_p = Retornos de la acción o *portfolio*.

r_m = Retornos del mercado.

COV = función de covarianza.

V = función de varianza.

De esta manera un Beta de 1 indica que el retorno de la acción se mueve con el del mercado en igual proporción. Esto es, si el mercado aumenta un 5% entonces el retorno de la acción aumentará un 5% también. Un Beta inferior a 1 significa que la acción es menos susceptible a cambios en el mercado y de forma contraria un Beta mayor a 1 indica que la acción es muy susceptible a cambios en mercado.

1.3.c Alpha

El coeficiente Alpha es una unidad de medida de rendimiento de una acción o *portfolio*. Compara el rendimiento de una inversión con el del mercado o *benchmark* en particular. El exceso de retorno de una inversión relativo al retorno del mercado es el Alpha de la inversión. Un Alpha de 1 nos indica que el retorno de nuestra inversión superó el retorno del mercado en un 1%. Si tuviésemos un Alpha negativo, éste reflejaría que la inversión tuvo un rendimiento peor comparado con el del mercado.

Es de suma importancia elegir correctamente el mercado o *benchmark* contra el cual se calcula éste coeficiente ya que de no estar relacionados no representará nada significativo.

El coeficiente se calcula mediante la siguiente fórmula:

$$\alpha = r_p - r_f - \beta(r_m - r_f)$$

Donde r_p = *Retorno del portfolio*
 r_f = *Retorno libre de riesgo*
 β = *El coeficiente Beta*
 r_m = *Retorno del mercado o benchmark*

2. Comercio de alta frecuencia

El comercio de alta frecuencia (o *High Frequency Trading* en inglés) es un tipo de comercio que se lleva a cabo en los mercados financieros. Éste hace uso de una plataforma de comercio automatizada (usada por bancos, inversores y brokers) la cual utiliza computadoras para realizar una gran cantidad de transacciones a altas velocidades. Estas plataformas permiten a los comerciantes ejecutar millones de órdenes y escanear múltiples mercados en cuestión de segundos, dándoles a los usuarios de estas plataformas una gran ventaja en el mercado.

Estos sistemas se valen de algoritmos para analizar los mercados e identificar cuales son los más rentables así como también predecir cambios en ellos al corto plazo. Estos algoritmos capaces de predecir el comportamiento del mercado, en nuestro caso particular acciones, son denominados estrategias y su implementación es el objetivo final del presente trabajo. Sin embargo antes de mostrar en qué consisten las estrategias implementadas es necesario entender algunos conceptos básicos del comercio de alta frecuencia así como también la plataforma en la que se desarrollaron.

2.1 Backtesting

Backtesting es el proceso por el cual se incorporan datos históricos a una estrategia ya desarrollada con el fin de evaluar cuál hubiera sido su rendimiento en el pasado. Basándonos en los resultados pasados podemos obtener información sobre cómo se ha comportado nuestra estrategia ante diferentes variables, desde ciclos económicos hasta eventos mundiales repentinos. Hay que tener en cuenta que por más fidedignos que sean nuestros datos, no podemos asumir que en un futuro nuestra estrategia se comportara de la misma manera que en su histórico. Cuanto más viejos sean nuestros históricos menos utilidad poseen.

A la hora de analizar una estrategia sobre un portfolio para ver su retorno esperado debemos tener en cuenta los siguientes posibles sesgos que pueda llegar a tener nuestro histórico de datos.

2.1.a Survivorship bias

Una base de datos de precios de acciones que no incluye las acciones que desaparecieron debido a bancarrotas, a unificaciones con otras empresas o adquisiciones de otras entidades sufre de *survivorship bias* (o sesgo de sobreviviente) ya que dicha base de datos solo posee a las acciones que “sobrevivieron” ante estos eventos. Más aún, estas acciones en su mayoría fueron removidas debido a un bajo rendimiento, lo que hace que sean de vital importancia a la hora de probar una estrategia.

Por ejemplo, los fondos de inversión de una compañía solo incluirá los fondos que han sido exitosos en el pasado. Muchos fondos cuyo rendimiento fue malo son cerrados y

unificados con otros fondos para ocultar su mal desempeño. Estos datos son muy relevantes a la hora de analizar el rendimiento pasado de dichos fondos.

2.1.b Look Ahead Bias

Al realizar el *backtesting* de una estrategia, es importante considerar el conjunto de datos que se tiene en cuenta para tomar las decisiones dentro de nuestro algoritmo. La estrategia siempre debe observar datos pasados, relativos a la fecha en la que se realizan las operaciones. Una posibilidad para evitar este inconveniente es utilizar los datos hasta el cierre del último período de comercio; estos datos se denominan *lagged*.

Para comprobar la existencia de este sesgo no deseado se puede realizar una prueba con dos conjuntos de datos, uno subconjunto del otro. Si las decisiones tomadas para los dos en su intersección son distintas, entonces el backtest sufre de *Look-Ahead bias*.

2.1.c Data Snooping Bias

Al optimizar distintos parámetros de la estrategia, se puede introducir una mejora basándose en datos históricos particulares. Esta mejora posiblemente muestre un mejor rendimiento en los *backtest* pero probablemente no se verá reflejada al tratar con datos reales. Esto se debe a que la optimización se basó en datos pasados que no necesariamente reflejan la situación real en el presente. Para evitar este tipo de sesgo se puede optar por calcular dichas mejoras en base a una ventana de tiempo que se mueve acorde avanzan las operaciones en el tiempo. También se puede dividir la muestra en dos, para optimizar el *backtesting* en la primera y ver el rendimiento en la segunda.

Paper trading, utilizar el modelo sobre datos desconocidos, nos permite detectar, además del sesgo de *data snooping*, sesgos *look ahead*. Otra alternativa es no fijar los parámetros, sino permitir que se ajusten dinámicamente dependiendo de la muestra.

2.1.d Standard & Poor's Index

El índice S&P 500 está compuesto por quinientas compañías con un gran capital en el mercado accionario, el cual ronda los diez billones de dólares. Las acciones de las compañías dentro de este índice poseen la particularidad de ser altamente líquidas.

Los índices S&P 400 y S&P 600 refieren a compañías con un capital mediano (entre 750 millones y 3 billones) y un capital bajo en el mercado respectivamente. Contrarias al índice S&P 500, cuanto menor es su capital en el mercado menor es su liquidez.

Si bien existen otros índices, como el *Russell index*, usualmente se utiliza el índice S&P debido a su fidelidad a la hora de mostrar el estado del mercado accionario de EEUU. Esto no quiere decir que en este índice están la totalidad de las empresas norteamericanas que posean un capital suficiente para estar en alguna de las tres categorías del S&P. Las empresas que se

encuentran dentro de dicho índice son elegidas por el comité de economistas del S&P en función a su participación, impacto y capital que poseen en el mercado.

2.2 Mecanismo de recolección de datos

Para recolectar datos financieros, existen varias alternativas en la red, lamentablemente gran parte de ellas son pagas. Al ser éste un trabajo meramente académico, el equipo no dispone de recursos para la compra de datos, por lo que las únicas opciones posibles son fuentes de datos abiertas.

2.2.a Google Finance

Google tiene disponible un *endpoint* al que se le pueden realizar consultas mediante HTTP. Este *endpoint* muestra para una acción el histórico de distintos valores (open, close, high, low, volumen) hasta los últimos 15 días.

El formato de la respuesta consiste en distintos encabezados, que especifican los argumentos de la consulta y los datos en un formato que se asemeja a un CSV, con la particularidad de que el valor de la primera fila correspondiente a la fecha es un timestamp de segundos precedido por el carácter a. Para las siguientes filas, este campo es un entero, que indica el offset en función del timestamp anteriormente mencionado.

A modo de ejemplo, la respuesta podría tener el siguiente formato:

```
< Múltiples headers >
a1526953902479, 234.0, 234.1
1,                234.2, 234.3
2,                233.9, 234.2
```

Es necesario destacar que este servicio ya no recibe mantenimiento por parte de Google, por lo que es poco confiable. El formato de la respuesta también es cuestionable, ya que no sigue ningún estándar conocido, y tampoco se encuentra documentación disponible.

