



Universidad Nacional Autónoma de México

Facultad de Contaduría y Administración

*Redes Neuronales Artificiales aplicadas al análisis
del Riesgo de Mercado: Caso de AMXL y AZTECA*

Tesis

Que para obtener el título de:

Licenciado en Contaduría

Presenta:

Raúl Axel Castillo Morales

Asesor:

Dr. Francisco López Herrera



Ciudad Universitaria, Cd. de Mx. 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A la Universidad Nacional Autónoma de México y a la Facultad de Contaduría y Administración, por brindarme los conocimientos y herramientas necesarios para mi desarrollo profesional.

A mi tutor, el Dr. Francisco López, quien me brindó su atención, paciencia, asesoría, y cuya dirección me ayudó a hacer posible este trabajo.

A los profesores, por su dedicación y entrega en cada una de sus clases.

A mis amigos por su amistad y apoyo constante.

Y por sobre todo, a mi familia, por brindarme su cariño, apoyo incondicional y siempre alentarme a seguir superándome.

Índice General

Introducción	9
Resumen.....	15
Capítulo 1. Valor en Riesgo y Redes Neuronales	17
1.1. Riesgo.....	19
1.1.1. <i>Volatilidad</i>	21
1.1.2. <i>Modelos de Rango</i>	29
1.2. Valor en Riesgo	32
1.2.1. <i>Metodologías</i>	35
1.3. Inteligencia Computacional.....	45
1.3.1. <i>Redes Neuronales</i>	50
1.4. Redes Neuronales aplicadas al pronóstico de volatilidad y al análisis del Riesgo de Mercado.....	59
Capítulo 2. Metodologías para el análisis del Riesgo de Mercado de una acción	63
2.1. VaR usando modelo propuesto de pronóstico	65
2.1.1. <i>Predicción de volatilidad usando Redes Neuronales</i>	65
2.2. Modelos de VaR Tradicionales	84
2.2.1. <i>Simulación Histórica con alisado</i>	84
2.2.2. <i>Simulación Monte Carlo</i>	85
2.3. Pruebas de Backtesting.....	86
Capítulo 3. Aplicación del Valor en Riesgo en las acciones mexicanas: Caso de América Móvil y TV Azteca	89
3.1. Metodología mediante Redes Neuronales	91
3.1.1. <i>Análisis y preprocesamiento de datos</i>	91
3.1.2. <i>Pronóstico de Volatilidad</i>	99
3.1.3. <i>VaR paramétrico mediante pronóstico de volatilidad</i>	100

3.2. VaR mediante Simulación Histórica con alisado	101
3.3. VaR mediante Simulación Monte Carlo	102
Conclusiones.....	107
Anexos	111
Referencias	117

Índice de Tablas

Tabla 1. Días de datos históricos para un Nivel de Tolerancia	24
Tabla 2. Distribución Normal Acumulativa	36
Tabla 3. Tabla de Kupiec.	44
Tabla 4. Zonas de Penalización de Basilea	45
Tabla 5. Estadística Descriptiva de los rendimientos y rangos logarítmicos.	92
Tabla 6. Resumen de la Eficiencia de los Estimadores de	94
Tabla 7. Variables independientes.....	95
Tabla 8. VIFs de las variables independientes de AMXL y AZTECA	96
Tabla 9. Resumen de importancia de los Componentes Principales (AMXL)	97
Tabla 10. Resumen de importancia de los Componentes Principales (AZTECA)	97
Tabla 11. Evaluación de Pronósticos de Volatilidad.....	100
Tabla 12. Pruebas de Backtesting (AMXL)	102
Tabla 13. Pruebas de Backtesting (AZTECA)	103
Tabla 14. Número de excesos.....	103

Índice de Figuras

Figura 1. Principales enfoques de la Inteligencia Computacional	47
Figura 2. Representación biológica de una “neurona básica” y su equivalente computacional.....	52
Figura 3. Modelo de neurona perceptrón.....	53
Figura 4. Ejemplo de Red FeedForward	57
Figura 5. Ejemplo de Red Neuronal Recurrente.....	59
Figura 6. RNN estándar y su cómputo hacia adelante	67
Figura 7. Diagrama del estado oculto de una RNN estándar y una red LSTM	70
Figura 8. Ejemplo de k-CV aplicado de forma secuencial	80
Figura 9. Comparativo de estimadores de volatilidad histórica de AMXL ..	93
Figura 10. Comparativo de estimadores de volatilidad histórica de AZTECA	93
Figura 11. Matriz de correlación de las variables independientes	96
Figura 12. Validación de Componentes con el menor RMSE (AMXL).....	98
Figura 13. Validación de Componentes con el menor RMSE (AZTECA)...	98

Introducción

La necesidad humana de responder a la incertidumbre se ha dado principalmente en forma cuantitativa, esto nos ha llevado al desarrollo de modelos estadísticos que han permitido gestionar el riesgo de que los resultados futuros no sean conforme a los estimados en un principio. La “desviación estándar” ha sido el modelo base para estimar, en cierta medida, dicha incertidumbre en los mercados financieros; a este modelo (en el área financiera) comúnmente se le conoce como la “volatilidad” de un instrumento, y su pronóstico es de gran interés tanto teórico como práctico. A su vez, es una de las principales variables de análisis en modelos de administración de riesgos, gestión de portafolios de inversión, valuación de instrumentos derivados, y estrategias de cobertura. Por ello, investigadores y analistas buscan crear modelos que generen un mejor estimado futuro de la misma.

Además, debido a las crisis financieras que han ocurrido en los últimos años (e.g., la Crisis “Punto-Com” en el año 2000, la “Gran Recesión” en el 2008, la crisis de la Deuda Soberana Griega en 2009, entre otras) y el desarrollo de nuevos productos e instrumentos de inversión (volviéndose cada vez más complejos con el objetivo de obtener mayores ganancias, diversificar el riesgo, o especular sobre el movimiento del mercado), la correcta administración del riesgo financiero se ha vuelto un tema de vital importancia en las empresas e instituciones que se encuentran expuestas a eventualidades (i.e., eventos con repercusiones tanto económicas como financieras) que resulten en pérdidas potenciales. Es por lo anterior que, a la Dirección de Finanzas, ya sea en empresas o en instituciones, le concierne el desarrollo de metodologías cada vez más avanzadas que permitan cuantificar dicho riesgo de manera correcta y así, las instituciones que se encuentran expuestas puedan tomar las medidas

adecuadas para mitigarlo. Y en el sector financiero en particular, esta tarea le correspondería a la Unidad de Administración Integral de Riesgos, la cual responde directamente al Consejo de Administración (en algunos casos se encuentra bajo la Dirección de Finanzas), y así, éste último pueda tomar las medidas pertinentes con respecto a la situación actual de la compañía.

El mayor impacto respecto a la administración de riesgos financieros ocurrió al darse a conocer el modelo “Value-at-Risk” (Valor en Riesgo, o VaR), en el documento técnico *RiskMetrics* (1996), publicado por el banco estadounidense JP Morgan. Uno de los grandes beneficios que trajo dicho modelo fue que permitió tener una cifra monetaria respecto a la posible pérdida que podría generarse en un horizonte de tiempo dado, lo cual resultó en su fácil entendimiento conceptual, aún para los tomadores de decisiones sin conocimientos técnicos sobre el tema. A su vez, es un modelo ampliamente aceptado a nivel mundial, al punto que los reguladores bancarios de diversos países requieren o promueven su uso, como es el caso del Comité de Basilea que propuso en los años 90’s su adopción con ciertas reglas para su implementación: considerar un horizonte de 10 días y un nivel de confianza del 95%; aunque, también permitía que cada institución usará su propio modelo siempre que cumpliera con los límites establecidos en las pruebas de backtesting.

Si bien, el VaR ha probado ser un método muy versátil que permite la evaluación de instrumentos individuales y portafolios, así como el tratamiento de una gran variedad de riesgos, aún es un modelo que depende en gran medida del juicio subjetivo del administrador de riesgos, tanto en la manera de analizar los datos de los que dispone, así como en la metodología a utilizar para su cálculo. Entre los diversos métodos que existen para calcular el VaR en el

presente trabajo se abordarán los siguientes: a) Delta-Normal, b) Simulación Histórica sin alisado (este modelo no se utilizará como modelo base de comparación), c) Simulación Histórica con alisado, y d) Simulación Monte Carlo. Con éstos, se realizará una aplicación al análisis del riesgo de mercado de las acciones de América Móvil y TV Azteca por el período de enero 2005 a marzo 2017.

Es por lo anterior y gracias al rápido desarrollo tecnológico, que nuevas metodologías para el pronóstico y análisis de series de tiempo se han considerado para resolver problemas concernientes al área de Finanzas. Tal es el caso de los modelos de “autoaprendizaje” o “aprendizaje de máquina” (así llamados debido a su autocorrección y auto calibración de parámetros). Dichos modelos fueron creados por investigadores de las áreas de Inteligencia Artificial e Inteligencia Computacional. Debido a estos modelos, en años recientes se ha podido dar un mejor uso a los datos disponibles y, actualmente, algunas instituciones e investigadores buscan mejorar e innovar los modelos existentes en Finanzas haciendo de estos modelos ya sea en problemas de regresión o de clasificación. El tema central del presente trabajo será analizar su aplicabilidad en un problema de regresión de series de tiempo, donde la información histórica toma relevancia en la ocurrencia de hechos futuros, y para ello, los modelos de Redes Neuronales Artificiales se han considerado ser los ideales para resolver este problema.

En tal sentido, esta tesis tiene la intención de realizar una valuación del VaR mediante el pronóstico de la volatilidad de una acción haciendo uso de herramientas de Inteligencia Artificial, en concreto, de Redes Neuronales Artificiales, y corroborar si el uso de estos modelos contribuye de forma positiva a la detección temprana de pérdidas potenciales en el mercado.

Asimismo, el pronóstico se realizará para distintos modelos de volatilidad (desviación estándar y modelos de rango) con el fin de comparar su nivel de estimación futura bajo el argumento de que para aquellos estimadores de volatilidad cuyos cambios sean más fáciles de pronosticar, serán de mayor utilidad en la práctica.

Siendo así, la hipótesis general del presente trabajo plantea que, a través de la implementación de un modelo de Redes Neuronales Artificiales es posible obtener un mejor pronóstico mensual de la volatilidad de un instrumento financiero, lo que conducirá a un mejor estimado de la máxima pérdida esperada, en contraste con las metodologías de VaR más usadas (Simulación Histórica y Simulación Monte Carlo). Y se deriva como hipótesis específica que el uso de otros modelos de volatilidad, como los modelos de rango, permitirá obtener un mejor modelo de pronóstico al considerar más datos dentro del movimiento del precio del instrumento y no sólo los valores al cierre.

Dentro de las limitantes de la investigación realizada en el presente trabajo se encontraron las siguientes:

- i) Datos históricos. En algunas de las series de datos que se obtuvieron para esta investigación, existe una falta de información significativa tanto para los datos diarios como mensuales;
- ii) Diseño de la red neuronal. Diseñar una red neuronal más compleja (i.e., agregar capas ocultas y/o células de procesamiento) implicaría un mayor costo computacional o el uso de equipos más sofisticados, aunque, en el presente trabajo, se experimentó con algunas modificaciones a la red (e.g., agregar o quitar número de iteraciones, células de procesamiento, y utilizar de 1 a 3 capas ocultas), encontrando que el modelo empleado en el presente trabajo fue uno de los que mejor desempeño tuvo.

Finalmente, esta tesis aporta la aplicación metodológica de los estudios adquiridos del área de Finanzas en la Carrera de Contaduría y, centrándose en la importancia que ha tenido el desarrollo tecnológico-estadístico, pretende aportar apoyo en el desarrollo de modelos que optimicen la gestión de los recursos y la exposición al Riesgo de Mercado al que están sujetas tanto de instituciones financieras como empresas; del mismo modo, operadores y asesores de inversión independientes podrían beneficiarse por el desarrollo del presente trabajo.

Esta investigación consta de tres capítulos. En el primer capítulo se aborda el tema de administración del riesgo financiero, comenzando con la definición del riesgo y el riesgo financiero en las empresas, siguiendo con una breve explicación de los modelos de volatilidad más conocidos, y el desarrollo del marco teórico del Valor en Riesgo aplicado en el Riesgo de Mercado así como sus metodologías (modelos paramétricos y no paramétricos); a su vez, se realiza una introducción conceptual del área de Inteligencia Computacional, sus subáreas de desarrollo y se da una explicación de los fundamentos de las Redes Neuronales Artificiales y los modelos que se han desarrollado; se concluye con las investigaciones de redes neuronales que se han aplicado a los temas de gestión del Riesgo de Mercado y pronóstico de volatilidad.

En el segundo capítulo se desarrolla la metodología empleada en el presente trabajo, partiendo desde la explicación del funcionamiento de los modelos de redes neuronales recurrentes, siguiendo con la justificación del modelo de redes neuronales seleccionado y el detalle de su cálculo. Después, continúa con la descripción de las variables seleccionadas (dependientes e independientes) a ingresar al modelo de pronóstico, al igual que su tratamiento. Asimismo, se da

una breve explicación de los modelos de Simulación Histórica con alisado y Simulación Monte Carlo, que serán utilizados como base de comparación respecto al modelo desarrollado en el presente trabajo. Y el capítulo culmina con la explicación de las pruebas de backtesting a emplear para evaluar el desempeño de los modelos de pronóstico y poder hacer una comparación entre éstos.

En el tercer capítulo, se presenta un breve análisis de las series de rendimientos y volatilidades de los instrumentos de América Móvil y TV Azteca, por el período de enero 2005 a marzo 2017. Posteriormente, se muestra el proceso de la selección y preprocesamiento de las variables independientes, partiendo del detalle de las variables seleccionadas, siguiendo con un análisis correlacional y de multicolinealidad entre los datos a modo de verificar el correcto uso del Análisis de Componentes Principales. Por último, se presentan los resultados obtenidos de los pronósticos realizados para los distintos modelos de volatilidad (desviación estándar y modelos de rango), así como los resultados obtenidos de la aplicación de las metodologías de Valor en Riesgo, del mismo modo que una comparación entre dichas metodologías.

Al final, se presentan las conclusiones obtenidas y recomendaciones derivadas de la presente tesis.

Resumen

Esta tesis muestra la investigación que se realizó para corroborar la eficiencia de las de Redes Neuronales Artificiales aplicadas al tema de Análisis del Riesgo de Mercado y al pronóstico de distintos modelos de volatilidad, teniendo como objetivo principal encontrar un modelo eficiente que permitiera tener una mejor estimación mensual del Valor en Riesgo, en contraste con los otros modelos normalmente usado para la gestión de riesgos: Simulación Monte Carlo y Simulación Histórica con alisado.

Debido a los hallazgos obtenidos, se rechaza la hipótesis principal dado que el pronóstico de datos no fue cercano a los valores y el número de excesos fue mayor a los aceptados por las pruebas de backtesting. Si bien, fue posible detectar una tendencia en los datos, así como anticipar algunos de los movimientos que presentaban los mismos, no superaron a los modelos base de comparación y, por el contrario, se demostró que no era una manera eficiente para estimar la máxima pérdida esperada de una acción para un horizonte de tiempo mensual.

Capítulo 1. Valor en Riesgo y Redes Neuronales

*Sin embargo, ninguna queja es más común
que la de una escasez de dinero.*

Adam Smith

1.1. Riesgo

La palabra “riesgo” tiene su origen del latín *risicare*, siendo originalmente una palabra náutica cuyo significado era roca peligrosa o precipicio (Donkin, 1864), asimismo, ésta es definida por la Real Academia Española (2014) del antiguo *riesco* (risco, por el peligro que suponen), entendiéndose como una contingencia o proximidad de un daño. Es por ello que, en Finanzas, esta palabra se entiende como el peligro que existe ante cambios en el valor de los activos financieros; mientras que Jorion (2004) la define como la volatilidad de los flujos financieros no esperados, derivada del valor de los activos y pasivos. El riesgo surge de la incertidumbre respecto de las futuras pérdidas y ganancias de una entidad. Si bien existe una relación de compensación entre el riesgo y el rendimiento, es de mayor importancia el análisis de la variabilidad de la pérdida, especialmente cuando ésta llega a niveles no previstos.

La medición de riesgos contempla el uso de probabilidad y estadística, donde a partir de una distribución de frecuencias trata de explicar el comportamiento histórico de los rendimientos de algún activo. Los parámetros más importantes que definen a una distribución son la media, representada por μ , y la desviación estándar (dispersión de datos alrededor de la media), representada por σ . En materia de portafolios de inversión, la media es el rendimiento promedio, y la desviación estándar es definida como la volatilidad del activo. La medición de dichos parámetros puede ser expresada considerando los datos una vez que ha ocurrido el evento (*ex-post*), o tomando en cuenta los datos esperados previo al evento (*ex-ante*):

$$\begin{array}{c} \textit{Ex-post} \\ \mu = \frac{\sum_{i=1}^m r_i}{m} \end{array} \quad (1.1)$$

$$\begin{array}{c} \textit{Ex-ante} \\ E[\mu] = \sum_{i=1}^m p_i r_i \end{array} \quad (1.3)$$

$$\sigma_{SD} = \left[\frac{\sum_{i=1}^m (r_i - \mu)^2}{m-1} \right]^{1/2} \quad (1.2) \quad \sigma_{SD} = [\sum_{i=1}^m p_i (r_i - E[\mu])^2]^{1/2} \quad (1.4)$$

Donde,

r_i es el rendimiento del activo i definido como: * $r_i = \ln \left(\frac{\text{Precio}_t}{\text{Precio}_{t-1}} \right)$;

p_i es la probabilidad de ocurrencia;

m es el número de observaciones (días).**

Existen diferentes naturalezas de riesgos, y de conformidad con la normatividad del sistema financiero mexicano, en el Artículo 66 de la Circular Única de Bancos (CUB) (2005), éstos se pueden clasificar como: a) Riesgos cuantificables, aquellos para los cuales sea posible medir sus pérdidas potenciales mediante bases estadísticas; y, b) Riesgos no cuantificables, aquellos derivados de eventos imprevistos para los cuales no sea posible medir las pérdidas potenciales. A su vez, los Riesgos cuantificables se subdividen en Discrecionales (resultantes de la toma de una posición de riesgo; e.g., riesgo de crédito, de liquidez, o de mercado) y No Discrecionales (resultantes de la operación del negocio, pero que no son producto de la toma de una posición de riesgo; e.g., riesgo operativo, tecnológico, o legal). Esta tesis se enmarca en el análisis del Riesgo de Mercado, el cual Alfonso de Lara (2013) define como la pérdida que puede sufrir un inversionista debido a la diferencia en los precios que se registran en el mercado o en movimientos de los llamados factores de riesgo (tasas de interés, tipos de cambio, etc.).

* En riesgos usualmente se usa el rendimiento logarítmico, que da un resultado casi similar a $r_i = \frac{\text{Precio}_t}{\text{Precio}_{t-1}} - 1$; empleado para el monitoreo de la volatilidad diaria (Hull, 2012).

** El Banco Internacional de Liquidaciones (BIS) recomienda considerar un horizonte de 250 días de operación (hábiles), equivalentes a un año calendario.

1.1.1. Volatilidad

Hull (2012) define la volatilidad como la desviación estándar de los rendimientos de una variable por unidad de tiempo, cuando el rendimiento es expresado con interés compuesto continuo. La volatilidad es un indicador fundamental en la medición de riesgos, puesto que representa la dispersión de los rendimientos respecto a su media. Por ello, el análisis de la volatilidad y el diseño de modelos para su pronóstico juegan un papel central para la medición del riesgo de mercado. Para el monitoreo diario de la volatilidad la fórmula de la varianza puede ser simplificada bajo los supuestos (GARP, 2016):

- a) r_i es definido como el cambio porcentual en la variable del mercado entre el final del día $i - 1$ y el final del día i

$$r_i = \frac{S_i}{S_{i-1}} - 1;$$

- b) μ es igual a cero. La justificación para este supuesto es que el cambio esperado de una variable en un día es muy pequeño comparado con la desviación estándar de los cambios, por lo que tiene poco efecto en los estimados de la varianza;
- c) $m - 1$ es reemplazado por m . Esto para cambiar el estimado imparcial de volatilidad por uno de máxima verosimilitud.