2.2.b Yahoo Finance

Esta alternativa quedó deprecada a lo largo del desarrollo de este proyecto por sobreutilización del servicio por parte de agentes externos. Esto es evidencia de lo frágil que puede ser la integración con otros servicios gratuitos.

2.2.c Alpha Vantage

Alpha Vantage es un servicio con menos de un año de antigüedad. Posee una API REST con todos los datos necesarios de los símbolos del mercado.

El servicio es gratuito y sin límite de consulta. El rendimiento ronda los 500 ms tomando los valores entre los percentiles 25% y 75% según lo especificado en la documentación. El intervalo temporal más pequeño soportado por la plataforma es de bins de 1 minuto. Se sugiere que no se realice más de un pedido por segundo a la API, para evitar problemas de performance.

Con respecto a información intra-diaria, existe el mismo límite impuesto por Google Finance, de 10 a 15 días de datos disponibles. Si se busca información diaria, y no intra-diaria, se ofrece 20 años de datos para consultar.

2.2.d Intrinio

Intrinio es un servicio pago, que ofrece distintos datasets vía API; análisis de sentimiento o indicadores técnicos además de los convencionales con valores de high, low, open y close.

La tarifa depende del dataset, cada dataset ronda los USD 50 mensuales. Se tiene un período de prueba sin realizar pagos para luego comenzar con la facturación.

Una gran ventaja de esta plataforma es el soporte del uso de sockets. Esto permite tener en tiempo real los cambios de precio de un símbolo cualquiera; basta con establecer una conexión y recibir los mensajes.

2.2.e Módulo automático de recolección de datos

Para tener independencia de los datos de las plataformas de trading algorítmico, se decidió realizar un módulo automático de recolección de datos de alguna de las fuentes alternativas previamente mencionadas.

El desarrollo se lo hizo en Python, y consta de dos funciones distintas. La primera se encarga de descargar los símbolos que se encuentran en una bolsa de valores, como por ejemplo NYSE. Se realizó un *scraper* para consultar la página oficial de la respectiva bolsa de valores, que busca alfabéticamente los símbolos disponibles y los almacena en un archivo de texto plano. Cada símbolo se encuentra en una línea distinta, haciendo posible la lectura de los valores con cualquier lector de archivos CSV.

La segunda funcionalidad nos permite descargar los distintos valores relevantes para nuestro análisis (como pueden ser el volumen, precio, precio de cierre, etc.) para un determinado símbolo a lo largo del tiempo. Desarrollamos un conector para Google Finance,

que al recibir un símbolo, puede descargar todos los datos que disponibiliza el servicio, parsearlos y almacenarlos en un CSV.

Cada fila de este documento contiene la fecha y hora (con precisión de 1 minuto) en la que se realizó la medición, el precio de apertura, el precio de cierre, el precio mínimo, el precio máximo y el volumen.

El documento generado se puede consumir fácilmente con librerías de procesamiento de datos como *Pandas*, ya que el formato usado es un formato estándar.

Para la puesta en producción de este módulo se debería programar diariamente una tarea que lo ejecute, ya sea un simple cronjob de un sistema operativo o algo más elaborado como el uso de Apache Airflow, que permite manejar cronjobs de una manera más robusta, al controlar errores y tener un log de ejecuciones.

2.3 Quantopian

Quantopian es una plataforma que le permite a sus usuarios desarrollar, testear y poner en funcionamiento diversas estrategias de HFT. Para ello, brinda un IDE online junto con una API pública para la programación de las estrategias, y datos históricos de alta fidelidad de los distintos mercados accionarios para realizar pruebas sobre ellas. Estos servicios los ofrece de forma gratuita dentro de su aplicación web y fue la plataforma utilizada para la realización del presente trabajo.

Si bien parte de las tecnologías utilizadas por la compañía son open source, como por ejemplo su motor para realizar backtests de estrategias llamado Zipline. Los algoritmos desarrollados por sus usuarios se mantienen en secreto a menos que el propietario desee hacerlos públicos. Esto comprende un aspecto clave de la plataforma ya que el funcionamiento de dichos algoritmos son la “forma de hacer dinero” de sus propietarios y si se divulgaran perderían eficiencia. Recordemos que en las estrategias su confidencialidad hace a su eficiencia, si múltiples comerciantes detectan las mismas oportunidades y realizan las mismas acciones de mercado, el retorno de dichas acciones será inferior.

El modelo de negocios de la plataforma consiste en ofrecer financiamiento a las mejores estrategias creadas por sus usuarios (usualmente inversiones de más de un millón de USD). Quantopian luego se queda con la mayor parte las ganancias generadas pagando un porcentaje de las mismas al creador (usualmente 10%). Para seleccionar las mejores estrategias Quantopian organiza concursos mensuales en los que los usuarios pueden participar. Luego la estrategia ganadora es evaluada en detalle para determinar si cumple con los requisitos de sus inversores y de serlo realizar una oferta. Esto último es un poco controversial ya que Quantopian requiere conocer el funcionamiento de la estrategia para determinar si realiza una oferta o no, en otras palabras, el usuario está obligado a revelar los secretos de su algoritmo. Según la plataforma en el caso de no realizar una oferta el secreto sigue siendo guardado. Por otro lado, el usuario tienen la posibilidad de invertir su propio

capital en su estrategia para no depender totalmente de Quantopian el cual pasaría a prestar servicios similares a los de un broker. En nuestra opinión, esta opción es un tanto riesgosa no por los posibles errores de que posea la estrategia, sino porque nos resultó difícil encontrar documentación clara respecto al funcionamiento de la plataforma con dinero real. Pareciera que éste caso es desalentado por la misma plataforma.

2.3.a Limitaciones

Permitir que usuarios ejecuten código arbitrariamente en servidores presenta una serie de desafíos en términos de seguridad informática e infraestructura. Python es un lenguaje con miles de librerías destinadas a diversos usos. Cada una de estas librerías debe ser analizada teniendo en cuenta su potencial uso malicioso. Obviamente Quantopian no puede analizar todas las librerías de python por lo que restringe su uso a solo unas pocas, estas son: `bisect`, `cmath`, `collections`, `copy`, `cvxopt`, `datetime`, `functools`, `heapq`, `itertools`, `math`, `numpy`, `operator`, `pandas`, `pytz`, `Queue`, `random`, `re`, `scipy`, `sklearn`, `statsmodels`, `time`, `talib`, `zipline`, `zlib`. Cabe destacar que tampoco mantiene al día las librerías que soporta por lo que se debe tener cuidado de no importar una versión no soportada.

Por otro lado, la plataforma posee una limitación en términos de tiempo de ejecución de código, no permite la ejecución continua de código por más de 50 segundos. Esta limitación es entendible dada la naturaleza plataforma, apunta a no malgastar recursos por parte de sus usuarios. Sin embargo, reduce en gran medida los tipos de algoritmos informáticos a utilizar especialmente aquellos relacionados con *machine learning* o *big data* como pueden ser redes neuronales *LSTM*.

Por último, la plataforma desalienta el desarrollo de código por otro medio que no sea el IDE online. Si bien un usuario podría desarrollar la estrategia en forma local utilizando el IDE que prefiera se ve forzado a copiar el código en el IDE de la plataforma para ejecutarlo o utilizar el *debugger*. Esta limitación dificulta el desarrollo ya que el IDE provisto por Quantopian no posee muchas de las herramientas que poseen la mayoría de los IDEs del mercado como autocompletado de código o indicadores de errores de compilación.

3. Estrategias

A lo largo del desarrollo del presente trabajo, se implementaron múltiples estrategias haciendo foco en aspectos estadísticos y computacionales. La mayoría de las estrategias implementadas, aunque efectivas en retorno, eran deficientes en términos de riesgo y apalancamiento.

De todas las estrategias implementadas, presentamos en este trabajo las tres que, en nuestra opinión, dieron los mejores resultados en retorno, riesgo y apalancamiento. Para ésta selección se prestó especial atención a los indicadores de rendimientos descritos en la sección 1.3.

3.1 Convergencia y Divergencia por Media Móvil (*Moving Average Convergence Divergence* o *MACD*)

La hipótesis detrás del uso de *Moving Average Convergence Divergence* (MACD) indica que, haciendo hincapié en el comportamiento visto en una regresión a la media, se puede inferir el comportamiento futuro comparando los precios a corto y largo plazo [9].

Partiendo de la base de que los precios de los últimos minutos tienen mayor relevancia que los observados hace varios días, se utilizará una *EMA* (*exponential moving average*) para suavizar la curva de precios, ya que pondera con mayor peso los inmediatamente anteriores.

Definimos:

$$EMA_K(0) = p_{-K} \quad (3.1)$$

$$EMA_K(t) = \alpha p_{-K+t} + (1 - \alpha) EMA_K(t - 1)$$

Donde: K : *pasos temporales pasados a considerar para la media*
 α : *decaimiento del peso a lo largo del tiempo, entre (0, 1)*

Para realizar una comparación entre los precios en el corto plazo y los precios en el largo plazo, definimos:

N : *cantidad de pasos temporales pasados a corto plazo*
 M : *cantidad de pasos temporales pasados a largo plazo*
 $EMA_N(x)$: *EMA a corto plazo*
 $EMA_M(x)$: *EMA a largo plazo*

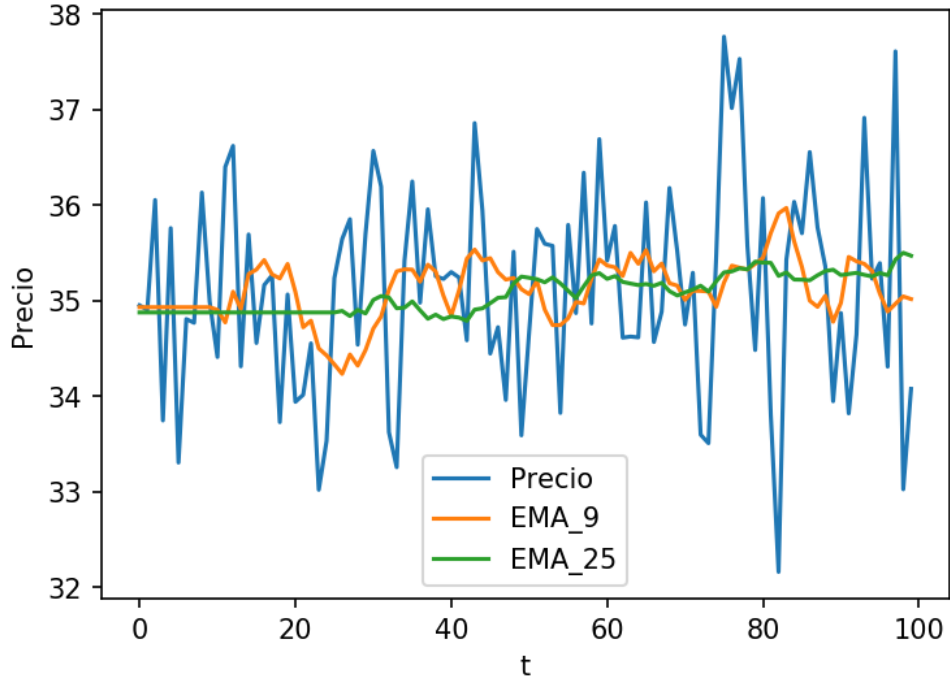


Figura 3.1: Comparación entre una serie temporal de precios y sus correspondiente EMA_9 y EMA_{25} .

Con EMA_N y EMA_M definimos:

$$MACD_{N,M}(t) = EMA_N(t) - EMA_M(t) \quad (3.2)$$

Este valor permite comparar la diferencia entre la tendencia a largo y corto plazo del precio en función del instante t .

Además, se puede calcular el EMA sobre los valores del mismo índice de $MACD_{N,M}(t)$, obteniendo una función de señal $SIG_{N,M}$ definida en la ec (3.3):

$$SIG_{N,M}(0) = MACD_{N,M}(0) \quad (3.3)$$

$$SIG_{N,M}(t) = \alpha MACD_{N,M}(t) + (1 - \alpha) SIG_{N,M}(t - 1)$$

La serie $SIG_{N,M}$ será el indicador para realizar órdenes, tanto short como long.

Para ejecutar una orden long, se debe cumplir la siguiente condición:

$$SIG_{N,M}(t - 1) > MACD_{N,M}(t - 1) \wedge SIG_{N,M}(t) < MACD_{N,M}(t) \quad (3.4)$$

Caso contrario, para ejecutar una orden short, se debe cumplir la siguiente condición:

$$SIG_{N,M}(t-1) < MACD_{N,M}(t-1) \wedge SIG_{N,M}(t) > MACD_{N,M}(t) \quad (3.5)$$

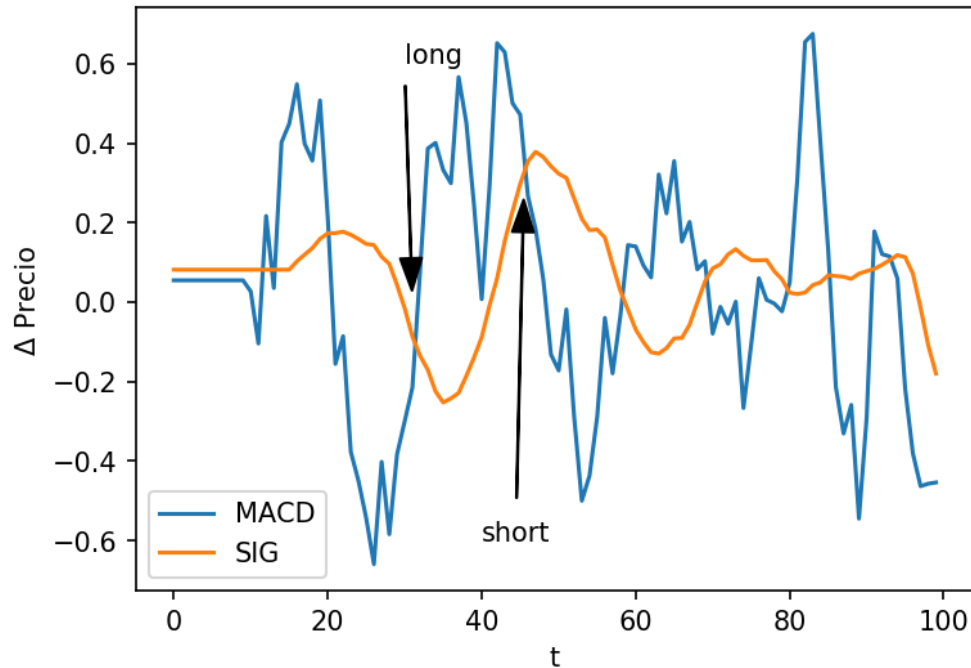


Figura 3.2: Toma de decisiones en base a los cruces de las curvas MACD y SIG.

Las variantes que ofrece esta estrategia son diversas. Se puede utilizar un promedio simple en vez de uno exponencialmente ponderado. También es posible modificar los parámetros k , N , M que se utilizan para el cálculo de los promedios para los períodos de corto y largo plazo. Por ejemplo, se puede probar con un corto plazo de $N = 9$ y un largo plazo de $M = 24$ o con una configuración de 15 y 26 respectivamente.

La estrategia se puede implementar ya sea para una sola acción o para diversas acciones. Si se utilizan diversos símbolos, en cada tick se opera long o short con la misma cantidad de acciones para cada uno de los símbolos, así se realiza el balanceo del *portfolio*.