Dado que estos cambios hacen una pequeña diferencia para el estimado de volatilidad, podemos simplificar la fórmula de la varianza diaria considerando únicamente el cuadrado de los rendimientos, siendo σ_n^2 la varianza diaria de la variable del mercado en el día n , estimada al final del día $n - 1$, con las observaciones m más recientes de r_{n-i}

$$\sigma_n^2 = \frac{1}{m} \sum_{i=1}^m r_{n-i}^2 \quad (1.5)$$

En el documento técnico *RiskMetrics* (JP Morgan & Co., 1996) usan análisis histórico de series de tiempo para derivar el estimado de volatilidad, por lo que asumen que la distribución de rendimientos pasados puede ser modelada para tener un pronóstico razonable de éstos en un horizonte de tiempo.

Volatilidad Histórica

En este método el pronóstico se basa en las observaciones históricas y que éstas tienen el mismo peso específico. En la ecuación (1.5) se le da el mismo peso a todas las variables $r_{n-1}^2, r_{n-2}^2, \dots, r_{n-m}^2$. Debido a que las series de tiempo en finanzas son de memoria “corta” (las personas tienden a darle mayor peso a los hechos más recientes), esta metodología no es muy útil en la práctica profesional (De Lara, 2013). Esto ha motivado al mercado a aplicar otros métodos como son los modelos EWMA y GARCH.

Volatilidad Implícita

La volatilidad implícita se refiere al parámetro σ utilizado en la fórmula de valuación de derivados de Black-Scholes (ecuación 1.6), en donde éste es el valor que da como resultado el precio de mercado de la opción. El uso de este modelo proviene de la idea de que la volatilidad implícita es aquella volatilidad que los mercados de opciones “estiman” se realizará. Su confiabilidad depende de la liquidez del mercado del subyacente, lo cual no siempre se cumple en la práctica, ya que no todos los instrumentos tienen contratos de opciones y tampoco son todos líquidos (Alatorre, 2015), aunado al hecho de que no existe una manera analítica para despejar el parámetro σ .

$$c = SN(d_1) - Ke^{-rT}N(d_2) \quad (1.6)$$

Siendo,

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \quad d_2 = d_1 - \sigma\sqrt{T}$$

Donde,

S es el valor del bien subyacente;

K es el precio de ejercicio de la opción;

r es la tasa libre de riesgo;

T es el período de la opción;

σ es la volatilidad implícita del bien subyacente;

$N(\cdot)$ es la distribución normal estándar acumulativa, es decir, las probabilidades ajustadas por el riesgo.

Modelo de Volatilidad Dinámica

El modelo de volatilidad dinámica o modelo de media móvil exponencialmente ponderada (EWMA, por sus siglas en inglés), asigna mayor peso a las observaciones más recientes contemplando, generalmente, un período de entre 6 meses hasta 2 años. Debido a la ponderación de los datos se obtienen mejores resultados en épocas de alta volatilidad y un mejor estimado de volatilidad en un día, puesto que se capturan rápido las altas variaciones (De Lara, 2013). Partiendo del supuesto de que la media de los rendimientos es igual a cero, a la ecuación (1.5) se le agrega un peso específico w para cada observación i días anteriores:

$$\sigma_n^2 = \sum_{i=1}^m w_i r_{n-i}^2 \quad (1.7)$$

Los pesos w deben sumar la unidad y, si $w_i = \lambda^{i-1}(1 - \lambda)$, donde $0 < \lambda < 1$. Entonces la varianza se define como

$$\sigma_n^2 = (1 - \lambda) \sum_{i=1}^m \lambda^{i-1} r_{n-i}^2 \quad (1.8)$$

Donde λ es el factor de descuento que se utiliza para determinar los pesos que se aplicarán a las observaciones y la cantidad efectiva de datos que se

emplearán para estimar la volatilidad. También se puede expresar mediante una ecuación de tipo recursiva (De Lara, 2013)

$$\sigma_n^2 = (1 - \lambda)r_{n-1}^2 + \lambda\sigma_{n-1}^2 \quad (1.9)$$

Así, mientras retrocedemos en los días de observación, los pesos disminuyen a un valor λ , por lo que cada peso es λ veces el peso anterior. Un elevado λ provocará que los estimados de volatilidad respondan lentamente a la nueva información, y un bajo valor de λ resultará en que los estimados de la volatilidad para días posteriores sean altamente volátiles. Con un Nivel de Tolerancia (medida que permite saber para cierto λ , cuántas observaciones se consideran para el cálculo de la volatilidad), dado por $NT = \lambda^k$, las observaciones históricas que recoge este modelo, con base en el documento *RiskMetrics* (JP Morgan & Co., 1996), son:

Tabla 1. Días de datos históricos para un Nivel de Tolerancia

λ	0.001%	0.01%	0.1%	1%
0.90	109	87	66	44
0.95	224	180	135	90
0.96	282	226	169	113
0.97	378	302	227	151
0.98	570	456	342	228
0.99	1146	916	687	458

Fuente: RiskMetrics (JP Morgan & Co., 1996).

Para determinar un valor λ óptimo (λ^*) existe el método de “Raíz del Error Medio Cuadrático” (RMSE, por sus siglas en inglés) (De Lara, 2013). Este método permite obtener una λ^* que minimice el error pronosticado de σ^2

$$RMSE_{\lambda^*} = \sqrt{\frac{1}{m} \sum_{i=1}^m [r_{i+1}^2 - \hat{\sigma}_{i+1|i}^2(\lambda)]^2} \quad (1.10)$$

En general esta metodología tiene la ventaja de que no requiere de una gran capacidad de almacenamiento de información y tampoco establece supuestos explícitos sobre la distribución de los rendimientos del activo.

Modelos ARCH y GARCH

En lo referente a series de tiempo, un supuesto básico del modelo clásico de regresión lineal es el de *homoscedasticidad* (*homo*-igual; *cedasticidad*-dispersión), el cual menciona que la varianza de cada término de perturbación ε_i es un número constante igual a σ^2 (Gujarati, 2010), es decir, que la varianza del término de error es la misma para todas las observaciones:

$$E(\varepsilon_i^2) = \sigma_\varepsilon^2, \quad i = 1, \dots, n$$

Sin embargo, esto no siempre se cumple. En el mercado se presentan datos atípicos (un dato atípico es una observación que es muy diferente en relación con las demás observaciones en la muestra) (Gujarati, 2010), por lo que la varianza de la serie tendrá cambios sistemáticos en el tiempo (es común observar períodos de alta turbulencia seguidos de periodos de calma), a este fenómeno se le denomina *heteroscedasticidad*. Las implicaciones básicas en un modelo de volatilidad son que ésta es heteroscedástica y depende de sí misma, por lo que es persistente (autorregresiva); hay agrupamiento de la volatilidad o “*volatility clustering*” (captura períodos de turbulencia y períodos de calma) y; la volatilidad tiende a regresar a la media (*mean-reversion*) (De Lara, 2013). Para dicha situación se emplean los modelos ARCH (heteroscedasticidad condicional autorregresiva) y GARCH (ARCH generalizado).

El modelo ARCH está diseñado especialmente para modelar y pronosticar la volatilidad, éste fue primeramente sugerido por Robert Engle (1982) y

extendido a GARCH por Bollerslev (1986) (como fueron referenciado por Alatorre, 2015). Dicho modelo está basado en la especificación de densidades condicionales en períodos sucesivos con un proceso de volatilidad dependiente del tiempo. Para el modelo ARCH(m) tenemos una extensión de la ecuación (1.7), donde asumimos que existe una varianza promedio a largo plazo y que se le debería asignar algún peso específico. Así, en este modelo el estimado de la varianza está basado por una varianza promedio a largo plazo y m observaciones. A mayor antigüedad, menor peso se le asigna a la observación. Por lo tanto, el modelo queda como:

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^m w_i r_{n-i}^2 \quad (1.11)$$

Donde,

V_L es la varianza promedio a largo plazo y;

γ es el peso asignado a la varianza a largo plazo;

Y los pesos $\gamma + \sum_{i=1}^m w_i$ deben sumar la unidad.

En el modelo GARCH(1,1), se define el rendimiento de un activo como $r_n = \mu + \varepsilon_n$, siendo ε_n un ruido blanco con varianza cero que, a pesar de no estar serialmente correlacionado, la serie ε_n no necesita ser serialmente independiente (The Volatility Institute NYU, 2013). Este modelo asume una forma paramétrica específica para la heteroscedasticidad condicional que puede presentar. Entonces, decimos que $\varepsilon_n \sim GARCH$ si

$$\varepsilon_n = \sigma_n z_n, \quad z_n \sim^{IID} N(0,1)$$

Denotando que $\omega = \gamma V_L$ tenemos que

$$\sigma_n^2 = \omega + \alpha \varepsilon_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (1.12)$$

Donde,

α es el peso asignado a ε_{n-1}^2 ;

β es el peso asignado a σ_{n-1}^2 .

De esta manera, se estiman los parámetros (ω, α, β) simultáneamente mediante métodos de máxima verosimilitud.*** Una vez calculados, podemos obtener γ como $1 - \alpha - \beta$; y V_L se puede obtener de dividir ω entre γ . Derivado de lo anterior, observamos que el modelo captura las implicaciones antes mencionadas: si la volatilidad en el tiempo n es alta, entonces también será elevada en $n + 1$ (*volatility clustering*); asimismo, para tener un proceso GARCH(1,1) estable es requerido que $\alpha + \beta < 1$ evitando tener una ω negativa; al igual que los pronósticos de volatilidad tenderán a su media en el mediano y largo plazo (*mean-reversion*); y al estar muy cercanos a 1 (*persistencia en la volatilidad*), ésta fluctuará alrededor de σ , la raíz de la varianza no condicionada:

$$\sigma^2 = \frac{\omega}{1-\alpha-\beta}$$

Si sustituimos para σ_{n-1}^2 en la ecuación (1.12) resulta

$$\sigma_n^2 = \omega + \beta\omega + \alpha\varepsilon_{n-1}^2 + \alpha\beta\varepsilon_{n-2}^2 + \beta^2\sigma_{n-2}^2$$

Y para σ_{n-2}^2

$$\sigma_n^2 = \omega + \beta\omega + \beta^2\omega + \alpha\varepsilon_{n-1}^2 + \alpha\beta\varepsilon_{n-2}^2 + \alpha\beta^2\varepsilon_{n-3}^2 + \beta^3\sigma_{n-3}^2$$

Por lo que obtenemos

$$\sigma_n^2 = \omega \sum_{i=1}^m \beta^{i-1} + \alpha \sum_{i=1}^m \beta^{i-1} \varepsilon_{n-i}^2 + \beta^m \sigma_{n-m}^2 \quad (1.13)$$

*** Si asumimos que la distribución de probabilidad de ε_i condicional a la varianza es normal y, al tomar en consideración las m observaciones en el orden en que ocurren, sacando los logaritmos de la ecuación resultante tenemos la siguiente expresión a maximizar:

$$L_m = \sum_{i=1}^m \left[-\ln(\sigma_i^2) - \frac{\varepsilon_i^2}{\sigma_i^2} \right] \quad (\text{Hull, 2012}).$$

Como podemos observar en este modelo, β es similar a λ (en el modelo EWMA) actuando como factor de descuento, mientras que α sería igual a $(1 - \lambda)$, y ω tendría un valor de cero. Por lo tanto, el modelo EWMA es similar a una forma particular del modelo GARCH, conocido como IGARCH (GARCH Integrado) (The Volatility Institute NYU, 2013). El modelo GARCH(1,1) también puede ser utilizado para estimar la volatilidad para opciones desde datos históricos (Hull, 2012).

Sin embargo, de acuerdo con lo mencionado en *RiskMetrics* (1996), West y Cho encontraron que los modelos GARCH no superan significativamente los estimados de volatilidad igualmente ponderada en pronósticos fuera de la muestra, excepto para horizontes de tiempo muy cortos; y además, Heynen y Kat mostraron en su investigación que los modelos GARCH tienen mejor habilidad de predicción para tasas de tipo de cambio que para rendimientos de acciones, aunado a esto, cuando se considera un período de observación mayor a 20 días, la ventaja sobre una estimación de caminata aleatoria simple, desaparece.

Debido a lo anterior y sumando otras restricciones que presentan los gestores de riesgo, en *RiskMetrics* optaron por utilizar el modelo EWMA (JP Morgan & Co., 1996): "... la volatilidad es definida como la desviación estándar de un proxy de una distribución normal para observaciones diarias históricas exponencialmente ponderadas con factores de descuento del 0.94 (para datos diarios, 74 días de límite 1%) y 0.97 (para datos mensuales, 151 días de límite 1%)". La compañía obtuvo como resultados que, utilizando esos valores para λ a través de un rango de diferentes variables del mercado, obtenían un pronóstico de la varianza cercano a la varianza histórica.

1.1.2. Modelos de Rango

Los modelos de estimación previos sufren de la limitante que sólo utilizan los precios de cierre y no contemplan el movimiento dentro del periodo de referencia. Dicha situación la señalan Alizadeh, Brandt, Diebold, Chou, y otros autores (como fueron referidos por Chou, Chou, & Liu, 2009) en sus investigaciones, demostrando que el modelo GARCH de esta manera no es exacto y, por lo tanto, se vuelve ineficiente. Así entonces, surgen como alternativa a la desviación estándar los modelos de volatilidad de rango, los cuales son capaces de capturar este movimiento ya que toman en cuenta los precios de apertura (O), máximo (H), mínimo (L), y cierre (C) del periodo de referencia (dichos precios es común verlos en las velas japonesas así como en otras técnicas empleadas en el análisis técnico). En general, estos modelos dependen del supuesto de que los precios siguen un *Movimiento Geométrico Browniano*[†] en tiempo continuo.

En los últimos años, se han diseñado distintos estimadores de volatilidad cuya base es cubrir los rangos existentes (diferencial bid-ask, saltos de apertura, saltos durante la noche, rebote por no-operación, entre otros), para efectos de esta tesis únicamente se consideraron los modelos más influyentes en este tema, los cuales han demostrado ser capaces de generar resultados superiores, en el pronóstico de la volatilidad, a los obtenidos por medio de la desviación estándar (dos Santos Maciel & Ballini, 2017).

Cabe señalar que, en términos de eficiencia teórica (que requiere una menor cantidad de observaciones para obtener la misma precisión estadística), estos

[†] Proceso estocástico (sistema probabilístico que tiene una evolución aleatoria a lo largo de otra variable, que normalmente es el tiempo) en el que el logaritmo natural de una variable aleatoria sigue un *Proceso de Wiener*, donde la media y la varianza son proporcionales a la variable.

estimadores presentan mejoras en contraste con el estimado “clásico” de desviación estándar. Se entiende por eficiencia como el máximo desempeño teórico frente a una distribución idealizada y, se define la medida de eficiencia de un estimador de volatilidad $\hat{\sigma}_i^2$, como la razón de la varianza del estimador clásico σ_{SD}^2 entre la varianza de un estimador $\hat{\sigma}_i^2$

$$Eff(\hat{\sigma}_i^2) = \frac{var(\sigma_{SD}^2)}{var(\hat{\sigma}_i^2)} \quad (1.14)$$

Parkinson

Parkinson (1980) fue uno de los primeros en proponer un estimado de volatilidad que utilizara el rango de precios (la diferencia entre el precio máximo y mínimo) con el fin de reflejar toda la información incluida en las fluctuaciones del precio. Su modelo parte del supuesto de que el logaritmo de los precios sigue una caminata aleatoria con tendencia cero; éste reportó una relativa eficiencia que va de 2.5 a 5 veces superior respecto del estimado tradicional de desviación estándar.

$$\sigma_{Pk} = \sqrt{\frac{1}{4m \cdot \ln 2} \sum_{i=1}^m (\ln H_i - \ln L_i)^2} \quad (1.15)$$

Garman-Klass

Con base en el trabajo realizado por Parkinson, Garman y Klass (1980) diseñaron un modelo de rango utilizando todos los precios (OHLC) disponibles para cualquier inversionista. Consideraron que utilizar 4 puntos de datos les daría información adicional; obtuvieron una eficiencia de aproximadamente 7.4.

$$\sigma_{GK} = \sqrt{\frac{1}{m} \sum_{i=1}^m \frac{1}{2} \left(\ln \frac{H_i}{L_i} \right)^2 - (2 \ln(2) - 1) \left(\ln \frac{c_i}{o_i} \right)^2} \quad (1.16)$$

Rogers-Satchell

Dado que los estimados anteriores asumían que existía una tendencia cero en el movimiento de los precios logarítmicos (ese supuesto los lleva a sobreestimar la volatilidad cuando la tendencia es diferente de cero), Rogers y Satchell (1991) agregaron un término de tendencia en el proceso estocástico. La eficiencia que obtuvieron fue de 6 para el caso de la tendencia cero y superior a 2 para cualquier tendencia.

$$\sigma_{RS} = \sqrt{\frac{1}{m} \sum_{i=1}^m \ln\left(\frac{H_i}{C_i}\right) \ln\left(\frac{H_i}{O_i}\right) + \ln\left(\frac{L_i}{C_i}\right) \ln\left(\frac{L_i}{O_i}\right)} \quad (1.17)$$

Extensión de Garman-Klass Yang-Zhang

Yang y Zhang modificaron el modelo de Garman-Klass (Bennet & Gil, 2012) para que éste incorporara saltos, sin embargo, aún asume una tendencia cero. Este estimado presentó una relativa eficiencia de 8.

$$\sigma_{GKYZ} = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(\ln\frac{O_i}{C_{i-1}}\right)^2 + \frac{1}{2} \left(\ln\frac{H_i}{L_i}\right)^2 - (2 \ln(2) - 1) \left(\ln\frac{C_i}{O_i}\right)^2} \quad (1.18)$$

Yang-Zhang

Posteriormente, Yang y Zhang (2000) propusieron un modelo que además de incorporar una tendencia, manejaba los saltos de apertura. Su modelo es un estimador de volatilidad multiperiodo que consiste en la suma de la volatilidad que ocurre durante la noche más, el promedio ponderado de la volatilidad de Rogers-Satchell y la volatilidad de apertura-a-cierre. Se caracteriza por ser aquel con la menor varianza entre los estimadores con propiedades similares; obtuvieron una eficiencia típica de alrededor de 7.3 con una estimación bisemanal de la varianza, y un valor de máximo de 8.5.

$$\sigma_{YZ} = \sqrt{\sigma_0^2 + k\sigma_{OC}^2 + (1-k)\sigma_{RS}^2} \quad (1.19)$$

Donde,

$$\begin{aligned}\sigma_o^2 &= \frac{1}{m-1} \sum_{i=1}^m \ln \left(\frac{o_i}{c_{i-1}} \right) - \overline{\ln \left(\frac{o_i}{c_{i-1}} \right)} \\ \sigma_{oC}^2 &= \frac{1}{m-1} \sum_{i=1}^m \ln \left(\frac{c_i}{o_i} \right) - \overline{\ln \left(\frac{c_i}{o_i} \right)} \\ k &= \frac{\alpha}{\alpha + \frac{m+1}{m-1}}, \quad \alpha = 1.34 \text{ suele ser el mejor valor.}\end{aligned}$$

Dado que sólo se pueden utilizar observaciones finitas para construir el rango, éste podría tener sesgos inesperados, especialmente para activos con poca liquidez y un volumen de transacción finito, lo que resultará en la apertura tardía así como en el cierre anticipado de las posiciones (Chou et al., 2009).