3.2 ARMA

Para esta estrategia se utilizó un modelo ARMA (acrónimo del inglés *Autoregressive Moving Average*) para la predicción de los retornos de diversas acciones. Este modelo se basa en la unión de dos modelos, uno autorregresivo y otro de media móvil con el fin de predecir el siguiente valor de la serie.

La serie sobre la cual se realizan las predicciones es la conformada por los retornos logarítmicos de la acción a predecir con un determinado horizonte de predicción. En nuestro caso se analizaron los horizontes $h \in \{10 \text{ min}, 20 \text{ min}, 30 \text{ min}, 60 \text{ min}\}$ utilizando $n = 150$ retornos logarítmicos. En otras palabras la serie se compone por n retornos logarítmicos

calculados cada h minutos, utilizando un *lookback* $l = nh$ dándonos como resultado la siguiente serie:

$$S = [R_1, R_2, \dots, R_n] \quad (3.6)$$

Como se dijo anteriormente, ARMA consta de dos modelos principales: AR y MA. A continuación describiremos en más detalle en que constan cada uno de ellos.

El modelo AR que se traduce a *Autoregressive* y cuyo orden lo denotaremos con la letra “a” nos indica que la serie a analizar se la puede adaptar a una combinación lineal de sus *lags* (valores anteriores) junto con un término de ruido blanco y una constante. El orden (a) del modelo expresa el número de observaciones de la serie temporal que intervienen en la ecuación. De esta forma, un modelo $AR(a)$ tiene la siguiente expresión [4]:

$$S_t = c + \sum_{i=1}^a \phi_i S_{t-i} + \varepsilon_t \quad (3.7)$$

Donde $c = \text{constante}$
 $\phi = \text{parametros del modelo}^1$
 $\varepsilon = \text{ruido blanco}^2$

El modelo MA por *Moving Average*, su orden se denota con la letra q , que nos indica que el valor de una determinada variable en un instante t se lo puede obtener en función de el valor medio de la serie sumado a una combinación lineal de valores de ruido blanco correspondientes a períodos anteriores. De esta manera, un modelo $MA(q)$ tiene la siguiente expresión [3]:

$$S_t = \mu + \varepsilon_t + \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad (3.8)$$

Donde $\mu = \text{media de la serie}$
 $\varepsilon = \text{ruido blanco}$
 $\theta = \text{parametros del modelo}^3$

Uniendo estos dos modelos independientes obtenemos el modelo ARMA(a, q) el cual posee la siguiente expresión [5]¹:

$$S_t = c + \varepsilon_t + \sum_{i=1}^a \phi_i S_{t-i} + \sum_{j=0}^q \theta_j \varepsilon_{t-j} \quad (3.9)$$

Para modelar una serie temporal utilizando ARMA debemos asegurarnos de que esta sea estacionaria. Una serie estacionaria significa que no muestra tendencia, es decir, posee una

¹ Existen varios métodos para determinar los valores de estos parámetros, el más usado es cuadrados mínimos.

² En este caso el ruido blanco es una señal discreta con media cero y varianza constante.

³ Calculados de igual manera que los del modelo autorregresivo.

media y variancia constante a lo largo del tiempo. Para asegurarnos que nuestra serie cumpla ésta propiedad realizamos una prueba de raíz unitaria de *Dickey-Fuller aumentada* (o *ADF* por su siglas en inglés) sobre la misma. Si el *p-value* resultante de la prueba ADF es menor a 0.05 (o 5%) se puede concluir que la serie es estacionaria, de lo contrario nos indica que la serie posee una raíz unitaria y que por lo tanto es no-estacionaria. En el caso de que suceda lo último no se puede aplicar el modelo por lo que se saltea la predicción -de todas formas, en las pruebas realizadas se vio que siempre se cumplía ésta condición.

Para determinar los valores de los órdenes a y q correspondientes a los modelos *AR* y *MA* se realiza una búsqueda orientada utilizando el valor AIC del modelo. El criterio de información de Akaike (AIC por su sigla en inglés) es una medida de calidad relativa de un modelo estadístico. El criterio no indica qué tan bueno es un modelo en sí mismo sino que nos proporciona una unidad de comparación entre modelos. Dados un conjunto de datos y varios modelos candidatos estos pueden ser comparados entre sí utilizando su valor de AIC, el modelo con el menor AIC es el más adecuado [6].

Es importante destacar que los valores (a, q) varían dependiendo de la acción, el horizonte de predicción, el *lookback* y del comportamiento de la acción en el periodo el cual se están realizando las predicciones. Dado que éste último parámetro es imposible de determinar con anterioridad, los valores (a, q) del modelo se computan día a día a medida que la estrategia avanza en el tiempo.

A la hora de realizar el rebalanceo del *portfolio* la lógica es muy simple, si el valor predicho del retorno es mayor a cero para una determinada acción ejecutamos una orden *long*, de lo contrario se ejecuta una orden *short* con una porción de las acciones si es que se posee alguna. Este proceso se ejecuta secuencialmente para todas las acciones del *portfolio*.

3.3 Máquina de soporte vectorial

Las máquinas de soporte vectorial (o SVM por su siglas en inglés) es un modelo de clasificación binario. A continuación explicaremos su funcionamiento a grandes rasgos y su utilización para la predicción del comportamiento del mercado accionario.

Las SVM [7] son un conjunto de algoritmos que caen dentro de la categoría de aprendizaje supervisado⁴. Dado un conjunto de puntos en un espacio en donde cada uno de ellos pertenece a una de dos categorías el algoritmo construye un modelo capaz de predecir si un nuevo punto pertenece a una categoría u otra. Para realizar esta clasificación, la SVM busca el hiperplano que mejor diferencie dichas categorías en el espacio. A la hora de realizar esta separación las SVMs no buscan simplemente uno -de los infinitos- hiperplanos que diferencian ambas categorías de puntos, sino que buscan el hiperplano que maximiza dicha distancia. Por esto también se los conoce como *clasificadores de margen máximo*.

⁴ El aprendizaje supervisado es una técnica para deducir una función a partir de datos de entrenamiento. Usualmente pares de vectores de entrada y salida.

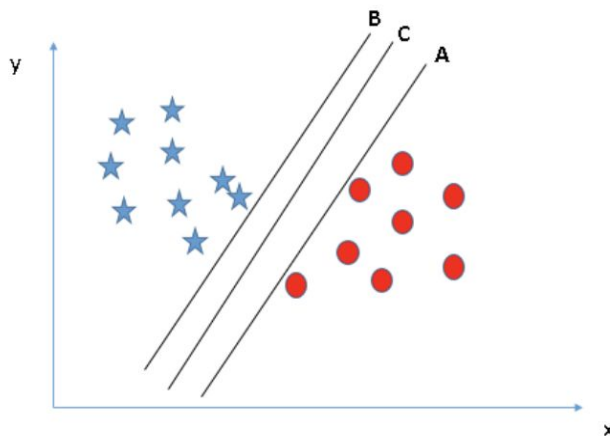


Figura 3.3 : representación de la separación de dos conjuntos de puntos en un plano por un hiperplano.

En la Fig.3.3 podemos ver como las dos categorías (representadas por círculos y estrellas) se ven separadas por el hiperplano C el cual posee un margen máximo con los hiperplanos B y A. Los puntos que en el espacio lindan con los hiperplanos B y A son denominados vectores de soporte ya que la distancia entre ellos es la mayor posible.

Este tipo de separación nos lleva a identificar una de las limitaciones que poseen las SVMs, solo pueden trazar hiperplanos lineales. Esta limitación se soluciona tomando funciones, conocidas como "kernel" que permiten mapear los valores a una dimensión mayor (o distinta) para luego clasificar en ese espacio. Una clasificación lineal en un espacio de dimensión mayor es una clasificación no lineal en el espacio original.

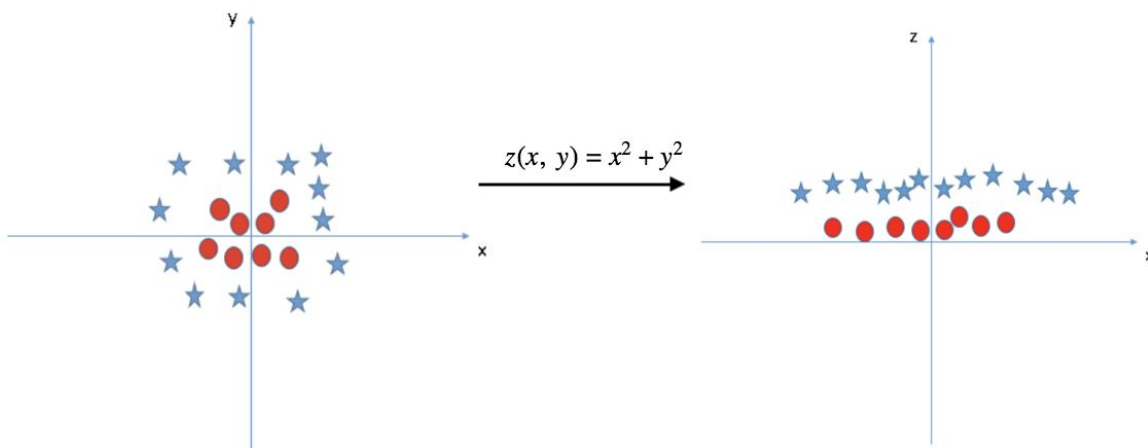


Figura 3.4: representación de dos conjuntos de puntos antes y después de aplicarles la función de kernel $k(x, y) = x^2 + y^2$.

Como se puede ver en la Fig.3.4 inicialmente era imposible trazar un hiperplano que separe los conjuntos de puntos debido a su distribución circular concéntrica. Al aplicarles la función de kernel $k(x, y) = x^2 + y^2$ vemos como ambos conjuntos se distribuyen de forma horizontal paralelos al eje x de forma tal que es posible dividirlos por un hiperplano.

3.3.a Datos y Modelo

En este trabajo utilizamos un SVC (*Support Vector Classifier*) con RBF (*Radial Basis Function*) como función de kernel implementados en la librería *sklearn* de *Python*. El objetivo es poder clasificar el retorno de una acción como positivo o negativo valiéndonos de diversos parámetros explicados más adelante. El retorno de una acción es positivo si aumenta su valor y negativo si decrece.

En cuanto a las acciones seleccionadas para llevar a cabo las predicciones optamos por elegir acciones que pertenezcan a un mismo sector, en nuestro caso NASDAQ-100. Esto nos permite incorporar al modelo una noción de contexto en el cual se desempeñan estas acciones mediante la utilización de métricas propias del sector.

Como se mencionó anteriormente la función de kernel utilizada es RBF, es una de las funciones de kernel más populares y se define como [7]:

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (3.10)$$

El numerador del exponente se lo puede identificar como la distancia euclidiana entre dos vectores de input mientras que σ en el denominador es un parámetro libre. Al utilizar la distancia euclidiana de los vectores de *input* se pondera la similitud de los mismos. Es decir que la clasificación está altamente basada en la semejanza de los *inputs* lo cual nos permite tomar ventaja de los patrones en los datos.

3.3.b Vectores de input/output

Para esta estrategia se utilizaron seis *features* (o características) de *input* para clasificar el retorno de una acción: retorno medio, volatilidad del retorno, momento, retorno medio del sector, volatilidad del retorno del sector y momento del sector -definidas en las ecs. 3.11, 3.12 y 3.13. Como *output* se utilizó un único valor, 1 si el retorno de la acción es positivo o -1 si es negativo. De forma similar a ARMA, si se predice que el valor de la acción subirá se ejecuta una orden *long* de lo contrario se ejecuta una orden *short* una porción de las acciones si se posee alguna.

A la hora de calcular los pares *input/output* se optó por utilizar una ventana que recorra los valores obtenidos para un determinado *lookback*. La longitud de esta ventana se calcula mediante la cantidad de observaciones n para un determinado horizonte de predicción denotado con la letra h . En nuestro caso $n \in \{10, 20, 30, 40\}$ y

$h \in \{10 \text{ min}, 20 \text{ min}, 30 \text{ min}, 60 \text{ min}\}$ de esta forma, la longitud w de la ventana (en minutos) viene dada por la siguiente ecuación $w = nh + h$. Cabe destacar que se le suma h al final de la ecuación para calcular el valor de output.

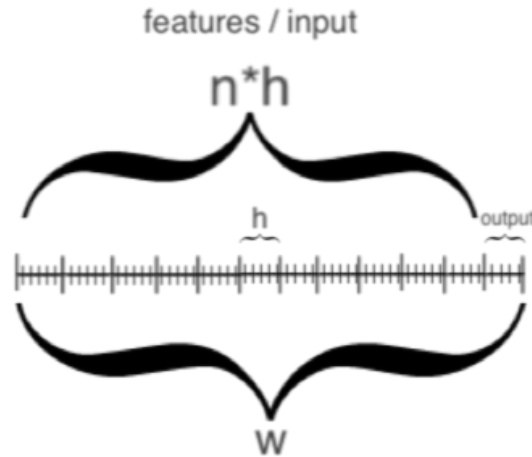


Figura 4: representación de una ventana indicando que secciones son utilizadas para el cálculo de las *features* y *output*.

Prestando atención a la Fig.4 podemos ver que para $n = 10$ y $h = 10$ obtendremos diez retornos logarítmicos destinados al cálculo de las características de *input* más un retorno extra que será utilizado para el cálculo del *output*. Para este último solo nos interesa su signo, si fue positivo o negativo. De esta forma, a medida que avanza la ventana se irán obteniendo los pares *input/output* que luego serán utilizados para el entrenamiento del SVC.

A continuación se detalla el cálculo de cada una de las características elegidas [8]:

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i \quad (3.11)$$

$$\sigma = \frac{1}{n} \sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \quad (3.12)$$

$$M = \sum_{i=1}^n \text{sign}(R_i) \quad (3.13)$$

Donde \bar{R} es el *retorno logarítmico medio*.

σ es la *volatilidad del retorno*.

M es el *momento*

De forma análoga se calculan las características referentes al sector.

Para determinar la cantidad de pares *input/output* a utilizar en el entrenamiento del SVC se optó por realizar sucesivos *backtestings* prestando especial atención al valor de *hit rate* de la estrategia. De esta forma se aumenta o disminuye la cantidad de elementos en el conjunto de entrenamiento hasta dar con el *hit rate* máximo. Al estar trabajando con una SVM usted se podría preguntar porqué no determinamos este valor como el máximo número de elementos antes de llegar al overfitting. Si bien esa es la forma tradicional de determinar dicho valor creemos que no se ajusta a las condiciones particulares del algoritmo. En nuestro caso hay que tener presente la antigüedad de los datos utilizados. Demasiados datos de entrenamiento pueden ser perjudiciales para la precisión de la estrategia, no porque se genere overfitting sino porque al ser más antiguos pueden no reflejar el comportamiento reciente de la acción de manera precisa.

A partir de la cantidad deseada e de elementos del conjunto de entrenamiento se calcula el valor de *lookback* como $l = e h + w$. Los respectivos conjuntos de entrenamiento se obtienen a través de los siguientes pasos:

1. Obtención del histórico de precios de cada acción y del sector según el valor de *lookback*.

$$prices = [p_1, p_2, p_3, \dots, p_l]$$

2. Cálculo de los retornos logarítmicos para cada acción y el sector con un determinado horizonte de predicción.

$$returns = [R_1, R_2, R_3, \dots, R_{e+n+1}]$$

3. Cálculo de las características sobre los retornos logarítmicos utilizando una ventana de longitud w dándonos como resultado los arreglos de *input/output*. Se utilizó la letra j para denotar las características referentes al sector.

$$inputs = [[\bar{R}, \sigma, M, \bar{R}_j, \sigma_j, M_j]_1, \dots, [\bar{R}, \sigma, M, \bar{R}_j, \sigma_j, M_j]_e]$$

$$outputs = [sign(R_0), \dots, sign(R_e)]$$

Valiéndonos de estos conjuntos de entrenamiento se realizaron diversos backtests para verificar la efectividad del modelo variando los parámetros n y h .