Asimismo, estadísticamente los modelos de rango son muy sensibles a los datos atípicos, sin embargo, muy pocos estudios se han realizado al respecto; como solución Chou (2005) sugiere un rango cuantil, es decir, usar la diferencia entre el promedio del mayor 5% y el promedio del menor 5% de observaciones, en lugar del rango tradicional. En lo que respecta al desempeño relativo, Shu y Zhang (2006) detectaron que, si bien estos estimadores tienen un buen desempeño si el precio sigue un proceso *GBM*, existen diferencias significativas entre éstos si la distribución de los rendimientos presenta saltos de apertura o una elevada tendencia; no obstante, los resultados empíricos respaldan su uso para calcular la volatilidad histórica.

1.2. Valor en Riesgo

El Valor en Riesgo (VaR, por las siglas en inglés) es una medida estadística que causó gran conmoción en los mercados desde su publicación en 1994, mediante el documento técnico *RiskMetrics* y cuya autoría le pertenece al

banco J. P. Morgan; su impacto comenzó al ser aprobado por el G-30 como parte de las “mejores prácticas” para el tratamiento de derivados en 1993. Fue un método novedoso para calcular el Riesgo de Mercado, dado que no sólo tomaba en cuenta la diversificación del portafolio y el nivel de apalancamiento, también el resultado era un solo número fácil de interpretar.

Así entonces, el VaR se define en *RiskMetrics* (JP Morgan & Co., 1996) como: “... una medida del máximo cambio potencial en el valor de un portafolio de instrumentos financieros, dada una probabilidad en un horizonte de tiempo establecido.” Esta medida responde a la pregunta “¿cuánto puedo perder con una probabilidad $x\%$ sobre un horizonte de tiempo?”. Es por dicha respuesta que una mejor definición sería (RiskMetrics Group, 1999): “El VaR se define como la pérdida prevista en el peor escenario con un nivel de confianza específico (e.g., 95%) en un horizonte de tiempo establecido (e.g., 1 día).”

Para cuantificar el VaR de un portafolio, se define W_0 como la inversión inicial y r como su tasa de rendimiento, entonces, el valor del portafolio al final del plazo sería $W = W_0(1 + r)$. Teniendo que para r , su rendimiento esperado y volatilidad sean μ y σ , respectivamente. El valor más bajo del portafolio con un nivel de significancia α sería

$$W^* = W_0(1 + r^*) \quad (1.20)$$

El VaR se define como la pérdida monetaria, relativa a la media

$$VaR = E(W) - W^* = -W_0(r^* - \mu) \quad (1.21)$$

Por lo que encontrar el VaR es el equivalente a buscar el valor mínimo para W^* , o el rendimiento crítico r^* . Expresado de una forma más general, el VaR puede derivarse de la distribución de probabilidad del valor futuro del

portafolio $f(x)$. Deseamos encontrar la peor realización posible W^* tal que la probabilidad de exceder dicho valor sea α , y definiendo el VaR en términos de la probabilidad de la cola izquierda de la distribución, tal que la probabilidad de un valor inferior a W^* , $p = P(w \leq W^*)$, sea $1 - \alpha$

$$1 - \alpha = \int_{-\infty}^{W^*} f(x)dx = p \quad (1.22)$$

Donde,

x es la ganancia o pérdida en cifras monetarias y;

$(1 - \alpha)$ se define como el nivel de confianza.

Como advertencias del modelo del VaR tenemos los siguientes puntos (Jorion, 2009):

- No describe la peor pérdida. Sólo podemos saber que la medida del VaR será superada con una frecuencia p ;
- No describe las pérdidas en la cola izquierda de la distribución, únicamente indica la probabilidad de que dichos valores ocurran;
- Es medido con algún error. Dicha cifra está sujeta a la variación de muestreo normal, es decir, un cambio en las muestras seleccionadas (período, longitud, metodologías estadísticas) arrojará diferentes valores de VaR.
- No otorga certidumbre sino una expectativa, la cual es válida sólo bajo condiciones normales del mercado, en momentos de crisis o turbulencia utilizan pruebas complementarias (pruebas de estrés o de valores extremos) (De Lara, 2013).

Hay tres tipos de mediciones relativas de VaR: (1) VaR relativo, (2) VaR marginal, y (3) VaR incremental. El VaR relativo mide el riesgo de obtener un rendimiento inferior relativo a un índice de referencia, o un punto de pérdida

máximo a alcanzar. El VaR marginal mide cuánto riesgo una posición agrega a un portafolio; específicamente, el cambio del VaR del portafolio si la posición fuera eliminada. Y el VaR incremental mide el impacto de pequeños cambios en la ponderación de la posición, por lo que está muy relacionado con el VaR marginal.

Para el cálculo del VaR están los métodos de valuación local (también conocidos como métodos analíticos) y de valuación completa. Los métodos de valuación local son modelos lineales (matriz de covarianzas y modelos de factor) que son empleados para instrumentos financieros tradicionales, tales como acciones, bonos, futuros, y swaps; y no-lineales (e.g., gamma, convexidad) empleados para instrumentos cuyo precio cambia de manera desproporcional en relación al activo subyacente, como es el caso de las opciones (Jorion, 2009). Los métodos de valuación completa son Simulación Histórica y Simulación Monte Carlo, estos modelos son apropiados para todos los instrumentos, lineales y no-lineales.

1.2.1. Metodologías

Método Delta-Normal

Este método lineal, también llamado “paramétrico” o de “varianza-covarianza”, es la aproximación de VaR más simple. Asume que las exposiciones del portafolio son lineales y que los factores de riesgo están distribuidos conjuntamente de manera normal (Jorion, 2007). El modelo tiene la ventaja de que no requiere gran cantidad de datos históricos, solo necesita la volatilidad y la matriz de correlación.

Para traducir la distribución general $f(x)$ en una distribución normal estándar $\phi(\epsilon) \sim N(0,1)$, partiendo de la ecuación (1.20) asociando W^* con r^* , debido a

que generalmente r^* es negativo, se puede escribir como $-|r^*|$, y hacemos una transformación estándar

$$-z = \frac{-|r^*| - \mu}{\sigma} \quad (1.23)$$

Que es equivalente a

$$1 - \alpha = \int_{-\infty}^{W^*} f(x) dx = \int_{-\infty}^{-z} \phi(\epsilon) d\epsilon \quad (1.24)$$

Entonces, si consideramos que el nivel z lo podemos conocer dado un nivel de confianza $(1 - \alpha)$ y a partir de las tablas de *función de distribución normal acumulativa* (ver tabla 2), r^* se define como

$$r^* = -z\sigma + \mu \quad (1.25)$$

Asumiendo que μ y σ están expresados de forma anual, el horizonte de tiempo considerado es Δt , en años. Sustituyendo en la ecuación (1.21) la ecuación (1.25) tenemos que el VaR individual es

$$VaR = W_0 z \sigma \sqrt{\Delta t} \quad (1.26)$$

Donde z toma los siguientes valores dependiendo del nivel de confianza deseado

Tabla 2. Distribución Normal Acumulativa

α	$1 - \alpha$	z_α	$z_{\alpha/2}$
10%	90%	1.282	1.645
5%	95%	1.645	1.960
2%	98%	2.054	2.326
1%	99%	2.326	2.576

Fuente: Elaboración propia.

Para el caso de un portafolio de activos, consideramos la teoría desarrollada por Markowitz (como fue citado por Reilly & Brown, 2011) donde establece

que el riesgo de un portafolio puede reconstruirse a partir de una combinación de los riesgos de los valores subyacentes. Siendo Σ la matriz de varianza-covarianzas del portafolio y x es el vector del valor de los activos, sustituyendo en la ecuación (1.26) el VaR del portafolio se puede obtener de la siguiente manera:

$$VaR_p = z[x^T \Sigma x]^{1/2} \sqrt{\Delta t} \quad (1.27)$$

Que sería igual a

$$VaR_p = [VaRI * C * VaRI^T]^{1/2} \quad (1.28)$$

Donde,

$VaRI$ es el vector de VaR individuales $1 \times n$;

C es la matriz de correlación $n \times n$;

$VaRI^T$ es el vector de VaR individuales transpuesto $n \times 1$.

A este VaR también se le conoce como *VaR Diversificado* puesto que toma en cuenta las correlaciones entre los rendimientos de los activos, por lo que el riesgo del portafolio es menor a la suma aritmética de los VaR individuales, suponiendo que la correlación entre los activos sea menor a 1; el estimado de *RiskMetrics* es similar a este método. El principal beneficio de este método es su simplicidad, sin embargo, debido a ello no toma en cuenta efectos no-lineales (instrumentos financieros como las opciones) y subestima la ocurrencia de grandes observaciones negativas por el supuesto de la distribución normal (Jorion, 2009).

Método de Simulación Histórica

Este método es uno de los que proporciona una implementación directa de valuación completa. Consiste en utilizar una serie histórica de datos como guía

de lo que pasará en el futuro. Se le aplican las ponderaciones actuales a una serie de tiempo de rendimientos históricos del activo.

El tiempo actual se define como t ; y observamos la información de 1 a t . El valor del portafolio actual es P_t , el cual es una función de los factores de riesgo actuales

$$P_t = P[f_{1,t}, f_{2,t}, \dots, f_{N,t}] \quad (1.29)$$

Se obtienen los cambios logarítmicos de los factores a partir de la distribución histórica

$$\Delta f_i^k = \{\Delta f_{i,1}, \Delta f_{i,2}, \dots, \Delta f_{i,t}\} \quad (1.30)$$

A partir de esta información, se determina una serie de tiempo simulada para construir un valor hipotético del portafolio actual a partir de un nuevo escenario

$$P^k = P_t(1 + \Delta f_i^k) \quad (1.31)$$

Entonces, se procede a obtener la serie de pérdidas y ganancias simuladas $P^k - P_t$. Por último, calculamos el VaR tomando el percentil que corresponda al nivel de confianza deseado.

La ventaja que tiene este método es que no hace supuestos específicos de la distribución de los rendimientos, más que confiar en la información histórica. Ésta es una mejora sobre la distribución normal dado que los datos históricos generalmente siguen una distribución de colas pesadas. A su vez, supone que para el horizonte seleccionado, el rendimiento próximo a ocurrir ya ha sido observado en la historia, lo que resulta en su principal desventaja; da por hecho que no habrá algún rendimiento mayor que aquellos observados en la ventana de tiempo considerada para el cálculo, por lo que el modelo es muy sensible a

la amplitud de la ventana usada, así como a la magnitud de los rendimientos observados.

Método de Simulación Monte Carlo

Este método es similar al método anterior de Simulación Histórica, excepto que los cambios hipotéticos de factores de riesgo Δf_i^k son generados por movimientos aleatorios a partir de un proceso estocástico predefinido, en lugar de la información histórica. Para implementar el modelo Monte Carlo primero es necesario seleccionar un proceso estocástico para el precio del factor de riesgo. Posteriormente, se generan simulaciones con variables aleatorias artificiales que representen el factor de riesgo en el horizonte objetivo.

Dado el supuesto de que los precios de los instrumentos reflejan un patrón de caminata aleatoria en mercados eficientes (basado en la forma débil de la Hipótesis de Mercados Eficientes) (Fama, 1965), entonces se asume que los precios siguen un *Proceso de Márkov*.^{††} Un caso particular de dicho proceso, sería el *Proceso de Wiener* (o *movimiento Browniano*), que describe una variable Δx , y cuyo cambio es medido en el intervalo Δt de tal manera que su cambio promedio sea cero y varianza proporcional a Δt , por lo que la variable x_t debe cumplir la siguiente ecuación (Jorion, 2007):

$$x_t = x_{t-1} + \Delta z, \quad \Delta z \sim N(0, \Delta t)$$

Si ξ es una variable normal estándar $N(0,1)$, entonces tenemos que

$$\begin{aligned} \Delta z &= \xi_t \sqrt{\Delta t} \\ \therefore x_t - x_{t-1} &= \Delta x = \xi_t \sqrt{\Delta t}, \quad \Delta x \sim N(0, \Delta t) \end{aligned}$$

^{††} Tipo particular de proceso estocástico en el que la variable aleatoria “carece de memoria”, es decir, la distribución de probabilidad del valor futuro de una variable aleatoria es únicamente dependiente de su valor actual, por lo que es independiente de su historia.

Siendo todo incremento Δx independiente a través del tiempo.

Una generalización del *Proceso de Wiener*, también conocida como *Movimiento Browniano Aritmético* (ABM, por sus siglas en inglés), sería definir la variable Δx en términos de un proceso de Wiener, agregando una tendencia μ por unidad de tiempo y volatilidad σ

$$x_t - x_{t-1} = \Delta x = \mu\Delta t + \sigma\Delta z \quad (1.32)$$

Donde,

μ y σ son constantes;

$\Delta z = \xi_t\sqrt{\Delta t}$ es un *Proceso de Wiener*.

Por lo que la variable aleatoria Δx se distribuye según una normal $N(\mu\Delta t, \sigma^2\Delta t)$. Al asignarle una dependencia a μ y σ respecto al valor actual de la variable, el resultado es un proceso *GBM*, por lo que éste está dado por

$$\Delta x = \mu x_{t-1}\Delta t + \sigma x_{t-1}\Delta z \quad (1.33)$$

Despejando para x_t tenemos

$$x_t = x_{t-1} + x_{t-1}(\mu\Delta t + \sigma\Delta z) \quad (1.34)$$

Por lo tanto, este proceso es una ecuación recursiva, cuyos valores de μ y σ permanecen constantes. El proceso es geométrico porque todos los parámetros son escalados por el valor actual de x , esto es comúnmente el caso para los precios de las acciones. Tomando en cuenta la ecuación (1.33), cabe resaltar que el cociente del cambio de la variable dividido entre el valor anterior sigue un proceso *ABM*

$$\frac{\Delta x}{x_{t-1}} = \mu\Delta t + \sigma\Delta z \quad (1.35)$$

$$\therefore \frac{\Delta x}{x} \sim N(\mu\Delta t, \sigma^2\Delta t) \quad (1.36)$$

Dado el problema que se presenta si el horizonte considerado es muy amplio (e.g., en años), este proceso definido en tiempo discreto, es decir, como en la ecuación (1.34), podría presentar errores dada la posibilidad de que la variable x tome valores negativos. Para evitar este problema en las simulaciones, el proceso es aproximado por pequeños pasos con una distribución normal con media y varianza definidas en la ecuación (1.36). O bien, también podría definirse el proceso en tiempo continuo, donde el límite de dicho modelo sería una distribución normal para $dx/x = d\ln(x)$, donde x tendría una distribución lognormal. Entonces, el logaritmo para el precio final sobre un intervalo $T - t = \tau$ sería

$$\ln(x_T) = \ln(x_t) + \left(\mu - \sigma^2 * \frac{1}{2}\right)\tau + \sigma\sqrt{\tau}\varepsilon \quad (1.37)$$

Donde ε es una variable normal estandarizada.

Para simular la caminata del precio futuro S_{t+n} , comenzamos a partir del precio al tiempo t con referencia en la ecuación (1.34), y generamos una secuencia de valores aleatorios ξ_i , siendo $i = 1, \dots, n$. Por lo que S_{t+1} se define como

$$S_{t+1} = S_t + S_t(\mu\Delta t + \sigma\xi_1\sqrt{\Delta t}) \quad (1.38)$$

Y repetimos la ecuación (1.38) hasta alcanzar el horizonte objetivo T , el cual sería cuando $S_{t+n} = S_T$. Posteriormente, se calcula el valor del activo en el horizonte objetivo, $F_T(S_T)$. Se replica la generación de escenarios las veces que sean necesarias, digamos k veces, teniendo como resultado una distribución de valores F_T^1, \dots, F_T^k . Obtenemos el vector de pérdidas y ganancias, ordenamos los datos, y tomamos aquel que corresponda al c -ésimo cuantil.

Una de las ventajas que presenta este método es que es lo suficientemente flexible para incorporar la variación en el tiempo en la volatilidad o en los rendimientos esperados, colas pesadas, y escenarios extremos, así como considerar exposición a precios no-lineales (Jorion, 2007). Anteriormente, una de las principales desventajas del modelo era la limitación computacional, ya que se requería de un equipo sofisticado, en cambio, en la actualidad, con casi cualquier computadora es posible realizar esta simulación. No obstante, el modelo si presenta un costo computacional elevado, si la valuación del activo en la fecha objetivo involucra una simulación, entonces el método requiere que se haga una simulación dentro de otra simulación, lo cual se volvería muy oneroso para implementarlo de forma frecuente. Otra desventaja es el riesgo asumido por la implementación del modelo, ya que se basa en procesos estocásticos específicos para los factores de riesgo subyacentes, los cuales podrían estar mal (Jorion, 2007).

Dadas las limitaciones que presentan los modelos anteriores, se recomienda realizar pruebas de estrés y backtesting como complemento a la medición del Valor en Riesgo.

Prueba de Estrés

La Prueba de Estrés, o de valores extremos, muestra lo que se podría perder en situaciones de crisis o elevada turbulencia en los mercados, es decir, se identifican y gestionan situaciones que podrían resultar en pérdidas extraordinarias. Dicho proceso se puede llevar a cabo mediante (GARP, 2016):

- 1) análisis de escenarios;
- 2) modelos de estrés, volatilidades, y correlaciones;
- 3) políticas adecuadas.

El análisis de escenarios consiste en evaluar el portafolio en probables situaciones extremas en las que podrían estar los mercados, esto normalmente se realiza moviendo variables financieras clave, de manera secuencial y por grandes montos; también se le conoce como “prueba de sensibilidad”. El segundo punto concierne a los parámetros de entrada de los modelos, como son la volatilidad y las correlaciones, los cuales no son constantes en el tiempo y tienden a desviarse significativamente de sus promedios históricos en momentos de crisis, por lo que se modifican sus valores creando nuevos escenarios que se analizan respecto a la base original. El punto de políticas adecuadas se refiere a las acciones que se tomarán si, una vez realizadas las pruebas de estrés pertinentes, se obtiene una pérdida estresada inaceptablemente elevada; este curso de acción debe asegurar que la institución pueda soportar el peor escenario obtenido por las pruebas de estrés.

Backtesting

En la prueba de Backtesting, se evalúa la efectividad del modelo de VaR y la conveniencia de calibrarlo. Ésta se realiza comparando el VaR observado contra las pérdidas y ganancias reales, obteniendo la eficiencia a partir de la razón del número de observaciones que fueron superiores al VaR y el total de observaciones, para un período de tiempo determinado. En cuanto a la calibración, de acuerdo con De Lara (2013) uno de los métodos más utilizados es el desarrollado por Kupiec, el cual consiste en contar las veces que las pérdidas y/o ganancias exceden el VaR en un período determinado. De las observaciones obtenidas, se asume que éstas siguen una distribución binomial y, siendo T el número total de observaciones y p la probabilidad de un exceso para un nivel α (usado en el cálculo del VaR) dado, la función de densidad que indica la probabilidad de observar N número de observaciones que excedan las pérdidas o ganancias sería:

$$Pr(N|T, p) = \binom{T}{N} (1-p)^{T-N} p^N \quad (1.39)$$

Donde si la probabilidad estimada está por encima del nivel de significancia “nulo” deseado (i.e., 5-10% comúnmente), el modelo no se rechaza; caso contrario, si está por debajo del nivel de significancia, el modelo se rechaza y se concluye que el modelo no es correcto. Puesto que, la decisión de relevancia a tomar es, si la relación de excesos de pérdidas y/o ganancias contra las observaciones totales es estadísticamente diferente a la probabilidad α , Kupiec desarrolló unas regiones de confianza con la hipótesis nula de que p es estadísticamente igual a la probabilidad α , y con una hipótesis alternativa de que p es diferente a ésta (tabla 3).

Tabla 3. Tabla de Kupiec.