4. Resultados

Con el fin de poder comparar las tres estrategias en términos de eficiencia, riesgo y precisión se decidió utilizar el mismo conjunto de acciones en todos los *backtests*, estas son: AAPL, AMZN, MSFT y FB correspondientes a las empresas Apple, Amazon, Microsoft y Facebook. Dichas acciones conforman nuestro *portfolio* de inversión distribuyendo el capital del mismo en forma equitativa entre las 4 acciones; el capital inicial del *portfolio* fue de 100.000 USD. Se eligieron estas acciones ya que presentan una alta liquidez y son comerciadas frecuentemente -lo que aumenta la calidad de los datos provistos por Quantopian⁵-. Todos los *backtests* se realizaron durante el año 2017, esto es desde el 1/1/2017 hasta el 31/12/2017 inclusive. Para la comparación de las estrategias se utilizaron los indicadores de rendimiento descritos en la sección 1.3 junto con su valor *Max Drawdown* y retorno.

Como punto de comparación para el rendimiento de las estrategias utilizaremos un *backtest* realizado en el mismo periodo y con las mismas acciones en el cual se realiza una operación *buy & hold* para cada una de las acciones con la misma cantidad de capital. En otras palabras, invertiremos todo el capital del *portfolio* en el primer día y no realizaremos ninguna otra operación a lo largo del año. El resultado de referencia *buy & hold* es:

Alpha	Beta	Sharpe Ratio	Max Drawdown (%)	Retorno (%)
0.12	1.08	2.78	-4.61	38.13

Tabla 1: resultado de referencia *buy & hold*

4.1 MACD

Para los parámetros $N \in \{10, 30\}$, $M \in \{60, 90\}$ y $k \in \{40, 60\}$ variables, operando cada 10 minutos, la estrategia arroja los resultados en *backtesting* presentados en la tabla 2.

id	N	M	k	Alpha	Beta	Sharpe Ratio	Max Drawdown	Hit Rate	Retorno	Predicciones
1	10	60	40	0.10	0.65	2.16	-3.66%	58.12	24.68%	647
2	10	60	60	0.07	0.62	1.89	-3.66%	57.20	20.76%	580
3	10	90	40	0.07	0.62	1.86	-4.13%	57.30	20.44%	597
4	10	90	60	0.08	0.64	1.94	-3.94%	57.37	20.88%	511

⁵ Si una acción no es comerciada en un determinado instante Quantopian no registra su valor devolviendo Nan si se consulta su precio.

5	30	60	40	0.08	0.68	1.83	-4.11%	55.07	23.67%	437
6	30	60	60	0.07	0.6	1.98	-4.42%	54.44	21.82%	381
7	30	90	40	0.06	0.67	1.83	-4.65%	55.78	20.05%	361
8	30	90	60	0.05	0.68	1.79	-6.13%	54.23	19.33%	346

Tabla 2: resultados de los backtests realizados para la estrategia MACD

Estos resultados se realizaron sin ningún tipo de apalancamiento, realizando la compra de hasta 90 *stocks* por símbolo y por tick. A modo de observación, si se introduce un apalancamiento no mayor a 0.3, consecuencia de operar con hasta 100 *stocks* por símbolo y por tick, se pueden conseguir resultados notablemente mejores. En la tabla 3 se presentan las mejores tres configuraciones con apalancamiento:

id	N	M	k	Alpha	Beta	Sharpe Ratio	Max Drawdown	Hit Rate	Retorno	Predicciones
1'	10	60	40	0.26	0.75	2.65	-3.88%	58.12	48.9%	647
5'	30	60	40	0.18	0.72	2.12	-4.44%	55.07	36.3%	437
6'	30	60	60	0.21	0.75	2.16	-5.07%	54.44	40.83%	381

Tabla 3: resultados de los backtests realizados para la estrategia MACD con apalancamiento

Es importante recalcar que esta estrategia operó únicamente con órdenes long, sin realizar órdenes short, ya que los resultados operando con ambas órdenes o únicamente con órdenes short no arrojaban resultados positivos. En la figura 5 mostramos un gráfico comparando el retorno de la estrategia con su mejor configuración ($id = 1'$) respecto al *backtest buy & hold* a lo largo del año.



Figura 5: retorno a lo largo del tiempo de la estrategia MACD (configuración $id = 1'$) y *Buy & Hold*.

4.2 ARMA

Resultados de los *backtests* para los distintos horizontes de predicción

$h \in \{10 \text{ min}, 20 \text{ min}, 30 \text{ min}, 60 \text{ min}\}$ especificando sus indicadores de rendimiento y cantidad de predicciones realizadas.

id	h (min)	Alpha	Beta	Sharpe Ratio	Max Drawdown (%)	Hit Rate (%)	Retorno (%)	Predicciones
1	10	-0.02	1.12	1.66	-7.14	50.27	21.9	31224
2	20	0.05	1.17	2.1	-6.42	50.27	31.73	15105
3	30	0	1.24	1.83	-5.05	49.98	28.82	10355
4	60	0.1	1.3	2.53	-6.75	50.95	42.05	4581

Tabla 4: resultados de los backtests realizados para la estrategia ARMA.

En la figura 6 mostramos un gráfico comparando el retorno de la estrategia con su mejor configuración ($id = 4$) respecto al *backtest buy & hold* a lo largo del año.



Figura 6: retorno a lo largo del tiempo para la estrategia SVM (con la configuración $id = 1$) y *Buy & Hold*.

4.3 Máquina de soporte vectorial

Resultados de los *backtests* con $n \in \{10, 20, 30, 40\}$ y $h \in \{10 \text{ min}, 20 \text{ min}, 30 \text{ min}, 60 \text{ min}\}$ junto con la cantidad de predicciones realizadas y el tamaño del conjunto de entrenamiento e .

id	n	h (min)	Alpha	Beta	Sharpe Ratio	Max Drawdown (%)	Hit Rate (%)	Retorno (%)	Predicciones	e
1	10	10	0.14	1.2	3.06	-4.79	50.84	59.57	32930	500
2	10	20	0.09	1.12	2.97	-5.23	50.61	48.64	15980	500
3	10	30	0.01	1.19	2.30	-6.33	51.2	39.9	10985	200
4	10	60	0.05	1.24	2.62	-6.16	51.85	49.06	4995	500
5	20	10	0.05	1.15	2.61	-7.95	50.78	42.16	32973	500
6	20	20	0.05	1.13	2.65	-4.29	50.82	43.48	15982	700
7	20	30	-0.13	1.18	1.43	-11.02	50.8	22.19	10988	700
8	20	60	0.07	1.25	2.74	-6.03	51.15	51.49	4991	500
9	30	10	0.07	1.15	2.79	-4.76	50.38	46.86	32967	600
10	30	20	0.03	1.18	2.54	-4.67	51.15	43.27	15985	700
11	30	30	0.01	1.15	2.36	-6.95	50.57	39	10988	600
12	30	60	0.1	1.27	2.8	-6.41	51.23	56.92	4991	900
13	40	10	-0.01	1.21	2.22	-7.45	50.92	38.61	32965	1000
14	40	20	0.0	1.18	2.24	-6.08	51	38.3	15920	1000
15	40	30	0.04	1.09	2.56	-5.48	51.05	40.65	10985	500
16	40	60	0.1	1.2	2.0	-5.77	51.46	54.28	4994	700

Tabla 5: resultados de los *backtests* realizados para la estrategia SVM.

En la figura 7 mostramos un gráfico comparando el retorno de la estrategia con su mejor configuración ($id = 1$) respecto al *backtest buy & hold* a lo largo del año.