Región de No-Rechazo para N número de errores				
$1 - \alpha$	$T = 125$ días	$T = 252$ días	$T = 510$ días	$T = 1000$ días
99%	$N < 4$	$N < 7$	$1 < N < 11$	$4 < N < 17$
97.5%	$N < 8$	$2 < N < 12$	$6 < N < 21$	$15 < N < 36$
95%	$2 < N < 12$	$6 < N < 20$	$16 < N < 36$	$37 < N < 65$
90%	$6 < N < 20$	$16 < N < 36$	$38 < N < 65$	$81 < N < 120$

Nota: N es el número de veces que se excedió el límite de VaR sobre T observaciones sin rechazar la hipótesis nula. Fuente: FRM© Parte II (GARP, 2016).

A su vez, el Comité de Basilea estableció sus criterios para la prueba de backtesting, clasificando los resultados en tres zonas de colores: verde, amarillo, y rojo. La zona verde significa que el modelo no tiene problemas de calidad, por lo tanto, no requiere de ajustes; la zona amarilla indica que no se puede tener una conclusión respecto al modelo, por ende, hay incertidumbre si debe o no calibrarse; y, la zona roja significa que el modelo presenta problemas de calidad y precisión, por lo que es necesario modificarlo. Asimismo, el

Comité de Basilea propone aplicar penalizaciones de incremento del capital requerido a la institución, el cual es tres veces el VaR (De Lara, 2013), en caso de caer en las zonas amarilla y roja (tabla 4). La penalización en la zona amarilla dependerá del motivo en el que se presenten los errores (GARP, 2016):

- Integridad básica del modelo. La desviación ocurrió porque las posiciones se reportaron de manera incorrecta o por un error en el código del modelo.
- La precisión del modelo se puede mejorar. La desviación ocurrió porque el modelo no calcula el riesgo con suficiente precisión.
- Operaciones intradía. Las posiciones cambiaron durante el día.
- Mala suerte. Los mercados estuvieron particularmente volátiles o las correlaciones cambiaron.

Tabla 4. Zonas de Penalización de Basilea

Zona	Número de excepciones	Incremento de Capital
Verde	0 a 4	0%
Amarilla	5	40%
	6	50%
	7	65%
	8	75%
	9	85%
Roja	10+	100%

Fuente: FRM© Parte II (GARP, 2016).

1.3. Inteligencia Computacional

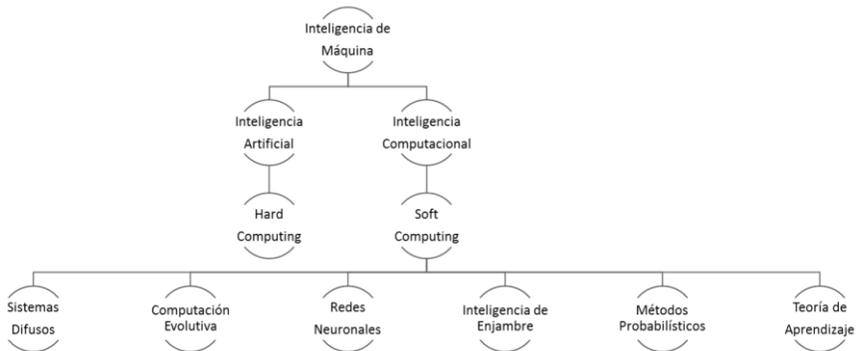
La Inteligencia Computacional es considerada por algunos expertos como el sucesor de la Inteligencia Artificial (IA) (Sumathi & Paneerselvam, 2010). Ésta hace referencia a la construcción de sistemas que tengan la capacidad de

adaptarse y de lidiar con procesos demasiado complejos para ser representados por el modelado matemático (tradicional).

El término de Inteligencia Computacional (IC) fue empleado por primera vez en 1990 por el IEEE Neural Networks Council, y definido en 1994 por Bezdek (como fue citado por Konar, 2005): “Un sistema es computacionalmente inteligente cuando: trata con datos numéricos (de bajo nivel), tiene componentes de reconocimiento de patrones, no usa conocimiento en el sentido de IA; y además, comienza a exhibir i) adaptabilidad computacional, ii) tolerancia a fallos computacionales, iii) velocidad semejante al cambio humano, y iv) tasas de error que se aproximan al desempeño humano.” Posteriormente, en 1998 Zadeh (como fue referenciado por Siddique & Adeli, 2013) propuso la distinción entre las técnicas de hard computing basadas en IA y las técnicas de soft computing basadas en IC. El hard computing consiste en la computación convencional y requiere de modelos analíticos perfectamente establecidos; mientras que el soft computing es tolerante a la imprecisión, incertidumbre, y verdad parcial.

La IC usa conocimiento basado en la experiencia sobre el proceso que produce un modelo en términos de comportamiento de entrada/salida. Engelbrecht (como fue referenciado por Siddique & Adeli, 2013) considera los siguientes enfoques básicos de la IC: i) Lógica difusa, ii) Computación evolutiva, iii) Redes neuronales, iv) Inteligencia de enjambre, v) Métodos probabilísticos, y vi) Teoría de aprendizaje (figura 1).

Figura 1. Principales enfoques de la Inteligencia Computacional



Fuente: Elaboración propia con base en Siddique & Adeli (2013).

Lógica Difusa

La lógica difusa aproxima el razonamiento humano usando términos lingüísticos imprecisos, o “difusos”; ésta permite la toma de decisiones con valores estimados a partir de información incompleta o incierta. Este concepto fue propuesto en 1965 por Lofti A. Zadeh para modelar el razonamiento humano a partir de información imprecisa e incompleta mediante definiciones y términos vagos permitiendo la construcción de una base de reglas (Siddique & Adeli, 2013). El objetivo principal de la lógica difusa es proporcionar un soporte formal al razonamiento basado en el lenguaje natural, el cual se caracteriza por usar proposiciones de carácter impreciso (Martín & Sanz, 2007).

Computación Evolutiva

La computación evolutiva se basa en gran medida en los principios de la teoría de evolución de Darwin. Consiste en la emulación del proceso de selección natural en un procedimiento de búsqueda (Siddique & Adeli, 2013). La búsqueda involucra realizar recombinaciones de los individuos para formar descendientes, mutaciones, y así obtener la selección de la siguiente generación

hasta que una solución emerge de la población. La computación evolutiva usa un progreso y desarrollo iterativos en una población, la cual, es posteriormente seleccionada en una búsqueda aleatoria guiada usando procesamiento en paralelo para lograr la población de soluciones deseada (Siddique & Adeli, 2013). Algunas metodologías cuya base es la computación evolutiva son: programación evolutiva, programación genética, estrategias evolutivas, y algoritmos genéticos.

Redes Neuronales

Los modelos de redes neuronales están inspirados por los procesos y estructura del cerebro en el nivel más bajo: las neuronas. Una red neuronal artificial es un análogo electrónico de la red neuronal biológica. El cuerpo de la red neuronal es modelado por una función de activación, la cual intenta mejorar la señal recibida por diferentes dendritas (Konar, 2005). Las redes neuronales artificiales tienen su origen en 1940, gracias al trabajo de Hebb (como fue referenciado por Siddique & Adeli, 2013). Éstas son sistemas biológicamente inspirados, masivamente paralelos y de procesamiento de información distribuida. Se caracterizan por su poder computacional, tolerancia a fallas, aprendizaje de datos experimentales, y capacidad de generalización (Siddique & Adeli, 2013).

Inteligencia de Enjambre

Es una técnica computacional basada en el estudio del comportamiento colectivo en sistemas descentralizados y auto-organizados, típicamente constituidos por una población de agentes simples que interactúan localmente entre sí y con su entorno (Siddique & Adeli, 2013). Algunos algoritmos diseñados a partir de esta técnica son: Optimización de enjambre de partículas, Optimización de colonias hormiga, y Algoritmo luciérnaga,

Métodos Probabilísticos

La teoría de probabilidad ha sido vista como una metodología de decisión para tratar con incertidumbre e imprecisión. El método probabilístico involucra considerar un espacio de probabilidad adecuado sobre una familia más amplia de estructuras, y demostrando que un mismo punto correspondiente a la estructura requerida tiene probabilidad positiva en este espacio. Este método fue introducido por Erdos y Spencer en 1974 (Siddique & Adeli, 2013). El comportamiento probabilístico o de estocasticidad (aleatoriedad) también se considera a veces como un atributo de los sistemas inteligentes (Siddique & Adeli, 2013).

Teoría de Aprendizaje

Las teorías de aprendizaje proveen explicaciones de los procesos de la entrada de información, su procesamiento y almacenamiento, así como la manera en que éstos ocurren (Siddique & Adeli, 2013). En psicología, Ormrod (como fue referenciado por Siddique & Adeli, 2013) define el aprendizaje como el proceso de reunir los efectos cognitivo, emocional y ambiental y las experiencias para adquirir, mejorar o cambiar conocimiento, habilidades, valores y percepciones.

A partir de las teorías del aprendizaje psicológico, surge la disciplina del Aprendizaje Automático (*Machine Learning*). Ésta puede ser definida como la disciplina que hace que la computadora cambie su estructura, programa o datos (con base en sus valores de entrada o en respuesta a información externa) de tal forma que mejore su desempeño futuro esperado (Nilsson, 1998). Una definición más formal y que plantee de mejor manera el problema de aprendizaje sería la de Tom Mitchell (1997): “Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna clase de tareas

T y medida de desempeño P, si su desempeño en las tareas en T, medido por P, mejora con la experiencia E.”

En esta disciplina hay tres formas básicas de aprendizaje: supervisado, no supervisado, y de reforzamiento. El aprendizaje supervisado, le enseña a la computadora cómo hacer algo; dado un conjunto de datos y con el conocimiento de cuál debería ser el valor de salida correcto (suponiendo que existe una relación entre los valores de entrada y de salida), se genera una función que acopla los valores de entrada con los valores de salida deseados. El aprendizaje no supervisado, le permite a la computadora aprender cómo hacer algo; permite aproximar problemas con poca o nula idea del resultado correcto (no hay retroalimentación basada en la predicción de resultados), por lo que la función modela una representación interna de los valores de entrada y los acopla con patrones similares. Por último, en el aprendizaje de reforzamiento, la computadora analiza y optimiza el comportamiento de un agente basado en la retroalimentación del ambiente; el valor de salida es una acción o secuencia de acciones donde una señal de supervisión es una recompensa ocasional.

1.3.1. Redes Neuronales

Las redes neuronales fueron inspiradas originalmente por el objetivo de tener máquinas que pudieran replicar el cerebro. Consisten en elementos computacionales simples que se conectan de varias maneras entre sí para servir una cierta función. Su arquitectura fue inspirada por la noción de la función cerebral que se tenía de mediados a finales del siglo XX, debido a esto, se les denomina “neuronales” (Wallisch et al., 2014).

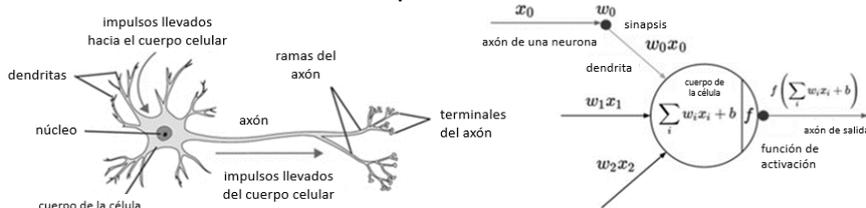
Redes Neuronales Biológicas

Biológicamente, existe una gran variedad de tipos de neuronas (e.g., sensoriales, motoras, interneuronas, espejo, etc.) en el cerebro humano, y un sinnfín de variaciones de éstas en otras especies; la teoría cerebral abstrae la complejidad de éstas con el fin de entender distintos aspectos de su desarrollo, aprendizaje, o función. Entonces, las neuronas artificiales son diseñadas como variaciones en las abstracciones de la teoría cerebral (Arbib, 2003)

El cerebro humano tiene aproximadamente 100 mil millones de neuronas, donde cada una de éstas se encuentran conectadas entre sí. A las conexiones entre las mismas se les llama “sinapsis”, y el cerebro humano contiene alrededor de 60 billones de dichas conexiones (Russell & Norvig, 2010). De manera general y simplificada, las neuronas reciben información de otras neuronas a través de las dendritas, los axones conectan con las dendritas en la sinapsis, y si la señal de entrada excede cierto límite, la neurona dispara un pulso eléctrico hacia la neurona post-sináptica.

En su equivalente computacional, las señales que viajan a lo largo del axón (e.g., x_0), interactúan de manera multiplicativa (e.g., w_0x_0) con las dendritas de otras neuronas basadas en la fuerza sináptica (e.g., w_0) (Karpathy, 2017) (figura 2). El cerebro humano tiene una propiedad denominada “plasticidad”, que significa que las neuronas pueden cambiar la naturaleza y número de sus conexiones con otras neuronas, en respuesta a los eventos que ocurren. Así, el cerebro es capaz de aprender.

Figura 2. Representación biológica de una “neurona básica” y su equivalente computacional



Fuente: Imagen traducida de notas de Karpathy (2017).

Redes Neuronales Artificiales

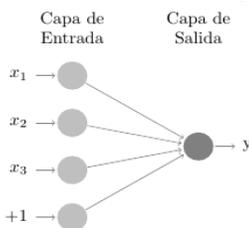
Las redes neuronales artificiales (ANN, por sus siglas en inglés) consisten en un número de neuronas artificiales, éstas tienden a tener menores conexiones que su análogo biológico y las redes son significativamente menores en contraste con el número de neuronas del cerebro humano (Russell & Norvig, 2010).

El primer modelo de neurona artificial surgió en 1943 cuando McCulloch y Pitts diseñaron modelos de redes neurológicas (como fueron referenciados por Kriesel, 2007). Posteriormente alrededor de 1950, Frank Rosenblatt (como fue referenciado por Siddique & Adeli, 2013) propuso el modelo de neurona conocido como “perceptrón” (figura 3). Este modelo calcula la suma ponderada de los valores de entrada, resta un valor límite y da como resultado uno de dos valores posibles. La función de activación se define como:

$$y = \begin{cases} 1, & \text{si } z \geq b \\ 0, & \text{caso contrario} \end{cases} \quad (1.40)$$

Siendo $z = \sum_{i=1}^n w_i x_i$; $i = 1, 2, \dots, n$; y b el umbral de disparo (valor de sesgo o “bias”, normalmente toma el valor de 1).

Figura 3. Modelo de neurona perceptrón



Fuente: Elaboración propia con base en Haykin (2009).

En 1960, surgen las primeras ANN ampliamente utilizadas comercialmente: las redes ADALINE (ADAPtive LINear Elements, o Elementos Lineales Adaptativos) y MADALINE (múltiple ADALINE); modelos diseñados por Bernard Widrow y Marcian Hoff, los cuales fueron entrenados por la regla de aprendizaje Widrow-Hoff, también conocida como “regla delta” (Kriesel, 2007), sin embargo, debido a los límites de la neurona perceptrón (límites expuestos por Minsky y Papert en su libro *Perceptrones* en 1969), se detuvo la popularidad de las redes y con ella, se pararon los fondos de investigación. No fue hasta la publicación “Aprendiendo representaciones por la retropropagación de errores” (*Learning representations by back-propagating errors*) de Rumelhart, Hinton y Williams en 1986, que se retomó la idea de las ANN (Kriesel, 2007).

El tema de esta última investigación es de importancia porque en ella desarrollaron el procedimiento de aprendizaje de “retropropagación del error” como una generalización de la “regla delta”; éste permitió resolver problemas que no eran linealmente separables. A su vez, se hace mención por primera ocasión de las “capas ocultas” que permitieron construir las Redes Perceptrón

Multicapa^{†††} (MLP, por sus siglas en inglés) (Rumelhart, Hinton, & Williams, 1986).

Tipos de ANN

Las ANN se pueden clasificar de acuerdo con su función de activación, paradigma de aprendizaje, o su arquitectura.

Función de Activación

Algunas de las funciones de activación más comunes son (Haykin, 2009):

- a) Lineal. El valor de salida en esta función es igual a su valor de entrada

$$f(z) = z \quad (1.41)$$

- b) Límite Binario. El valor de salida sólo puede tomar los valores de 1 y 0 dependiendo del valor de entrada. Con esta función las neuronas clasifican las entradas en dos categorías diferentes. Ejemplos de este tipo de función son la neurona de McCulloch y Pitts, y la neurona Perceptrón.

- c) Sigmoide. Toma un valor real y lo acopla en el rango [0,1]. Usualmente es la función logística

$$f(z) = \frac{1}{1+e^{-z}} \quad (1.42)$$

- d) Signum. Es similar a la función de límite binario, pero el rango de valores va [-1,1] en lugar de [0,1]. Una aproximación para su determinación, y en forma correspondiente a la función sigmoide, sería la función tangencial hiperbólica

^{†††} Erróneamente llamadas así debido a que la red no necesariamente sigue las mismas reglas o tiene la misma función de activación que la neurona Perceptrón (Hinton, 2012).

$$f(z) = \text{sgn}(z) \approx \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (1.43)$$

- e) ReLU (Rectified Linear Unit, o Unidad Lineal Rectificada) (Goodfellow, Bengio, & Courville, 2016). En las redes neuronales modernas es común que se recomiende el uso de esta función con redes de tipo feedforward, definida por la función de activación

$$f(z) = \max(0, x) \quad (1.44)$$

Es decir, la activación se limita en cero y el valor de salida es no lineal

$$y = \begin{cases} z, & \text{si } z > 0 \\ 0, & \text{caso contrario} \end{cases}$$

En todos los casos $z = \sum_{i=1} w_i x_i$

Paradigma de Aprendizaje

Hay 3 tipos de paradigmas de aprendizaje: Supervisado, No supervisado, y de Reforzamiento (Russell & Norvig, 2010).

Como ya se mencionó, el *Aprendizaje Supervisado* aprende a predecir un valor de salida objetivo, dado un vector de valores de entrada. El aprendizaje se puede dividir en dos clases: Regresión y Clasificación. Para el caso de la Regresión, el valor objetivo es un número real o un vector de números reales. Debido a que en el aprendizaje se busca reducir la diferencia entre el valor objetivo (o), en cada caso de entrenamiento y el valor obtenido (y); una medida de desempeño del modelo, comúnmente utilizada, es la del error medio cuadrático (MSE, por sus siglas en inglés) (Goodfellow, Bengio, & Courville, 2016)

$$L = \frac{1}{2} \sum_i (y_i - o_i)^2 \quad (1.45)$$

Otras medidas usuales son el RMSE y la distancia Euclidiana. En la Clasificación, el valor objetivo es una etiqueta de clase; usando la regresión logística como función de activación, una medida de desempeño podría ser la siguiente

$$L = -\frac{1}{m} [\sum_i o_i \ln(o_i) + (1 - t_i) \ln(1 - o_i)] \quad (1.46)$$

Algunos ejemplos de ANN de aprendizaje supervisado son las MLP, Redes de Base Radial, y algunas Redes Recurrentes.

El *Aprendizaje No Supervisado* identifica la mejor representación interna de los datos. Algunos ejemplos de aprendizaje de este tipo son el aprendizaje de Hebbian y los mapas de Kohonen.

Respecto al *Aprendizaje de Reforzamiento*, la máquina aprende a seleccionar cierta acción que maximice la recompensa; la información que provee el sistema no dice por qué o cómo realizó la acción correcta, sólo que lo hizo bien. Este tipo de aprendizaje es difícil en cuanto a su entrenamiento debido a que las recompensas son rezagadas. Un ejemplo son las Redes Hopfield.

Arquitectura

En lo concerniente a la clasificación de las redes conforme a su arquitectura, es decir, la manera en que se interrelacionan las neuronas dentro de la red, los principales tipos según Haykin (2009) son Redes Mono Capa, Redes Feedforward y Redes Recurrentes; adicionalmente, Hinton (2012) menciona como un cuarto tipo a las Redes conectadas Simétricamente.

Red Mono Capa. Se refieren a la arquitectura perceptrón.

Redes FeedForward (de Avance). Son el tipo más común de redes; la primera capa es de entrada y la última capa es de valores de salida, la capa(s) intermedia(s) se le denomina “oculta(s)” (figura 4). Si la red llega a tener más de una capa oculta se les llama “Redes Neuronales Profundas” (*Deep Neural Networks*, o DNN) (Goodfellow et al., 2016). En este tipo de red la conexión es únicamente delantera. El valor de salida Y de la red se puede escribir de forma vectorial de la siguiente manera

$$Y = f(W * x + B) \quad (1.47)$$

Donde,

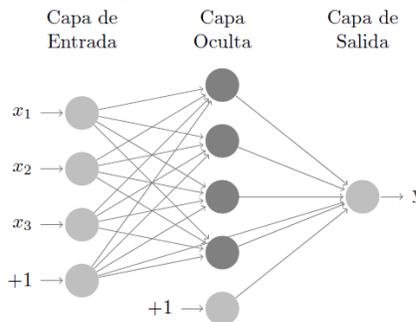
x es el vector de entrada;

W es la matriz de pesos sinápticos;

B es el vector de sesgo.

Algunos ejemplos de este tipo de redes son: MLP, Red de Base Radial, Red de Regresión Generalizada, y Red de Creencia (*Belief Network*).

Figura 4. Ejemplo de Red FeedForward



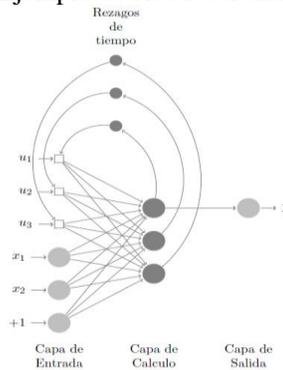
Fuente: Elaboración propia con base en Haykin (2009).

Redes Recurrentes (RNN, por sus siglas en inglés). También se les conoce como Redes de Realimentación (Feedback). Este tipo de redes tienen ciclos dirigidos en sus conexiones, esto implica que pueden tener dinámicas complicadas pudiendo dificultar su entrenamiento. Éstas se pueden obtener a partir de una red feedforward conectando sus valores de salida con los de entrada (figura 5). Son una representación más realista de la biología neuronal. Son equivalentes a redes muy profundas con una capa oculta por cada corte. Tienen la habilidad de recordar información en estado oculto por mucho tiempo. Este tipo de redes se consideran ideales para modelar series de tiempo y que usa los valores de salida de las unidades de la red en el tiempo t , como valores de entrada de otras unidades en el tiempo $t + 1$ (Mitchell, 1998). La mayoría de este tipo de redes también pueden procesar secuencias de longitud variable (Goodfellow et al., 2016). Tomando en cuenta la notación descrita en la red feedforward, podemos definir la red como (Siddique & Adeli, 2013)

$$Y(t + \Delta) = \Gamma(W * Y(t) + B) \quad (1.48)$$

Teniendo el tiempo de retardo Δ con un significado simbólico - es una analogía con el período refractario de un modelo elemental de neuronas biológicas-. Algunos ejemplos de RNN son las Redes de Elman, Redes Recursivas, Redes de Jordan, y Redes de Estado de Eco; como modelos más recientes están los modelos LSTM (*Long Short Term Memory* ó Larga Memoria de Corto Plazo), y GRU (*Gated Recurrent Unit* ó Unidad Recurrente Cerrada).

Figura 5. Ejemplo de Red Neuronal Recurrente



Fuente: Elaboración propia con base en Haykin (2009).

Redes conectadas Simétricamente. Son como las RNN, pero las conexiones entre las unidades son simétricas (tienen el mismo peso en ambas direcciones). A este tipo de redes sin unidades ocultas se les denomina “Redes Hopfield”, cuando tienen unidades ocultas reciben el nombre de “Máquinas de Boltzmann”.

1.4. Redes Neuronales aplicadas al pronóstico de volatilidad y al análisis del Riesgo de Mercado

Para corroborar la eficiencia del uso de modelos de redes neuronales para mejorar el nivel de pronóstico de datos financieros, los investigadores Locarek-Junge y Prinzler (1998) implementaron un modelo de Redes de Densidad Mezclada (MDN, por sus siglas en inglés)[‡] para estimar la distribución la serie de rendimientos del tipo de cambio Marco Alemán/Dólar EE.UU. y poder predecir el VaR de un portafolio; obtuvieron un mejor desempeño que el modelo base de *RiskMetrics*, sin embargo, el modelo no fue evaluado con

[‡] Modelo de redes neuronales propuestas por Bishop en 1996, como una extensión al modelo MLP para calcular las probabilidades condicionales.

métodos lo suficientemente adecuados, aunado al hecho de que el diseño de la red MDN no es consistente con los atributos de las series de tiempo (e.g., el efecto de retroalimentación de volatilidad) lo que puede llevar a tener sesgos en el modelo de redes. Por su parte Miazhynskaia, Dorffner, y Dockner (2003), usaron un modelo MDN recurrente lineal y otro no-lineal para estimar y pronosticar las distribuciones condicionales, y, a partir de éstas, calcular el VaR para las series de tiempo usadas (el índice Dow Jones Industrial average ó DJIA, el índice FTSE 100 de la bolsa de valores de Londres, y el índice de Japón NIKKEI 225); el modelo se comparó contra un modelo GARCH usando distribuciones normales y Student- t^{ff} para cada modelo, y el análisis empírico se basó en las series de rendimientos de los índices. No obtuvieron resultados significativos en la mejora del pronóstico respecto de los modelos lineales y no-lineales, pero si en el comparativo de los modelos normales respecto de los no-normales, dominando los primeros significativamente en las pruebas de evaluación.

Tino, Schittenkopf y Dorffner (2001) se enfocaron en el pronóstico de volatilidad implícita. Para ello, optaron por usar una red recurrente Elman utilizando datos del DAX 30 (índice bursátil de la bolsa de Frankfurt) y de contratos de opciones sobre éste, así como de la información intradía de los contratos de opciones del FTSE 100, comparando dicho modelo contra un modelo GARCH(1,1) y un modelo de Márkov (MM). Los resultados mostraron que la red tuvo un mejor desempeño contra el modelo GARCH, pero no tuvo mejoras significativas respecto al modelo MM, lo que sugiere que un modelo más simple y fácil de ajustar, como fue el MM, es aquél que producirá los mejores pronósticos. Por otro lado, Hamid e Iqbal (2004) compararon la

^{ff} Distribución de probabilidad simétrica definida por el único parámetro denominado “grados de libertad”.

predicción generada por una red feedforward con el de la volatilidad implícita de las opciones sobre futuros del índice S&P 500, obtenida mediante el modelo de precios de opciones americanas sobre futuros Barone-Adesi y Whaley (BAW); el pronóstico generado por el modelo de redes neuronales superó al de volatilidad implícita y no fue significativamente diferente a la volatilidad realizada^{eee} en los horizontes probados de 15 y 35 días, mientras que en el plazo de 55 días ambos modelos tuvieron resultados positivos.

Chen, Lai y Yen (2009) usaron una red feedforward para pronosticar la volatilidad condicional, la cual determinaron con la estimación del segundo momento condicional y sustrayendo el cuadrado del rendimiento estimado por la red neuronal; ya estimada la volatilidad condicional, calcularon el VaR mediante la simulación de la distribución del rendimiento usando el método de Simulación Monte Carlo. El modelo se comparó con el modelo ARMA-GARCH utilizando base de datos los rendimientos logarítmicos del índice Hang Seng, teniendo resultados superiores el pronóstico con redes neuronales. Una limitante de esta investigación es que no detallan el entrenamiento y selección del modelo de redes.

Chaudhuri y Ghosh (2015) realizaron su investigación enfocada al pronóstico de volatilidad histórica en el mercado accionario hindú, mediante un comparativo entre 2 arquitecturas de redes feedforward y 9 algoritmos de aprendizaje diferentes. Los resultados que obtuvieron fueron favorables al tener un nivel de pronóstico significativo para el año 2015, pero fueron negativos al

^{eee} Modelo de volatilidad donde ésta se obtiene al considerar la raíz cuadrada de la suma del cuadrado de los rendimientos para un período determinado (e.g., diario para obtener la volatilidad del mes, o intradía para obtener la volatilidad de un día).

querer utilizar un período reciente para pronosticar un período previo al mismo (año 2015 contra año 2008).

Petneházi y Gáll (2018) investigaron el nivel de estimación entre distintos tipos de modelos de volatilidad de rango contra el modelo de Desviación Estándar utilizando un modelo de redes neuronales de Larga Memoria de Corto Plazo (LSTM, por sus siglas en inglés) para pronosticar los rendimientos diarios del DJIA considerando un período de mayo 2008 a mayo 2016; los resultados que obtuvieron fueron que los modelos de rango tuvieron un mejor desempeño respecto del modelo que sólo utiliza los precios de cierre con un nivel de exactitud cercano al 60%, y que el uso de poca información con este tipo de red neuronal es viable, sin embargo, entre los mismos modelos de rango no hubo gran diferencia en el nivel de pronóstico.

En el caso del mercado mexicano, Silva (2017) utilizó un modelo MLP para el pronóstico de la volatilidad implícita de los contratos de opciones de América Móvil, el cual, al comparar su desempeño respecto a una regresión por Mínimos Cuadrados Ordinarios, demostró una mejora significativa en el pronóstico, teniendo el segundo método de estimación un mayor error respecto a la volatilidad implícita real.

Capítulo 2. Metodologías para el análisis del Riesgo de Mercado de una acción

Que lo mejor de tu pasado sea lo peor de tu futuro.

Anónimo

2.1. VaR usando modelo propuesto de pronóstico

El cálculo del VaR se realizó usando la ecuación (1.26), siendo el parámetro σ el valor obtenido mediante el pronóstico generado por el modelo de redes neuronales.

2.1.1. Predicción de volatilidad usando Redes Neuronales

De acuerdo con Mitchell (1997), las RNN son las redes más efectivas para emplear en el pronóstico de series de tiempo dada su arquitectura, puesto que éstas pueden hacer uso de una memoria interna para su procesamiento en una serie de datos que siguen una dependencia secuencial, es decir, además de utilizar como datos de entrada los datos suministrados de manera inicial, también toman en cuenta lo que han experimentado en el tiempo (la decisión que una RNN toma en un punto específico afectará las decisiones que tome posteriormente). Así, las RNN poseen dos fuentes de datos de entrada: los datos suministrados en el presente, y los recabados por el pasado inmediato. Las RNN son en esencia, redes de avance con circuitos que permiten que la información persista en la red (figura 6); dicho circuito permite que la información pase del estado t al estado $t + 1$. El mecanismo de avance en las RNN es similar al de las feedforward, teniendo que para cada paso de tiempo desde $t = 1$ hasta $t = T$, se aplican las ecuaciones:

$$h^{(t)} = H(b + Wh^{(t-1)} + Ux^{(t)}) \quad (2.1)$$

$$y^{(t)} = Y(d + Vh^{(t)}) \quad (2.2)$$

Donde,

$h^{(t)}$ es el estado oculto o la “memoria” de la red;

$y^{(t)}$ es el vector de valores de salida;

$x^{(t)}$ es la secuencia de valores de entrada;

b y d son los vectores de sesgo;

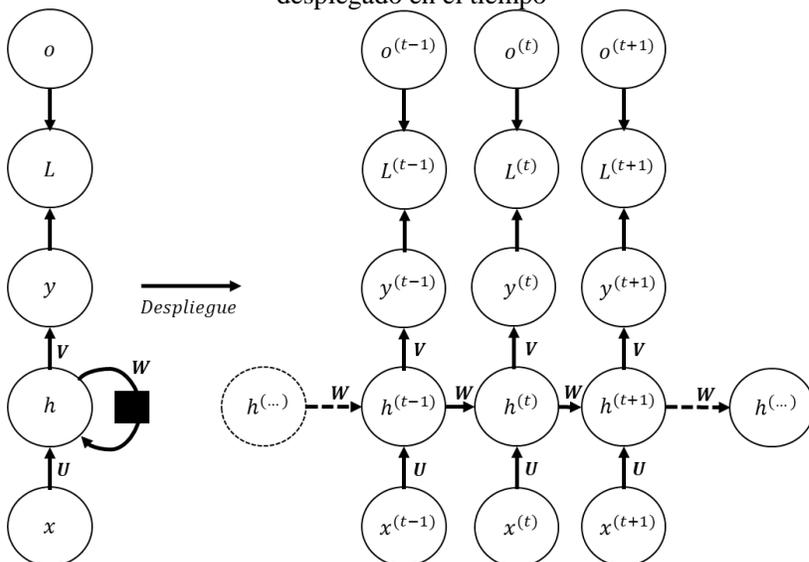
U , V y W son las matrices de pesos para los valores de entrada a ocultos, ocultos a salida, y ocultos a ocultos, respectivamente.

Para el caso del estado oculto, la función $H(\cdot)$ usualmente es una no lineal como $\tanh(\cdot)$ (ecuación 1.43) ó $ReLU(\cdot)$ (ecuación 1.44); para la función de salida Y , se empleó una función lineal (ecuación 1.41) dado el problema que la presente tesis busca resolver. Otra diferencia que tienen respecto de las redes feedforward es el método de aprendizaje, éste se da mediante la corrección del término de error aplicando el procedimiento de retropropagación (*backpropagation*) a través del tiempo (BPTT, por sus siglas en inglés), es decir, el gradiente¹ de cada valor de salida depende no sólo de los cálculos del paso de tiempo actual, sino también de los pasos de tiempo previos.

Sin embargo, dado que el objetivo buscado es poder capturar la memoria a corto y largo plazo, los modelos estándar de RNN pierden efectividad puesto que enfrentan un gran problema respecto a las dependencias a largo plazo: el gradiente se desvanece (o en raras ocasiones explota) dentro de una cadena de RNN en el tiempo (Bengio, Simard, & Frasconi, 1994). El gradiente se desvanece cuando los pesos sinápticos tienen pequeños valores que, en la multiplicación, van desapareciendo con el tiempo; mientras que explota cuando los pesos tienen grandes valores que siguen aumentando con el tiempo y conducen a la divergencia en el proceso de aprendizaje.

¹ Es un operador diferencial que se aplica a una función multivariable y da como resultado un vector cuyos componentes son las derivadas parciales de la función con respecto a sus variables. Éste apunta en la dirección del mayor incremento de la función; toma el valor de cero en un máximo o mínimo local. Normalmente se denota con el símbolo ∇ .

Figura 6. RNN estándar y su cómputo hacia adelante desplegado en el tiempo



Nota: El “despliegue” se refiere al desglose del diagrama del circuito (izq.) en la secuencia completa de la red (der.). Se observa una RNN estándar con una secuencia de valores de entrada x que corresponden a una secuencia de valores de salida y ; una medida de error L de los valores y con sus respectivos valores objetivo o ; y, la parametrización de las conexiones de los valores de entrada x con los valores ocultos h por una matriz de pesos U , de las conexiones recurrentes ocultas a ocultas por una matriz de pesos W , y de las conexiones de los valores ocultos h con los de salida y por una matriz de pesos V . Fuente: Elaboración propia con base en Goodfellow et al. (2016).

De entre los distintos tipos de RNN que existen, las que destacan son las redes LSTM, ya que estos modelos han lidiado mejor con el problema del desvanecimiento del gradiente (Goodfellow et al., 2016). Las redes LSTM fueron inicialmente propuestas por Hochreiter y Schmidhuber (1997), cuyo diseño fue explícitamente dirigido para lidiar con el problema de las dependencias a largo plazo. Su diseño no es muy diferente del de las RNN estándar, salvo que la determinación de su estado oculto tiene una diferente arquitectura: en el caso de las RNN estándar éste se compone de una capa oculta, es decir, sólo necesita de una función de activación para su cálculo; por

su parte, las LSTM constan de 4 capas ocultas -se les suele comparar con redes neuronales “muy profundas” dada la cantidad de capas ocultas que contiene cada estado- (figura 7). A su vez, a cada estado oculto se le denomina *celda de memoria*, la cual, de manera interna decide qué mantener (y qué eliminar) de la memoria. Ésta toma como valores de entrada la celda h_{t-1} previa y el valor actual x_t , y al final, combina el estado previo, la memoria actual, y los valores de entrada. Así entonces, las ecuaciones mediante las que se actualiza cada celda de memoria son las siguientes:

Puerta de Entrada

$$a_i^t = b^i + W^i h^{(t-1)} + U^i x^{(t)} \quad (2.3)$$

$$i^{(t)} = \sigma(a_i^t) \quad (2.4)$$

Puerta de Olvido

$$a_f^t = b^f + W^f h^{(t-1)} + U^f x^{(t)} \quad (2.5)$$

$$f^{(t)} = \sigma(a_f^t) \quad (2.6)$$

Puerta de Salida

$$a_q^t = b^q + W^q h^{(t-1)} + U^q x^{(t)} \quad (2.7)$$

$$q^{(t)} = \sigma(a_q^t) \quad (2.8)$$

Puerta de Externa

$$a_g^t = W^g h^{(t-1)} + U^g x^{(t)} \quad (2.9)$$

$$g^{(t)} = \tanh(a_g^t) \quad (2.10)$$

Estado Interno

$$s^{(t)} = s^{(t-1)} * f^{(t)} + g^{(t)} * i^{(t)} \quad (2.11)$$

Celda

$$h^{(t)} = H(s^{(t)}) * q^{(t)} \quad (2.12)$$

Donde,

$h^{(t)}$ es la celda de memoria;

$s^{(t)}$ es el estado interno de la celda;

$g^{(t)}$ es la puerta de entrada externa;

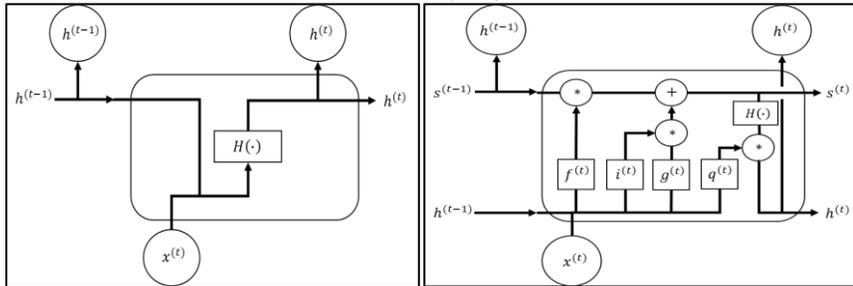
$i^{(t)}$, $f^{(t)}$ y $q^{(t)}$ son las *puertas*¹¹ de entrada, olvido y salida, respectivamente;

b , U y W son, respectivamente, los sesgos, los pesos de entrada, y los pesos recurrentes dentro de la celda de memoria;

Cada puerta controla el proceso de lectura, escritura y reinicio de la información. La puerta de entrada define los valores de actualización y cuánta información del estado actual se permite pasar. La puerta de olvido decide cuánta información del estado anterior se deja pasar. La puerta de salida define cuánta información del estado interno actual se expone a la red externa (capas superiores y el siguiente paso de tiempo), esto debido a que no toda la información en el estado interno puede ser relevante para la celda de memoria utilizada por otras unidades. Y la puerta de entrada externa (también denominada como el valor esperado del estado interno $s^{(t)}$) selecciona, del estado actual, los valores que podrían ser nuevos candidatos para el estado interno actualizado, éstos se adicionan al estado interno anterior junto con las modificaciones de la puerta de entrada. La función $\sigma(\cdot)$ denota la función de activación sigmoide (ecuación 1.42), y en el caso de la función $H(\cdot)$, como ya se mencionó, ésta debe ser una función no lineal, por lo que en el presente trabajo se optó por utilizar la función $\tanh(\cdot)$ para su cálculo ya que delimita el rango de valores en $[-1, 1]$.

¹¹ Se les denomina “puertas” porque la función sigmoide delimita los valores de estos vectores entre 0 y 1, y al multiplicar cada elemento con los elementos de otro vector se define cuánta información de este otro “pasa”.

Figura 7. Diagrama del estado oculto de una RNN estándar (izq.) y una red LSTM (der.)