Figura 7: retorno a lo largo del tiempo para la estrategia SVM (con la configuración $id = 1$) y *Buy & Hold*.

5. Análisis de los resultados

En esta sección analizaremos los resultados obtenidos para las distintas estrategias haciendo foco en sus indicadores de rendimiento y comparándolos con los del *backtest* de referencia mostrado en la sección 4. Para ello, se restan los indicadores obtenidos en los diversos *backtests* menos los del *backtest* de referencia encontrados en la tabla 1. Para una mejor visualización marcaremos con verde aquellos resultados obtenidos en las estrategias que superaron los del *backtest* de referencia y con rojo los que no. También prestaremos especial atención a los valores de *hit rate* obtenidos con el fin de refutar la hipótesis del mercado eficiente.

A la hora de comparar los *hit rates* de los distintos *backtests* nos encontramos con un problema: la cantidad total de predicciones realizadas varía dependiendo de la estrategia y sus parámetros. Comparar los *hit rates* sin tener en cuenta la cantidad de predicciones realizadas sería engañoso ya que es de esperar que con una cantidad mayor de predicciones la medición sea más precisa. En otras palabras, estaríamos comparando valores con distintos grados de precisión.

Para solucionar este problema de forma sencilla se podría aumentar el tiempo en el cual se realiza el *backtest* de manera tal que en todos se realicen la misma cantidad de predicciones. Esta solución, aunque efectiva, nos presenta con otro problema: estaríamos comparando *hit rates* que fueron calculados durante distintos periodos de tiempo. Si bien esto quizás parezca inofensivo, las acciones pudieron haber tenido comportamientos abruptos -debido a cambios repentinos del mercado o de las empresas⁶- los cuales muy difícilmente puedan ser predichos por los algoritmos deteriorando su performance.

Otra solución posible que aumenta el número de predicciones realizadas por las estrategias es aumentar la cantidad de acciones con las cuales trabaja. Sin embargo, esta solución también presenta un problema a la hora de comparar los *hit rates* obtenidos. De forma similar a la anterior estas nuevas acciones podrían tener comportamientos más erráticos o estables afectando la medición del *hit rate*.

De esta manera podemos concluir que no podemos variar el espacio temporal ni la cantidad de acciones para comparar los *backtests* de forma correcta. Es por ello que decidimos recurrir a herramientas estadísticas para comparar dichos valores, en particular a los intervalos de confianza. Compararemos los distintos valores de *hit rate* basándonos en los extremos de sus intervalos de confianza al 99%.

Dado que el *hit rate* es una variable que posee una distribución binomial realizamos el cálculo del intervalo de confianza aproximando dicha distribución a una distribución normal [10]

⁶ Ej: recientemente Facebook presentó una bajada muy grande en el valor de sus acciones debido a sus contratos con la empresa Cambridge Analytica.

utilizando el teorema del límite central⁷. De esta forma calculamos el intervalo de confianza como:

$$p \pm z \sqrt{\frac{p * (1-p)}{n}} \quad (5.1)$$

Donde p = probabilidad de éxito
 n = cantidad de predicciones
 z = valor critico

5.1 MACD

Para esta estrategia se utilizaron los resultados de las 3 configuraciones con y sin apalancamiento que arrojaron mayor retorno respectivamente.

Diferencia de los <i>backtests</i> realizados con el <i>backtest</i> de referencia					
id	Alpha	Beta	Sharpe Ratio	Max Drawdown (%)	Retorno (%)
1	-0.02	-0.43	-0.62	0.95	-13.45
5	-0.04	-0.4	-0.95	0.5	-14.46
6	-0.05	-0.48	-0.92	0.19	-16.31
1'	0.14	-0.33	-0.13	0.73	10.77
5'	0.06	-0.36	-0.66	0.17	-1.83
6'	0.09	-0.33	-0.63	-0.46	2.7

Tabla 6: diferencia de los *backtests* realizados con el *backtests* de referencia para la estrategia MACD.

⁷ Esta aproximación no es confiable si la cantidad de observaciones es pequeña o la probabilidad de éxito de la variable es cercana a 0 o 1. A nuestro favor el *hit rate* cumple con estas restricciones.

Intervalos de confianza al 99% del <i>hit rate</i> para cada uno de los <i>backtests</i> realizados			
id	Hit Rate (%)	Intervalo (-)	Intervalo (+)
1 y 1'	58.12	53.12356942	63.11643058
5 y 5'	55.07	48.9404176	61.1995824
6 y 6'	54.44	47.86744413	61.01255587

Tabla 7: intervalos de confianza para los *backtests* realizados para la estrategia MACD.

Esta estrategia exhibe un buen desempeño si observamos los valores beta que arroja, rondando el 0.6 y 0.7. Si se los compara al valor del *benchmark*, es una mejora notable ya que buscamos que este valor sea lo más cercano a 0 para no tener ningún tipo de correlación con el mercado. Esto nos permitiría tener retornos positivos sin depender de que el retorno del mercado general sea positivo o negativo.

Por otra parte, el *Sharpe Ratio* demuestra que por ser menor al de referencia, aun cuando los retornos son mayores a los del *benchmark*, el riesgo de esta estrategia es mayor. Esto se explica por el aumento de la varianza de los retornos en esta estrategia, que aumenta, haciendo que el *Sharpe Ratio* disminuya.

Los valores observados de *hit rate* son elevados, pero al calcular un intervalo de confianza obtenemos que únicamente uno de los tres relevantes se encuentran por encima del 50% de una manera estadísticamente significativa. Las estrategias siguientes realizan una cantidad elevada de predicciones por día, mientras que esta estrategia realiza predicciones condicionada al caso particular en la que dos curvas se cruzan. A lo largo del día este tipo de situaciones no se repiten con una alta frecuencia, a pesar de que el algoritmo opere cada 10, 5 o 1 minuto.

5.2 ARMA

Diferencia de los <i>backtests</i> realizados con el <i>backtest</i> de referencia					
id	Alpha	Beta	Sharpe Ratio	Max Drawdown (%)	Retorno (%)
1	-0.14	0.04	-1.12	-2.53	-16.23
2	-0.07	0.09	-0.68	-1.81	-6.4
3	-0.12	0.16	-0.95	-0.44	-9.31
4	-0.02	0.22	-0.25	-2.14	3.92

Tabla 8: diferencia de los *backtests* realizados con el *backtests* de referencia para la estrategia ARMA.

Intervalos de confianza al 99% del <i>hit rate</i> para cada uno de los <i>backtests</i> realizados			
id	Hit Rate (%)	Intervalo (-)	Intervalo (+)
1	50.27	49.54	50.99
2	50.27	49.22	51.32
3	49.98	48.71	51.24
4	50.95	49.05	52.85

Tabla 9: intervalos de confianza para los *backtests* realizados para la estrategia ARMA.

Como se puede ver en las tablas presentes en esta sección, los resultados obtenidos para esta estrategia fueron peores que los del *backtest* de referencia en todos los aspectos. Solo el *backtest* con $id = 4$ dio un retorno levemente superior que no justifica el incremento de riesgo reflejado en el resto de los indicadores.

Si analizamos el *hit rate*, vemos que si bien sobrepasa el 50% es casi despreciable. Teniendo en cuenta los intervalos de confianza al 99% no podemos asegurar que la estrategia está prediciendo los movimientos de las acciones ya que ninguno de los intervalos se encuentra totalmente por encima del 50%.