Nota: El estado oculto de una RNN estándar (izq.) procesa la información a través de una sola capa; mientras que en la red LSTM (der.), éste consta de cuatro capas interactuando entre sí. El procesamiento de la información en la celda de memoria sigue el siguiente flujo: 1) mediante la puerta de olvido $f^{(t)}$, la celda decide qué información de la celda anterior $h^{(t-1)}$ y los datos de entrada actuales $x^{(t)}$ necesitan ser eliminados y se actualizan en el estado interno anterior $s^{(t-1)}$; 2) la puerta de entrada $i^{(t)}$ decide qué valores se actualizarán y la puerta de entrada externa $g^{(t)}$ selecciona los nuevos valores candidatos que podrían agregarse al estado interno, posteriormente éstos se combinan para actualizar el estado $s^{(t-1)}$ y así obtener el estado interno actual $s^{(t)}$; y, 3) la puerta de salida $q^{(t)}$ filtra los valores del estado $s^{(t)}$ que se van a obtener como resultado final y que pasarán a la siguiente celda. Fuente: Elaboración propia con base en Olah (2015).

A partir del surgimiento del modelo de redes LSTM se han ido desarrollando variantes del mismo que buscan optimizar y/o mejorar la predicción que éste genera, algunos ejemplos son las redes GRU, la máquina neuronal de Turing, Sin embargo, conforme al estudio realizado por Greff et al. (2017), encontraron que dichas variantes, en general, no presentan mejoras significativas en contraste con el modelo estándar de redes LSTM. Asimismo, Jozefowicz, Zaremba y Sutskever (2015) realizaron un estudio similar, pero probando alrededor de diez mil arquitecturas distintas de RNN, obteniendo como resultados que un modelo arrojó mejores resultados que las LSTM en ciertas tareas, pero no de manera relevante. Del mismo modo, Kim y Won (2018) propusieron un modelo de pronóstico combinando el modelo de redes LSTM junto con varios tipos de GARCH para predecir los valores del índice KOSPI 2000, comparándolo contra otros modelos (GARCH, GARCH exponencial,

EWMA, DNN, y un híbrido entre un modelo DNN y un GARCH), teniendo como resultados que el modelo con el mejor desempeño fue el híbrido entre el modelo LSTM y tres diferentes tipos de GARCH, ya que fue el que tuvo el menor error de pronóstico en todas las pruebas. Por lo tanto, se concluyó que utilizar un modelo de redes LSTM estándar tendría un buen desempeño para el presente análisis.

Ya seleccionado el modelo, se procedió a delimitar los principales aspectos de la red: i) selección y preprocesamiento de datos (variable dependiente y variables independientes); ii) conjuntos de entrenamiento, validación, y prueba; iii) arquitectura de la red; iv) modelo de entrenamiento; y, v) evaluación de resultados.

Selección de Datos

Variable dependiente

A partir de los precios diarios O, H, L, y C, y con el fin de tener un marco comparativo del modelo clásico respecto de los modelos de rango, se consideró usar las siguientes volatilidades como variables a pronosticar:

- Desviación Estándar (ecuación 1.2);
- Parkinson (ecuación 1.15);
- Garman-Klass (ecuación 1.16);
- Rogers-Satchell (ecuación 1.17);
- Extensión de Garman-Klass Yang-Zhang (ecuación 1.18); y,
- Yang-Zhang (ecuación 1.19).

Éstas se obtuvieron para cada mes y se anualizaron por el factor $\sqrt{252}$, considerando que el año de operación consta de 252 días. Para medir la

eficiencia de cada estimador, se consideró utilizar la eficiencia relativa propuesta por Slepaczuk y Zakrzewski (2009), ya que la eficiencia propuesta por Garman-Klass (ecuación 1.14) hace la comparación en la base de diferentes intervalos y horizontes de tiempo, lo que puede resultar en un comparativo erróneo entre los estimadores de volatilidad

$$rEff = \frac{std(\sigma_{SD}^{Ann})}{media(\sigma_{SD}^{Ann})} / \frac{std(\hat{\sigma}_i^{Ann})}{media(\hat{\sigma}_i^{Ann})} \quad (2.13)$$

Donde,

$media(\sigma_{SD}^{Ann})$ es la media del estimador de volatilidad clásico anualizado;

$std(\sigma_{SD}^{Ann})$ es la desviación estándar del estimador clásico anualizado;

$media(\hat{\sigma}_i^{Ann})$ es la media de un estimador de volatilidad anualizado dado;

$std(\hat{\sigma}_i^{Ann})$ es la desviación estándar de un estimador de volatilidad anualizado dado.

Variables independientes

En cuanto a los datos que sirven como insumo para la red neuronal, Hellström y Holmström (1998), mencionan que, para el pronóstico del mercado accionario, éstos se dividen en técnicos y fundamentales. Los datos técnicos son aquellos que se utilizan en el estudio de la acción histórica del precio, mediante el uso de gráficas e indicadores, con el fin de predecir futuras tendencias en éste (conocido como Análisis Técnico). Y los fundamentales, conciernen a los datos provenientes de la actividad de la compañía así como de la situación del mercado.

De los datos técnicos, las series diarias que más se utilizan son las de los precios O, H, L, y C, además del volumen operado. A partir de éstos, se pueden obtener

datos derivados como son los rendimientos logarítmicos (ecuación 2.14) y los rangos logarítmicos (ecuación 2.15)

$$r_i^d = \ln\left(\frac{C_t}{C_{t-1}}\right) \quad (2.14) \quad R_i^d = \ln\left(\frac{H_t}{L_t}\right) \quad (2.15)$$

Respecto a los datos fundamentales, además de considerar los factores de la economía en general, toman la idea de que una acción es una parte alícuota del valor de una empresa, por lo que sabiendo el valor de la misma es posible tomar decisiones de inversión. Éstos se pueden subdividir en variables económicas, de la industria, y de la compañía. Como variables económicas se encuentran: la inflación, tasas de interés, balanza de pagos, el Producto Interno Bruto, entre otros; como variables de la industria están: el valor de un índice de referencia (e.g., Índice de Precios y Cotizaciones ó IPC), materias primas (e.g., petróleo y metales), y el valor accionario de los competidores; mientras que, como variables de la compañía tenemos: múltiplos (e.g., razón precio/utilidad, razón precio/valor en libros, razón, precio/ventas, etc.), razones financieras (e.g., de solvencia, liquidez, apalancamiento financiero, rendimiento del capital, etc.), la estructura de capital, entre otros.

El objetivo de usar variables adicionales a la propia volatilidad es debido a investigaciones como la de Chen, Roll y Ross (1986), que concluyeron que los rendimientos de las acciones están expuestos a noticias económicas sistemáticas, y las variables que analizaron (producción industrial, cambios en la prima de riesgo, cambios en la curva de tasas, y cambios en la inflación esperada y no esperada) resultaron significativas para explicar el rendimiento de las acciones; aunque ellos no encontraron relación entre los rendimientos de las acciones y el cambio de precios de un índice de petróleo, García (2016) así lo hace para el caso del mercado mexicano, donde las variables de la Mezcla

Mexicana de Exportación (MME) y el tipo de cambio peso-dólar (USDMXN) resultan ser variables con un gran efecto en el riesgo sistemático, y si bien, él no obtuvo un resultado positivo para el caso de la TIIIE 28 días, confirma que ésta tiene una fuerte relación con el movimiento del tipo de cambio USDMXN. A su vez, Boz, Menéndez, Orgaz, y Prior (2015), demostraron que el movimiento de la volatilidad tiene relación con variables independientes de tipo fundamentales; ellos concluyeron que, para el caso del mercado español, el riesgo de una acción se puede explicar utilizando variables contables y macroeconómicas, siendo las variables estadísticamente más significativas el Tamaño de la empresa, la razón EBIT (Utilidades antes de Intereses e Impuestos) / Ventas, la tasa Euribor, la tasa de Desempleo, y el Producto Interno Bruto (PIB), para los dos sectores que analizaron: industrial y financiero; asimismo, encontraron que la adopción de las *Normas Internacionales de Información Financiera* (NIIF) tuvo un impacto en la relación de las variables independientes (contables y económicas) y el riesgo de las acciones.

En lo que respecta en concreto al mercado mexicano, Trejo, Noguera y White (2015) hicieron un análisis de las razones financieras más utilizadas (Capitalización de Mercado, razón Valor de la Firma / EBITDA⁺⁺⁺, Margen EBITDA, Rendimiento de Inversión ó ROI, tasa de Dividendo, tasa de Flujo Libre de Caja, razón Deuda Neta / Capital, crecimiento de Utilidades por Acción ó UPA, y crecimiento de Ventas) por los analistas de capitales obteniendo como resultados que éstas tienen un poder de predicción de hasta un año para los rendimientos de las acciones analizadas.

⁺⁺⁺ Utilidades antes de Impuestos, Intereses, Depreciación, y Amortización.

La relevancia del uso de múltiplos parte del supuesto de que los inversionistas evalúan el precio de una acción (juzgan si está sobre, sub, o justamente valuada), considerando lo que ésta compra de la compañía (CFA Institute, 2016). Su principal ventaja como indicadores de valor es que resumen en un número el valor de mercado de la acción y un valor fundamental, y siendo fáciles de usar y comunicar, permiten un mejor análisis comparativo entre competidores, así como el desempeño de una compañía respecto a la industria que pertenece. Si bien, los estados financieros[‡] se reportan cada trimestre en la Bolsa Mexicana de Valores, dado que los múltiplos involucran una variable de mercado, es posible aumentar la frecuencia de los datos (e.g., de trimestral a mensual) y generar así, un mejor estimado de la perspectiva de los inversionistas respecto de la compañía y el valor de la acción.

Debido a la disponibilidad de la información y que se buscaba utilizar los datos en la menor frecuencia posible, las variables fueron recabadas con una frecuencia mensual. Esto también resultó en el uso de variables alternativas a las mencionadas previamente, como fue el Indicador Global de Actividad Económica (IGAE)^{‡‡} en lugar del PIB; o, el utilizar múltiplos que consideren la variable Ventas con otro dato, como fueron el tamaño de la compañía (Capitalización de Mercado) y el Valor de la Firma.

[‡] Reportes que son la representación estructurada de la situación y desarrollo financiero de una entidad a una fecha determinada o un período definido, cuyo propósito es proveer información de la misma acerca de su posición financiera, del resultado de sus operaciones, y de los cambios en su capital contable y recursos, que son útiles en el proceso de la toma de decisiones económicas (CINIF, 2015).

^{‡‡} Su cálculo sigue una metodología similar o cercana a la que se utiliza para determinar el PIB trimestral.

Preprocesamiento de datos

Primero se realizó un análisis de correlación entre las variables independientes con el fin de detectar una posible redundancia entre éstas. La medida de correlación muestra la relación lineal entre dos variables (ecuación 2.16), es decir, el grado de movimiento conjunto que éstas presentan. La correlación toma los valores que van de -1 a +1, siendo +1 una relación lineal perfecta positiva, -1 una relación lineal perfecta negativa, y con un valor de cero o cercano a cero mayor será el grado de independencia, aunque no significa que exista una independencia total entre las variables; el signo positivo indica que las variables se mueven en la misma dirección, mientras que el signo negativo indica que se mueven en direcciones opuestas.

$$\rho_{i,j} = \frac{Cov_{i,j}}{\sigma_i \sigma_j} \quad (2.16)$$

Donde,

σ_i es la desviación estándar del activo i ;

σ_j es la desviación estándar del activo j ;

$Cov_{i,j}$ es la covarianza entre los activos i y j , se define como

$$Cov_{i,j} = \frac{1}{m-1} \sum_{i=1}^m (r_i - \mu_i)(r_j - \mu_j)$$

A su vez, se llevó a cabo la prueba de *Factor de Inflación de la Varianza* (VIF, por sus siglas en inglés) para determinar si efectivamente existía un problema de *multicolinealidad*.^{***} Dicha prueba se define como la razón de la varianza observada y la que habría sido en caso de que x_n (la variable n -ésima) no estuviera correlacionada con el resto de las variables independientes (ecuación

^{***} Situación en la que se presenta una fuerte correlación entre las variables explicativas del modelo, lo que puede resultar en problemas estadísticos al realizar el pronóstico o determinar el efecto que tiene cada variable independiente en la variable dependiente.

2.17), es decir, ésta identifica la correlación entre las variables independientes y la fuerza de esa correlación. Los valores que puede tomar van de 1 (si no hay correlación de la variable respecto a otras) hasta infinito (hay presencia de una muy alta correlación), valores de 1 a 5 sugieren que hay una moderada correlación, pero no tan severa como para tomar medidas correctivas.

$$VIF_n = \frac{1}{1-R_n^2} \quad (2.17)$$

Siendo, R_n^2 el coeficiente de determinación⁺ cuando x_n es pronosticado contra todas las otras variables independientes.

Una vez realizadas las pruebas anteriores, y detectada la multicolinealidad entre los datos, en el presente trabajo se procedió con una extracción de factores mediante el método de Análisis de Componentes Principales (PCA, por sus siglas en inglés), el cual consiste en una transformación ortogonal (i.e., las correlaciones entre las variables individuales se descomponen a modo de poder evaluar la significancia de dichas variables de manera independiente) de los datos, y de esta manera, se calculan los vectores que tengan la mayor varianza asociada con ellos.

Para el PCA, la descomposición se hizo por el método de Valores Singulares (SVD, por sus siglas en inglés). El SVD consiste en la factorización de una matriz en “vectores singulares” y “valores singulares”. En particular, establece que cualquier matriz A que sea $m \times n$ puede descomponerse en tres matrices (ecuación 2.18), donde cada una tiene características particulares: Q y K son definidas como columnas ortonormales (i.e., las columnas son ortogonales

⁺ Es el porcentaje de la variación total en la variable dependiente, explicada por la variable independiente. Se obtiene de elevar al cuadrado el coeficiente de correlación ρ^2 .

entre ellas y la matriz multiplicada por su transpuesta resulta en la matriz de Identidad), mientras que S es definida como una matriz diagonal. Los elementos ubicados en la diagonal de S son conocidos como “valores singulares” de la matriz, y a las columnas de Q y K se les denomina como “vectores singulares” de la izquierda y derecha, respectivamente.

$$A_{m \times n} = Q_{m \times m} S_{m \times n} K_{n \times n}^T \quad (2.18)$$

La selección de los componentes principales que se utilizaron para la red neuronal se hizo mediante el uso del método *validación cruzada con k-iteraciones* (k-CV). El método de validación cruzada (CV) se basa en la idea de repetir los procesos de entrenamiento y evaluación en diferentes subconjuntos de la base de datos original. Si D_i y \bar{D}_i son los subconjuntos de datos de Entrenamiento y Validación de un conjunto D , y el proceso CV entrena un algoritmo dado, m veces para encontrar un modelo adecuado que minimice la función logarítmica de la verosimilitud, la ecuación se puede expresar como

$$CV = -\frac{1}{m} \sum_{i=1}^m \ln \left(E(\hat{M}(\bar{D}_i) | D_i) \right) \quad (2.19)$$

Donde,

$\hat{M}(\bar{D}_i)$ es los estimados de los parámetros de máxima verosimilitud en \bar{D}_i ;

$E(\hat{M}(\bar{D}_i) | D_i)$ es la verosimilitud evaluada en el conjunto D_i .

En específico, el método *k-CV* consiste en dividir de manera aleatoria el conjunto D en k subconjuntos no superpuestos y del mismo tamaño. Cada subconjunto es entonces usado como conjunto de Prueba, con las restantes $k - 1$ iteraciones como conjunto de Entrenamiento, para posteriormente obtener el error de prueba estimado del promedio de los k errores de prueba

obtenidos (ecuación 2.20). Como medida de error se utilizó el RMSE, y un valor $k = 10$.

$$CV_k = \frac{1}{k} \sum_{i=1}^k RMSE_i \quad (2.20)$$

$$RMSE_i = \left[\frac{1}{2} \sum_i (y_i - o_i)^2 \right]^{1/2} \quad (2.21)$$

Conjuntos de Entrenamiento, Validación y Prueba

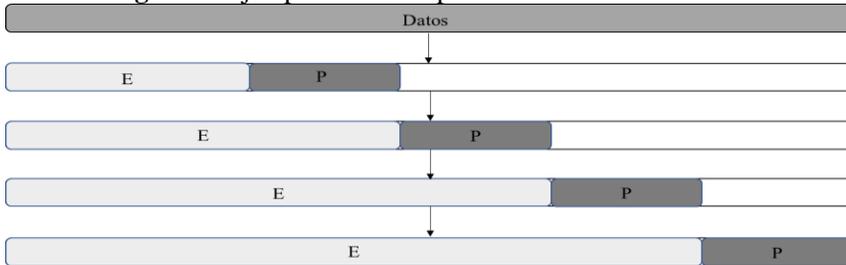
Para poder crear un modelo de redes neuronales óptimo, es necesario dividir la base de datos en 3 conjuntos: Entrenamiento, Validación, y Prueba. El primero, como su nombre lo indica, es empleado en la fase de “entrenamiento”, es decir, son los datos usados para desarrollar modelos. El segundo conjunto, es necesario a manera de “calibrar” los modelos generados en la fase de entrenamiento; con este conjunto se selecciona el modelo con el mejor desempeño, permite detectar si el modelo de redes se encuentra sub o sobre ajustado a los datos deseados. Y el tercer conjunto, es usado para obtener los resultados finales del modelo seleccionado de predicción.

Si bien Hagan et al. (2014) recomiendan dividir los conjuntos en una proporción de 70:15:15, en el presente trabajo se consideró utilizar el último 20% de los datos de las series para el conjunto de Prueba y, mediante el método de k -CV hacer la selección óptima de los conjuntos de Entrenamiento y Validación (Du & Swamy, 2014).

En el caso de series de tiempo, se hace una modificación al k -CV, ya que la serie podría tener patrones temporales, por lo tanto, éste se realiza de forma secuencial en los $k - 1$ subconjuntos de Prueba, es decir, en la primer iteración tomamos el primer subconjunto como conjunto de Entrenamiento y evaluamos en el segundo subconjunto; en la segunda iteración tomamos los subconjuntos

1 y 2 para la fase de entrenamiento, y evaluamos en el subconjunto 3; y así, sucesivamente hasta llegar a la iteración $k - 1$ (figura 7).

Figura 8. Ejemplo de k -CV aplicado de forma secuencial



Nota: El subconjunto E es el conjunto de Entrenamiento, el subconjunto P es el conjunto de Prueba, y el de color blanco son los subconjuntos que no se toman en consideración hasta la siguiente iteración. Fuente: Elaboración propia.

Arquitectura de la Red

Capas Ocultas

Las capas ocultas permiten que la red diseñe una función que generalice los datos analizados. En la práctica el uso de 1 o 2 capas ocultas es lo más común dado que el uso de un mayor número de capas, no sólo implica un aumento del tiempo computacional, sino también del riesgo de sobreajuste. El sobreajuste consiste en un modelo de predicción con pocos grados de libertad, es decir, la función se ajusta a los datos individuales en lugar de aprender un patrón general. Por ello se utilizó una capa oculta en el modelo de redes.

Modelo de Entrenamiento

El modelo de entrenamiento usado para el aprendizaje de la red es el de BPTT. Partiendo del modelo de retropropagación estándar, el cual, es el algoritmo para determinar cómo cada paso de entrenamiento debe calibrar los pesos (U , V y W) y los sesgos (b y d) en términos de la proporción relativa a aquellos cambios que causan el descenso más rápido del costo/error

$$\nabla_{\phi} L = \frac{\partial L}{\partial \phi} \quad (2.22)$$

Donde,

ϕ toma los valores de (U, V, W, b, y d).

El modelo BPTT se aplica de manera similar, salvo que primero se realiza el proceso de despliegue de la red, que consiste en replicar los tiempos de la red recurrente obteniendo una red feedforward equivalente (ver figura 6); posteriormente, la red feedforward resultante se puede entrenar usando el algoritmo de retropropagación estándar. Siendo la función de error la ecuación (1.45) el gradiente de los datos estimados en el tiempo t , para todo i , t , es

$$\nabla_{y^{(t)}} L = \frac{\partial L}{\partial y_i^{(t)}} = y_i^{(t)} - o_i \quad (2.23)$$

Comenzando al final de la secuencia, en el tiempo final T , $h^{(T)}$ sólo tiene a $y^{(T)}$ como descendiente, por lo que el gradiente se denota como

$$\nabla_{h^{(T)}} L = \frac{\partial y^{(T)}}{\partial h^{(T)}} = V^T \nabla_{y^{(T)}} L \quad (2.24)$$

Se procede con la iteración hacia atrás para retropropagar los gradientes a través del tiempo, de $t = T - 1$ hacia $t = 1$. Dado que $h^{(t)}$ tiene como descendientes $o^{(t)}$ y $h^{(t+1)}$, el gradiente está dado por

$$\nabla_{h^{(t)}} L = \left(\frac{\partial h^{(t+1)}}{\partial h^{(t)}} \right)^T (\nabla_{h^{(t+1)}} L) + \left(\frac{\partial y^{(t)}}{\partial h^{(t)}} \right)^T (\nabla_{y^{(t)}} L) \quad (2.25)$$

Integrándose la propagación hacia atrás en la celda de memoria para el tiempo t de la siguiente manera:

$$\epsilon_c^t = \frac{\partial L}{\partial h^t} \quad (2.26) \quad \epsilon_s^t = \frac{\partial L}{\partial s^t} \quad (2.27)$$

Salidas de la Celda

$$\delta_i^t = \sigma'(a_i^t) \sum_{c=1}^C g^t \epsilon_s^t \quad (2.28)$$

Puertas de Salida

$$\delta_f^t = \sigma'(a_f^t) \sum_{c=1}^C s^{t-1} \epsilon_s^t \quad (2.29)$$

Estados

$$\delta_c^t = i^t H'(a_g^t) \epsilon_s^t \quad (2.30)$$

Celdas

$$\epsilon_s^t = q^t H'(s_c^t) \epsilon_c^t + f^{t+1} \epsilon_s^{t+1} + w_{cl} \delta_l^{t+1} + w_{cf} \delta_f^{t+1} + w_{cq} \delta_q^{t+1} \quad (2.31)$$

Puertas de Olvido

$$\delta_q^t = \sigma'(a_q^t) \sum_{c=1}^C H(s_c^t) \epsilon_c^t \quad (2.32)$$

Puertas de Entrada

$$\epsilon_s^t = \sum_{k=1}^K w_{ck} \delta_k^t + \sum_{z=1}^Z w_{ch} \delta_z^{t+1} \quad (2.33)$$

Donde,

C es el conjunto de celdas de memoria, K el número de salidas, y Z el número de celdas en el estado oculto.

Dado que el cómputo completo de este modelo implicaría un gran costo computacional, por el tiempo que tomaría su entrenamiento, se subdividen los datos de manera aleatoria en mini-lotes y se hace el cómputo respecto a un mini-lote. Yendo de manera repetida a través de todos los lotes y haciendo estos ajustes, se convergerá hacia un mínimo-local de la función del error. A este proceso se le denomina *Descenso de Gradiente Estocástico*.

De manera general, el proceso del BPTT es el siguiente (Ciaburro & Venkateswaran, 2017):

1. Se despliega la RNN para contener n redes feedforward.
 2. Se inician los pesos w con valores aleatorios.
-

-
3. Los insumos para cada red se establecen con valores x_i .
 4. Los datos de entrada se propagan hacia adelante sobre toda la red desplegada.
 5. Se retropropaga el error a través de toda la red desplegada.
 6. Se actualizan todos los pesos en la red.
 7. Se promedian los pesos para encontrar el peso final en la red plegada (contraída).
 8. Los pasos 3 – 7 se repiten hasta que los criterios de alto sean alcanzados o que la cantidad de iteraciones se haya terminado.

Usualmente se consideran entre 100 y 500 iteraciones (dependiendo del problema en cuestión). En el presente trabajo se usaron de 100 a 150 nodos de entrenamiento, de conformidad con el nivel que necesitaban los datos empleados.

Evaluación de Resultados

Para evaluar la exactitud del pronóstico se usaron los métodos del coeficiente de Theil (ecuación 2.34) y el RMSE (ecuación 2.21). El coeficiente de desigualdad de Theil (U) es uno de evaluadores estadísticos de pronóstico comúnmente observado en la literatura (Bliemel & MacKay, 1973). Si existe el caso de un pronóstico “perfecto”, es decir, todos los cambios estimados i son iguales a los cambios observados i , el coeficiente tendrá un valor de cero. Cuando el coeficiente toma un valor de 1 se debe a que el método de pronóstico da estimaciones con el mismo error estándar que la simple extrapolación sin cambios; si el coeficiente supera el valor de 1 el método usado será rechazado puesto que no puede superar el método más simple de extrapolación sin cambios.

$$U_{II} = \frac{[\sum_{i=1}^n (P_i - A_i)^2]^{1/2}}{[\sum_{i=1}^n A_i^2]^{1/2}} \quad (2.34)$$

Teniendo,

(P_i, A_i) como los *cambios* estimados y observados, respectivamente.

2.2. Modelos de VaR Tradicionales

Los modelos de riesgo de mercado empleados como punto de comparación fueron: Simulación Histórica con alisado y Simulación Monte Carlo, ya que éstos son comúnmente usados por las áreas de riesgos financieros, ya sea por su fácil implementación y/o eficacia.

2.2.1. Simulación Histórica con alisado

La Simulación Histórica con alisado es una extensión al modelo de Simulación Histórica (Hull, 2012). En esta simulación, en contraste con la clásica Simulación Histórica, no se le da el mismo peso a los escenarios generados, sino que comúnmente se le asigna un mayor peso a los escenarios más recientes y uno menor a los más antiguos, esto se lleva a cabo mediante la aplicación de un factor de descuento similar al de la metodología *EWMA*. Conforme a los resultados obtenidos por Reyes (2015), ésta presenta una mayor eficiencia respecto de los métodos de Simulación Histórica sin alisado y la Simulación Monte Carlo, ya que incorpora más rápido la información de los eventos más recientes; lo anterior, va acorde a los resultados presentados por el Grupo BBVA en su informe financiero 2016-Pilar III (2017), donde señalan que la Simulación Histórica con alisado se adapta con mayor rapidez a los cambios en las condiciones de los mercados financieros, y que tenderá a ser superior que la Simulación sin alisado en los momentos que se produzcan repuntes en la incertidumbre de dichos mercados. Para su cálculo se realizó lo siguiente:

-
- A partir del vector de precios $S = [s_1, \dots, s_t]$, donde t es la fecha más reciente, se determinan los rendimientos x_i , siendo $i = 1, \dots, m$.
 - Se realiza la revaluación del instrumento con los rendimientos x_i y la observación más reciente s_t , generando un vector de precios $F = [f_1, \dots, f_m]$, donde $f_i = s_t(1 + x_i)$. Y se procede con la obtención del vector de pérdidas y ganancias $P = [(s_t - f_1), \dots, (s_t - f_m)]$.
 - Teniendo el vector P , se realiza el proceso de alisado mediante la ponderación de los escenarios utilizando la función $w_i = \lambda^{m-i}(1 - \lambda)m$; puesto que los pesos deben sumar la unidad, el peso dado el escenario i se puede expresar como $\lambda^{m-i}(1 - \lambda)/(1 - \lambda^m)$.
 - El VaR se obtiene ordenando la matriz P del peor al mejor resultado p_i y se obtiene el percentil deseado dado un nivel de confianza $(1 - \alpha)$, con base en la distribución w_i .

2.2.2. Simulación Monte Carlo

Como ya se mencionó, la Simulación Monte Carlo consiste en la simulación de escenarios, a partir de valores generados de manera aleatoria, con el fin de estimar la dispersión de los precios de un instrumento financiero. Para estimar el VaR se realizaron los siguientes pasos:

- A partir del vector de precios $S = [s_1, \dots, s_t]$, donde t es la fecha más reciente, se determinan los rendimientos x_i , siendo $i = 1, \dots, m$; de los cuales obtendremos la media μ y desviación estándar σ de los datos.
- Asumiendo que los precios siguen un proceso *GBM*, se genera una secuencia de números aleatorios distribuidos normalmente ξ_j , donde $j = 1, \dots, n$; que usaremos para la revaluación completa del instrumento considerando la observación del vector de precios en el tiempo t , una

tendencia $\mu\Delta t$ y el factor de incertidumbre $\sigma\xi_j\sqrt{\Delta t}$. La secuencia se realiza hasta alcanzar el horizonte objetivo T , es decir, cuando $s_{t+n} = s_T$.

- El paso anterior se repite hasta generar k escenarios, a partir de los cuales, se construye el vector de pérdidas y ganancias $P^k = [(s_t - s_T^1), \dots, (s_t - s_T^k)]$.
- Dada la distribución de pérdidas y ganancias, el VaR se obtiene tomando el percentil de la distribución considerando un nivel de confianza $(1 - \alpha)$.

2.3. Pruebas de Backtesting

Se aplicaron las pruebas de Kupiec (LRuc) y Christoffersen (LRcc) para evaluar la eficiencia de los modelos considerando hacerlo con los niveles de confianza del 95% (para tener una base comparativa de eficiencia) y 99% (para observar la eficiencia en los casos de mayor riesgo).

La prueba LRuc consiste en evaluar la correcta cobertura incondicional de la cola izquierda de la distribución del rendimiento marginal, es decir, examina la relación entre el número esperado de excesos respecto al valor del VaR (implícitos por el nivel α seleccionado durante el período de pronóstico, i.e., αM), y las violaciones al VaR realizadas (observadas a partir de los datos reales).

La prueba LRcc, hace la evaluación del modelo de VaR respecto de la correcta cobertura condicional de la distribución, contemplando no sólo una prueba de cobertura incondicional, sino también la independencia de los excesos. Lo anterior se refiere a que no sólo toma en cuenta la frecuencia de los valores excedentes al VaR, también toma en consideración sus dinámicas temporales, ya que los modelos de VaR asumen que estos excedentes deberían estar

independientemente distribuidos a través del tiempo. Si la correcta cobertura condicional se alcanza con el modelo empleado, los excesos del VaR deberían estar independientemente distribuidos.

Considerar que se cumpla esta condición es de tanta importancia como la propiedad de la cobertura incondicional, por el hecho de que, si los excesos se pueden pronosticar o presentan algún tipo de agrupamiento detectable, quiere decir que el modelo de riesgo no captura lo suficientemente rápido la nueva información ante las condiciones cambiantes del mercado, y que resultaría en que la institución carezca de fondos suficientes para cubrir los requerimientos de capital por períodos prolongados durante los momentos de un riesgo en aumento. Un beneficio adicional de esta prueba es que genera información sobre las probabilidades condicionales de observar un exceso seguido de otro en el modelo de riesgo seleccionado, así como el número promedio de días entre cada uno de éstos.

En cuanto a la relación de los excesos esperados contra los actuales, se define formalmente como la razón de excedentes Actuales contra los Estimados (AE)

$$AE = \frac{\sum_{n=1}^M d_{t+n}}{\alpha M} \quad (2.35)$$

Siendo $d_{t+s} \equiv \mathbf{1}_{\{r_{t+n} < VaR_{t+n}(\alpha)\}}$ y $\mathbf{1}_{\{\cdot\}}$ la función indicadora que es igual a uno si la condición se cumple, cero en caso contrario. Por lo tanto, si AE es mayor a uno, el modelo subestima el riesgo y podría resultar en una gran pérdida no esperada; y si éste es menor a uno, el modelo es demasiado conservador lo que conllevaría a la sobreestimación del requerimiento de capital.

A su vez, se utilizó el Error de Cuantil medio (Avg_QL) para determinar el rendimiento superior o inferior entre los modelos. El Error de Cuantil (QL) es una función de error utilizada para las regresiones de cuantil y comúnmente usada en el contexto de modelos de VaR (Ardia, Boudt, & Catania, 2016). Para un pronóstico de VaR, con un nivel α dado para el tiempo $t + 1$, el error de cuantil asociado se define como

$$QL_{t+1}(\alpha) = (\alpha - d_{t+1})(y_{t+1} - VaR_{t+1}(\alpha)) \quad (2.36)$$

Puesto que la función QL penaliza en mayor medida las observaciones que exceden al VaR, los QLs son promediados por el período de pronóstico, así, los modelos con los promedios más bajos se considerarán como los mejores. Para determinar la superioridad de un modelo respecto a otro, se realiza una división de sus Avg_QL asociados; e.g., para un modelo X y un modelo Z , si $\frac{Avg_QL_X}{Avg_QL_Z} < 1$, entonces el modelo X es superior al modelo Z , y viceversa.

Para aplicar las pruebas de backtesting, se creó una serie de ganancias y pérdidas ficticia (sin considerar comisiones ni impuestos) con base en las series de precios máximo y mínimo de los datos mensuales. Se seleccionó un precio de manera aleatoria sobre dicho rango y se obtuvo el rendimiento dividiendo el precio obtenido respecto al precio de cierre del mes anterior, a manera de simular la apertura y cierre de una nueva posición por mes para cada acción y durante el período de prueba seleccionado.

Capítulo 3. Aplicación del Valor en Riesgo en las acciones mexicanas: Caso de América Móvil y TV Azteca

Lo que es estático y repetitivo, es aburrido.

Lo que es dinámico y aleatorio, es confuso.

En medio, se encuentra el arte.

John Locke

3.1. Metodología mediante Redes Neuronales

El análisis de esta tesis se realizó para dos acciones, una muy bursátil (América Móvil Serie L, “AMXL”) y otra de baja bursatilidad (TV Azteca CPO, “AZTECA”).

3.1.1. Análisis y preprocesamiento de datos

La serie de precios seleccionada comprende el período del 31 de diciembre de 2004 hasta el 31 de marzo de 2017, a partir de éstos, se obtuvo la serie diaria de rendimientos logarítmicos, así como los rangos logarítmicos. La tabla 5 muestra el resumen de la estadística descriptiva de los datos diarios (rendimientos y rangos) de ambas acciones, sin considerar el conjunto de Prueba.

Si bien todas las series tienen una media cercana a cero, éstas presentan una distribución de datos con skew positivo (i.e., los datos tienen una asimetría hacia el lado derecho de la distribución) salvo los rendimientos logarítmicos de AZTECA que tienen un bajo skew negativo (con una ligera asimetría hacia el lado izquierdo de la distribución), y una curtosis leptocúrtica (indica que la distribución es de colas más pesadas que una distribución normal), sobre todo en el caso de los rangos logarítmicos, lo que se puede interpretar como una mayor probabilidad de ocurrencia de valores extremos; es común observar estas características en la literatura de volatilidad, por su comportamiento de colas pesadas. En cuanto a las pruebas de normalidad, todas rechazaron la hipótesis nula para todas las series, por lo tanto, se rechaza la idea de que las series tienen una distribución normal.

Tabla 5. Estadística Descriptiva de los rendimientos y rangos logarítmicos de AMXL y AZTECA, período del 3 de enero de 2005 al 31 de octubre de 2014

		r_{AMXL}^d	R_{AMXL}^d	r_{AZTECA}^d	R_{AZTECA}^d
m		2,462	2,462	2,461	2,461
Media		0.0005	0.0014	0.0000	0.0014
Mediana		0.0006	0.0004	0.0000	0.0007
Desv. Est.		0.0190	0.0108	0.0191	0.0024
Skew		0.0272	26.480	-0.0850	6.3027
Curtosis		5.8558	846.26	12.4884	62.7161
Mínimo		-0.1374	0.0000	-0.2179	0.0000
Máximo		0.1255	0.3951	0.1422	0.0403
Error Est.		0.0004	0.0002	0.0004	0.0000
Pruebas de Normalidad					
Anderson-Darling	Stat	23.871	800.05	35.207	320.87
	p-valor	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Jarque-Bera	Stat	3526.6	7.39E+07	16028	420330
	p-valor	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Kolmogorov-Smirnov	Stat	0.0621	0.4490	0.0811	0.2765
	p-valor	2.20E-16	2.20E-16	2.20E-16	2.20E-16

Fuente: Elaboración propia con datos de S&P Capital IQ.

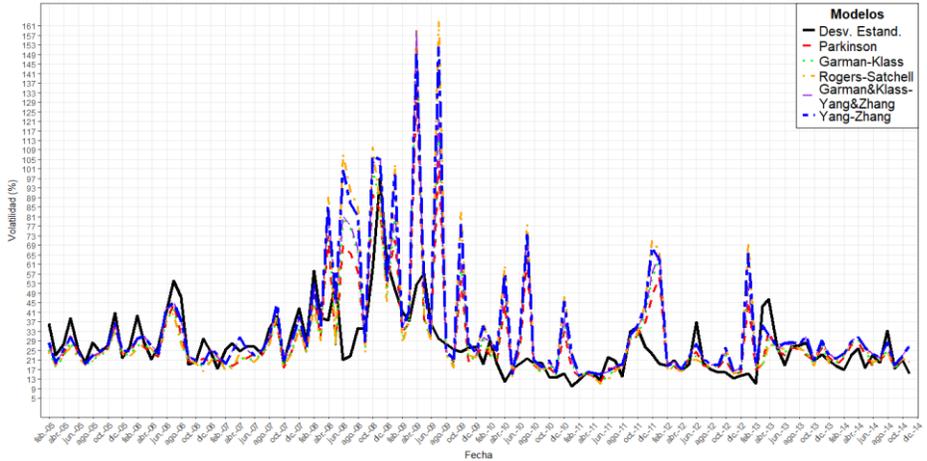
Volatilidad

La serie de volatilidad comprende del período de enero 2005 a marzo 2017, quedando delimitada a octubre 2014 en el conjunto de Entrenamiento-Validación. Las series (figuras 9 y 10) muestran la volatilidad anualizada de cada mes. Para AMXL se puede observar una diferencia muy marcada entre el modelo clásico de desviación estándar respecto de las volatilidades de rango, en particular en el 2009, año en el que AMXL había declarado y llevado a cabo su consolidación como negocio con la empresa Teléfonos de México, S.A.B. de C.V. (Telmex), además de que también cuenta con un ADR⁺⁺ (por sus siglas en inglés de *American Depositary Receipt*) en las bolsas de Estados Unidos de América (EE.UU.). En cuanto a AZTECA, se puede observar un comportamiento similar entre el modelo clásico y los de rango, siendo los picos

⁺⁺ Recibo de Depósito Americano que se emite en Bolsas de Valores estadounidenses al amparo de títulos accionarios emitidos en un país extranjero que se depositan en un fideicomiso, de tal manera que cada ADR está respaldado por una determinada cantidad de acciones a través de un fideicomiso.

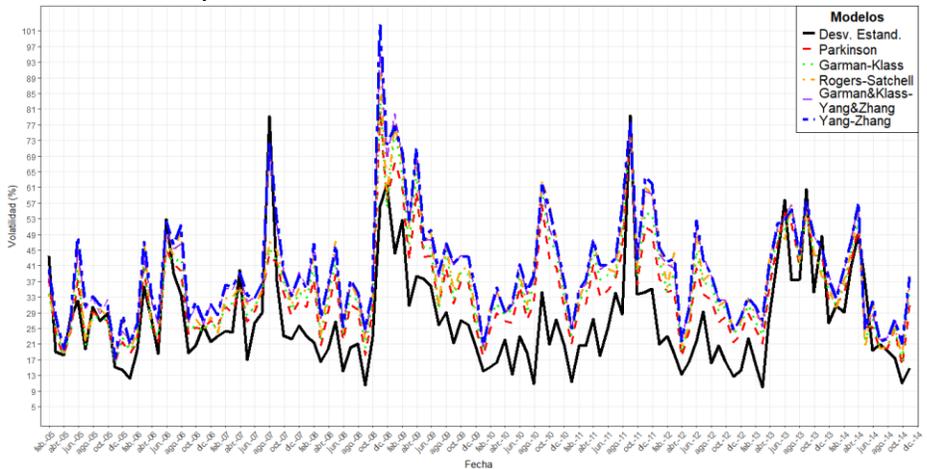
de volatilidad más marcados en el caso de los modelos de rango dado los datos adicionales que éstos contemplan.