5.3 Máquina de soporte vectorial

Diferencia de los <i>backtests</i> realizados con el <i>backtest</i> de referencia					
--	--	--	--	--	--

id	Alpha	Beta	Sharpe Ratio	Max Drawdown (%)	Retorno (%)
1	0.02	0.12	0.28	-0.18	21.44
2	-0.03	0.04	0.19	-0.62	10.51
3	-0.11	0.11	-0.48	-1.72	1.77
4	-0.07	0.16	-0.16	-1.55	10.93
5	-0.07	0.07	-0.17	-3.34	4.03
6	-0.07	0.05	-0.13	0.32	5.35
7	-0.25	0.1	-1.35	-6.41	-15.94
8	-0.05	0.17	-0.04	-1.42	13.36
9	-0.05	0.07	0.01	-0.15	8.73
10	-0.09	0.1	-0.24	-0.06	5.14
11	-0.11	0.07	-0.42	-2.34	0.87
12	-0.02	0.19	0.02	-1.8	18.79
13	-0.13	0.13	-0.56	-2.84	0.48
14	-0.12	0.1	-0.54	-1.47	0.17
15	-0.08	0.01	-0.22	-0.87	2.52
16	-0.02	0.12	-0.78	-1.16	16.15

Tabla 10: diferencia de los backtests realizados con el backtests de referencia para la estrategia SVM.

En base a la tabla 10 podemos ver como la mayoría de los indicadores de riesgo poseen valores peores que los del backtest de referencia. Todos los backtests -excepto el primero- dieron valores de alpha inferiores al del backtest de referencia aunque la diferencia en la mayoría de los casos es inferior a 0.1 con lo cual se los podría catalogar como insignificante.

Los valores de Beta son mayores en todos los *backtests* lo que indica que la estrategia es más susceptible a los movimientos del mercado. Esto se puede deber a que la estrategia incorpora para sus predicciones valores referentes al sector al que pertenecen. Si las predicciones tienen en cuenta el sector es de esperar que reaccionen acorde a los movimientos del mismo.

Respecto a los valores de *Sharpe Ratio* obtenidos vemos que para los *backtests* $id \in \{1,2\}$ presentan un aumento considerable respecto al *backtest* de referencia. Recordemos que este indicador es uno de los más importantes a la hora de evaluar el riesgo del *portfolio* ya que nos da una noción de que tan estables son los retornos que se obtienen a lo largo del tiempo. En especial destacamos los resultados del *backtest* $id = 1$ cuyo *Sharpe Ratio* es superior a 3 (ver sección 4.3) el cual es un valor excelente según la bibliografía⁸.

Intervalos de confianza al 99% del <i>hit rate</i> para cada uno de los <i>backtests</i> realizados			
id	Hit Rate (%)	Intervalo (-)	Intervalo (+)
1	50.84	50.13	51.55
2	50.61	49.59	51.63
3	51.2	49.97	52.43
4	51.85	50.03	53.67
5	50.78	50.07	51.49
6	50.82	49.80	51.84
7	50.8	49.57	52.03
8	51.15	49.33	52.97
9	50.38	49.67	51.09
10	51.15	50.13	52.17
11	50.57	49.34	51.8
12	51.23	49.41	53.05
13	50.92	50.21	51.63
14	51	49.99	52.02
15	51.05	49.82	52.28
16	51.46	49.64	53.28

Tabla 9: intervalos de confianza para los *backtests* realizados para la estrategia SVM.

⁸ Un *sharpe ratio* mayor a 1 se considera aceptable, mayor a 2 se considera muy bueno y mayor a 3 se lo considera como excelente.

Al analizar los intervalos de confianza para los valores de *hit rate* obtenidos podemos ver que para los *backtests* $id \in \{1, 4, 5, 10, 13\}$ todo el intervalo se encuentra por encima del 50% con lo cual podemos afirmar con un 99% de certeza que la estrategia predice los movimientos de las acciones con los parámetros correspondientes a dichos *backtests*.

Un aspecto interesante a destacar es que en algunos casos el *hit rate* obtenido no se condice necesariamente con un mejor rendimiento de la estrategia en general. Tomemos como ejemplo el *backtest* $id = 13$ anteriormente mencionado, si miramos su rendimiento respecto del *backtest* de referencia vemos que performa peor. Esto se puede deber a que el *hit rate* solo nos muestra los aciertos a la hora de predecir pero no nos dice nada sobre la calidad de dichos aciertos. No es lo mismo predecir que el valor de una acción va a subir y ésta incremente en un 1% a que incremente un 10%. Ambas predicciones son correctas y aportan de igual manera al cálculo del *hit rate*, aunque una predicción es claramente más significativa que la otra.

6. Conclusiones

Prestando atención a la estrategia MACD y comparando sus distintas configuraciones, observamos que el apalancamiento es un factor a tener en cuenta dentro de este tipo de estrategias, ya que permite operar con un mayor volumen y obtener un retorno más alto. El caso más evidente es la comparación entre la configuración 1 y 1'. La configuración sin apalancamiento no supera el 25% de retorno, mientras que la configuración con apalancamiento la duplica en rendimiento.

Para concluir con esta estrategia, podemos ver cómo un apalancamiento no tan pronunciado nos permite aumentar de forma considerable los retornos, a pesar de poseer un riesgo asociado al tener que utilizar dinero prestado. Vemos también, que la configuración 1' es la mejor de todas con un *hit rate* bastante elevado y ubicado cómodamente por encima del 50% considerando su intervalo de confianza.

Al mirar los resultados de la estrategia ARMA rápidamente vemos que fueron desalentadores en todos los aspectos. Concluimos entonces, que la estrategia -con los parámetros utilizados- no logra predecir los movimientos de las acciones y por lo tanto sería un error utilizarla en algún *portfolio* ya que degradará su rendimiento y aumentará el riesgo.

Respecto a la estrategia SVM, como se dijo anteriormente, vemos que en general los indicadores de rendimiento empeoran respecto del *backtests* de referencia a excepción del retorno. El concepto de “a mayor riesgo, mayor ganancia” se ve reflejado en estos resultados ya que se ve un claro incremento del retorno a expensas de un empeoramiento de los indicadores observados. Sin embargo podemos afirmar que para la mayoría de los casos presentes el aumento de riesgo es mínimo comparado con el incremento del retorno. En especial destacamos los resultados obtenidos en los *backtests* $id \in \{1, 2, 4, 8, 12\}$ los cuales aumentan en más del 10% del retorno del *portfolio* aumentando mínimamente el riesgo.

El concepto de riesgo depende de quien lo evalúe, los inversores analizan los indicadores dándole más importancia a algunos que a otros. En nuestra opinión los *backtests* con $id \in \{1, 2\}$ no solo aumentan el retorno sino que también disminuyen el riesgo total gracias al incremento del *sharpe ratio*.

Para finalizar con ésta sección, podríamos atrevernos a rechazar la hipótesis del mercado eficiente dado que se logró predecir de forma constante en el tiempo el futuro de las acciones. Esto es un indicio acerca de las ineficiencias del mercado, el cual la hipótesis niega. Para avanzar en este sentido sería necesario realizar predicciones y operaciones reales operando en un mercado real durante periodos más amplios y con un *portfolio* más diverso.

7. Referencias

- [1] Aragonés, J.R. and Mascareñas, J., 1994. La eficiencia y el equilibrio en los mercados de capital. *Análisis financiero*, 64 (1), pp.76-89.
- [2] Definiciones de términos financieros extraídas de: <https://www.investopedia.com/>
- [3] Enders, Walter, 2004. Stationary Time-Series Models. *Applied Econometric Time Series (Second ed.)*. pp. 48–107.
- [4] Pandit, Sudhakar M. and Wu, Shien-Ming, 1983. *Time Series and System Analysis with Applications*.
- [5] Box, George; Jenkins, Gwilym M.; Reinsel, Gregory C. 1994. *Time Series Analysis: Forecasting and Control (Third ed.)*.
- [6] Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), pp.716–723.
- [7] Cortes, Corinna and Vapnik, Vladimir N. 1995. Support-vector networks. *Machine Learning*, 20 (3), pp.273–297.
- [8] Saahil Madge. 2015. *Predicting Stock Price Direction using Support Vector Machines*. Princeton university press.
- [9] Appel, Gerald. 2005. *Technical Analysis Power Tools for Active Investors*.
- [10] Wallis, Sean A. 2013. Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics*, 20 (3), pp.178–208.