Figura 9. Comparativo de estimadores de volatilidad histórica de AMXL, período de enero de 2005 a marzo de 2017



Fuente: Elaboración propia.

Figura 10. Comparativo de estimadores de volatilidad histórica de AZTECA, período de enero de 2005 a marzo de 2017



Fuente: Elaboración propia.

Los estimadores de rango presentan una eficiencia superior para el caso de AZTECA que para AMXL (tabla 6), lo que sugiere que la segunda presenta grandes saltos (de apertura, durante la noche, máximo-mínimo), corroborándose así en la figura 9; lo anterior, aunado al hecho de que las series de rendimientos y rangos presentan una elevada curtosis y un alto skew, se puede traducir en una mayor probabilidad de ocurrencia de valores extremos resultando en un mayor nivel de riesgo.

Tabla 6. Resumen de la Eficiencia de los Estimadores de Volatilidad Histórica, período de enero 2005 a octubre 2014

	Precios	AMXL	AZTECA
Desv. Estándar	C	1	1
Parkinson	HL	0.7477	1.3520
Garman-Klass	OHLC	0.6606	1.3593
Rogers-Satchell	OHLC	0.5873	1.3456
GK-YZ	OHLC	0.7073	1.3863
Yang-Zhang	OHLC	0.6694	1.3722

Fuente: Elaboración propia.

Variables independientes

Además de los 4 factores de mercado (rendimiento de la acción, rango de precios apertura-cierre, rango de precios máximo-mínimo, volumen de operación), se seleccionaron 6 factores económicos, 4 de la industria, y 9 de la compañía. Las variables seleccionadas se describen en la tabla 7, tomando el período de enero 2005 a marzo 2017.

Tabla 7. Variables independientes

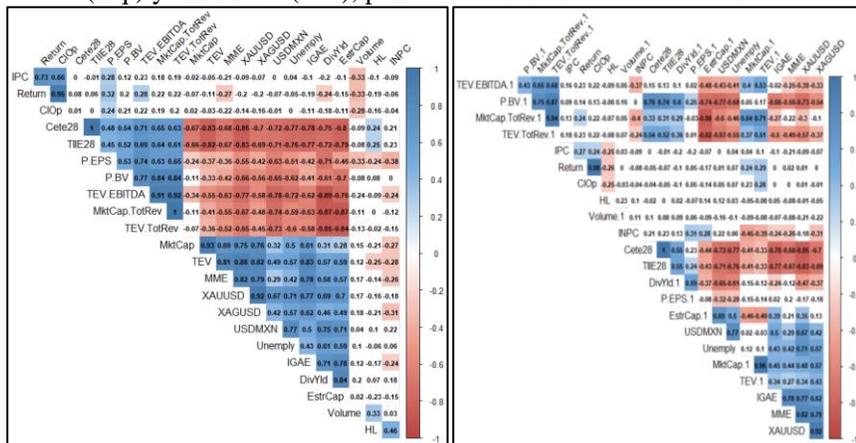
No	Variable	Tipo	Fuente
1	Rendimiento de la acción (Return)	<i>Mercado</i>	<i>Elaboración propia con datos de S&P Capital IQ</i>
2	Rango de precios de Cierre-Apertura de la acción (CIOp)	<i>Mercado</i>	<i>Elaboración propia con datos de S&P Capital IQ</i>
3	Rango de precios Max-Min de la acción (HL)	<i>Mercado</i>	<i>Elaboración propia con datos de S&P Capital IQ</i>
4	Volumen de Operación (Volume)	<i>Mercado</i>	<i>S&P Capital IQ</i>
5	Tipo de Cambio USDMXN Spot (USDMXN)	<i>Económica</i>	<i>S&P Capital IQ</i>
6	Tasa Cete 28 días (Cete28)	<i>Económica</i>	<i>Banco de México</i>
7	Tasa TIIE 28 días (TIIE28)	<i>Económica</i>	<i>Banco de México</i>
8	Tasa de Inflación (INPC)	<i>Económica</i>	<i>Banco de México</i>
9	Índice Global de Actividad Económica (IGAE)	<i>Económica</i>	<i>Banco de México</i>
10	Tasa de Desempleo (Unemploy)	<i>Económica</i>	<i>S&P Capital IQ</i>
11	Rendimiento del Índice de Referencia (IPC)	<i>Industria</i>	<i>Elaboración propia con datos de S&P Capital IQ</i>
12	Precio de Exportación del Petróleo Crudo (MME)	<i>Industria</i>	<i>PEMEX Base de Datos Institucional</i>
13	Precio del oro spot (XAUUSD)	<i>Industria</i>	<i>Investing</i>
14	Precio de la plata spot (XAGUSD)	<i>Industria</i>	<i>Investing</i>
15	Tasa de Dividendo (DivYld)	<i>Compañía</i>	<i>S&P Capital IQ</i>
16	Capitalización del Mercado (MktCap)	<i>Compañía</i>	<i>S&P Capital IQ</i>
17	Valor de la Firma (TEV)	<i>Compañía</i>	<i>S&P Capital IQ</i>
18	Razón Valor de Mercado/Ventas Totales (MktCap.TotRev)	<i>Compañía</i>	<i>S&P Capital IQ</i>
19	Razón Valor de la Firma/EBITDA	<i>Compañía</i>	<i>S&P Capital IQ</i>
20	Razón Valor de la Firma/Ventas Totales (TEV.TotRev)	<i>Compañía</i>	<i>S&P Capital IQ</i>
21	Razón de Precio/Valor en Libros (P.BV)	<i>Compañía</i>	<i>S&P Capital IQ</i>
22	Razón Precio/UPA (P.EPS)	<i>Compañía</i>	<i>S&P Capital IQ / Economática</i>
23	Estructura de Capital (EstrCap)	<i>Compañía</i>	<i>Economática</i>

Fuente: Elaboración propia.

Preprocesamiento de Datos

Se hizo el análisis de correlación (figura 11) entre las variables independientes; de los resultados obtenidos se puede concluir que una gran cantidad de los datos presentan altas correlaciones entre éstos (ya sea positiva o negativa). Pudiendo observar, la relación positiva entre las tasas (Cetes y TIIE) y las series de datos fundamentales de la compañía (salvo por la tasa de Dividendo, el Valor de la Firma y la Capitalización del Mercado), así como la relación negativa de éstas, respecto de las demás variables económicas, de la compañía y de la industria; cabe mencionar la baja correlación que presentan los rangos (Máximo-Mínimo, Cierre-Apertura) y el Volumen en toda la matriz, para el caso de ambas acciones.

Figura 11. Matriz de correlación de las variables independientes de AMXL (izq.) y AZTECA (der.), período de enero 2005 a octubre 2014



Nota: Las correlaciones mayores a 0.24 son significativas en un nivel de 0.01. Fuente: Elaboración propia con los datos descritos en la tabla 7.

Los resultados de la prueba VIF se muestran en la tabla 8, donde se observa que, si bien los rangos, el volumen, el IPC, la inflación, y la razón Precio/UPA tienen una moderada correlación, los demás datos presentan valores muy elevados, sin mencionar que el VIF promedio en ambas acciones es mucho mayor a 1, por lo que es necesario hacer un tratamiento de los datos para ambas acciones.

Tabla 8. VIFs de las variables independientes de AMXL y AZTECA, período de enero 2005 a octubre 2014

	HL	CIOp	Volumne	MktCap	DivYld	TEV	MktCap_TotRev	TEV_EBITDA	TEV_TotRev	P_BV	P_EFS	EstCap	IPC	USDMXN	Cete28	TIE28	INPC	MME	XAUUSD	XAGUSD	IGAE	Unemply
AMXL	2.37	2.21	1.76	508.18	17.39	706.33	2,156.82	56.72	1,520.23	10.28	3.60	23.99	2.37	14.86	446.44	401.01	3.54	17.00	49.17	19.56	10.92	6.15
AZTECA	1.50	1.47	1.47	2,230.93	5.41	1,313.33	2,247.71	9.30	1,974.31	24.36	3.65	6.66	1.37	12.03	419.36	380.29	3.64	16.70	56.46	17.97	9.92	5.99
Media																						
AMXL	271.86																					
AZTECA	397.45																					

Fuente: Elaboración propia con los datos descritos en la tabla 7.

Una vez, detectada la multicolinealidad entre las variables independientes, se procedió a hacer la extracción de factores mediante el método PCA. Así, los componentes obtenidos (tablas 9 y 10) no presentan multicolinealidad. Se observa que a partir del séptimo componente ya se cubre el 0.915 de la varianza total en el caso de AMXL; para AZTECA, es a partir del octavo componente donde ya se tiene una proporción acumulada de la varianza de 0.9153.

Tabla 9. Resumen de importancia de los Componentes Principales (AMXL), período de enero 2005 a octubre 2014

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Desv. Estánd.	3.3812	1.8272	1.5940	1.2128	0.9083	0.8852	0.8090	0.7424	0.5750	0.5542	0.4612	0.3663
Prop. de Var.	0.4971	0.1452	0.1105	0.0640	0.0359	0.0341	0.0285	0.0240	0.0144	0.0134	0.0093	0.0058
Prop. Acum.	0.4971	0.6422	0.7527	0.8166	0.8525	0.8866	0.9150	0.9390	0.9534	0.9667	0.9760	0.9818
	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	
Desv. Estánd.	0.3313	0.2961	0.2552	0.2193	0.1929	0.1829	0.1303	0.1246	0.0570	0.0373	0.0144	
Prop. de Var.	0.0048	0.0038	0.0028	0.0021	0.0016	0.0015	0.0007	0.0007	0.0001	0.0001	0.0000	
Prop. Acum.	0.9866	0.9904	0.9932	0.9953	0.9969	0.9984	0.9991	0.9998	0.9999	1.0000	1.0000	

Fuente: Elaboración propia.

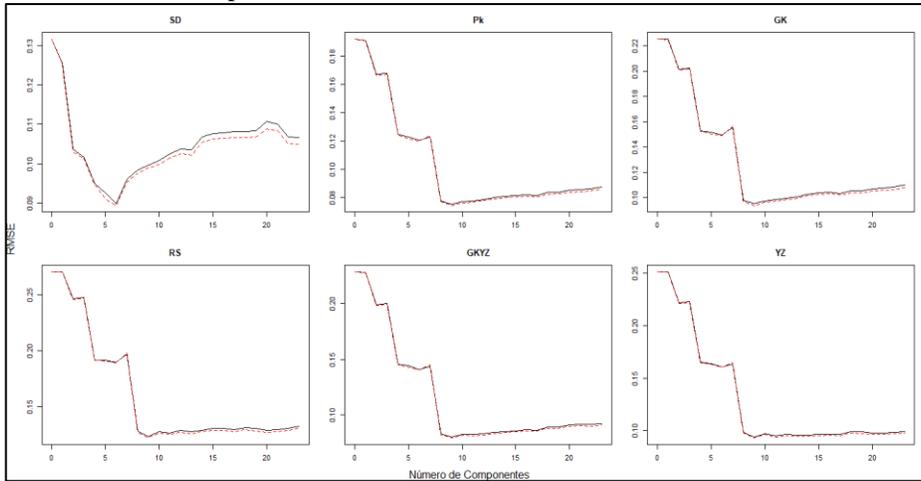
Tabla 10. Resumen de importancia de los Componentes Principales (AZTECA), período de enero 2005 a octubre 2014

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Desv. Estánd.	2.8940	2.1893	1.4940	1.3334	1.1141	0.9972	0.9318	0.8786	0.7655	0.6048	0.5527	0.4103
Prop. de Var.	0.3640	0.2084	0.0970	0.0773	0.0540	0.0432	0.0378	0.0336	0.0255	0.0159	0.0133	0.0073
Prop. Acum.	0.3640	0.5724	0.6695	0.7468	0.8007	0.8440	0.8817	0.9153	0.9407	0.9566	0.9699	0.9772
	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	
Desv. Estánd.	0.3878	0.3424	0.2637	0.2498	0.2278	0.1691	0.1255	0.1205	0.1082	0.0354	0.0113	
Prop. de Var.	0.0065	0.0051	0.0030	0.0027	0.0023	0.0012	0.0007	0.0006	0.0005	0.0001	0.0000	
Prop. Acum.	0.9838	0.9889	0.9919	0.9946	0.9969	0.9981	0.9988	0.9994	0.9999	1.0000	1.0000	

Fuente: Elaboración propia.

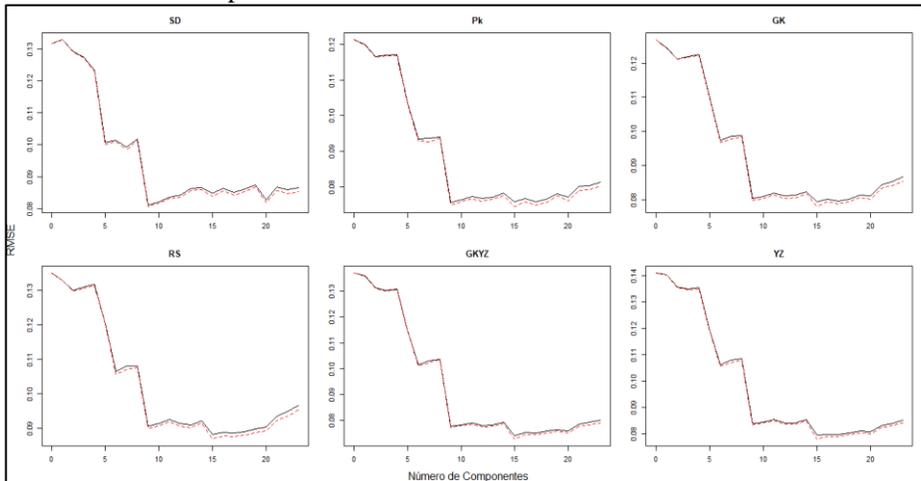
Entonces, se aplicó el método k -CV para la selección de los principales componentes. Los resultados obtenidos fueron: para AMXL, la selección de 6 componentes para el pronóstico de σ_{SD}^{Ann} , y 9 para los estimadores de rango (figura 12); en el caso de AZTECA, se seleccionaron 9 componentes para σ_{SD}^{Ann} , y 15 para los estimadores de rango (figura 13).

Figura 12. Validación de Componentes con el menor RMSE (AMXL),
período de enero 2005 a octubre 2014



Fuente: Elaboración propia.

Figura 13. Validación de Componentes con el menor RMSE (AZTECA),
período de enero 2005 a octubre 2014



Fuente: Elaboración propia.

3.1.2. Pronóstico de Volatilidad

Los resultados obtenidos por la red LSTM mostraron que el modelo no se rechaza en ninguno de los pronósticos al tener un coeficiente de Theil (U) menor a 1, destacando los modelos de Yang-Zhang y Parkinson con el menor coeficiente U y el menor RMSE, respectivamente, tanto en AMXL como en AZTECA (tabla 11).

Cabe señalar que, si bien el pronóstico no es cercano a los valores reales, en cierta medida el modelo detectó una tendencia en los datos, mostrándose con mayor claridad en el caso de AMXL, donde se puede observar que incluso en movimientos bruscos de alza y baja de la volatilidad fue capaz de anticipar la dirección de dicho cambio, salvo por los meses donde la volatilidad del mercado se vio principalmente afectada por temas cualitativos: e.g., los meses de abril, mayo y diciembre 2016, fechas en las que estaba latente la incertidumbre electoral de Estados Unidos de América (EE.UU.), así como la transición del ya electo presidente de EE.UU. Donald Trump. En el caso de AZTECA, el modelo no detectó una visible tendencia en los datos, pero en algunos meses fue capaz de anticipar el movimiento alcista/bajista de la volatilidad (ver Anexos 1 y 2).

Siendo el horizonte objetivo bastante amplio y dado los altos niveles de volatilidad que presentan ambas acciones, aunado a la limitada información de las series de tiempo que se disponía, se puede considerar como un modelo de pronóstico razonable, aunque debido a los resultados obtenidos en esta fase, se esperaba tener una subestimación del nivel de riesgo, situación que se corrobora y se detalla más adelante.

Tabla 11. Evaluación de Pronósticos de Volatilidad, período noviembre 2014 a marzo 2017

Volatilidad	AMXL		AZTECA	
	RMSE	U	RMSE	U
SD	6.8700	0.7723	13.8870	0.8610
Pk	3.8240	0.7441	11.2530	0.8594
GK	3.8550	0.7097	12.2640	0.8663
RS	4.1970	0.6924	12.7840	0.8261
GKYZ	4.9930	0.6763	11.8370	0.8525
YZ	5.1040	0.6645	11.7820	0.8228

Fuente: Elaboración propia.

3.1.3. VaR paramétrico mediante pronóstico de volatilidad

Conforme a los resultados previamente descritos, para realizar el cálculo del VaR paramétrico de ambas acciones se optó por utilizar los pronósticos obtenidos con los modelos de volatilidad de Yang-Zhang (LSTM_YZ) y Parkinson (LSTM_Pk).

Las pruebas de backtesting mostraron que ambos modelos no son efectivos para lidiar con el VaR, ya que los p-valores obtenidos en las pruebas de Kupiec (LRcc) y Christoffersen (LRuc) fueron de cero, o muy cercanos a éste tratándose de los resultados obtenidos por el modelo LSTM_YZ con un nivel de confianza del 95%.

Respecto a la prueba de Razón de excedentes Actuales contra los Esperados (AE), los modelos muestran claramente la subestimación del nivel de riesgo al tener un valor muy superior a 1; a su vez, ambos modelos presentan mejores resultados en el caso de AMXL al tener un menor valor y cuyo mejor desempeño es para un nivel de confianza del 95%. Lo anterior, sugiere que estos modelos no sólo no son buenos en la estimación del riesgo, sino que en los casos en los que éste sufra un incremento abrupto y por un período prolongado, van a subestimarlo demasiado y resultarán en pérdidas cuantiosas.

En términos de la Pérdida de Cuantil media (Avg_QL), el modelo LSTM_YZ sólo tuvo un rendimiento superior al LSTM_Pk por 11.76% en promedio, y siendo más marcada esta situación para un nivel de confianza de 99% en ambas acciones (tablas 12 y 13).

3.2. VaR mediante Simulación Histórica con alisado

Los resultados que se obtuvieron de las pruebas de backtesting para la Simulación Histórica con alisado (SHa) mostraron su efectividad al momento de hacer pronósticos para la gestión del riesgo, ya que los p-valores de las pruebas LRcc y LRuc con un nivel de confianza de 99% fueron los más altos en ambas acciones; para un nivel del 95% se ajustó mejor el pronóstico con AZTECA que con AMXL (siendo superior la Simulación Monte Carlo para esta acción), y que AZTECA con un α del 1%, por lo que el modelo capturó de mejor manera y con mayor velocidad la información secuencial de la serie de rendimientos de AZTECA usando un nivel de confianza del 95%, generando así una mejor estimación del nivel de riesgo.

Respecto a la prueba AE, con un nivel del 99% el modelo fue muy conservador, y para un nivel del 95%, en el caso de AMXL subestimó el nivel de riesgo, mientras que en AZTECA tuvo el mejor desempeño al tener un estimado conservador, pero cercano a 1. En cuanto al Avg_QL, este modelo presentó un rendimiento superior a los modelos LSTM por un 73.07% en promedio; para el modelo MC tuvo un rendimiento superior de un 42.57% en promedio, sin considerar el caso de AMXL donde, usando un nivel de confianza del 99%, el segundo fue ligeramente superior por un 0.43% (tablas 12 y 13).

3.3. VaR mediante Simulación Monte Carlo

En el caso del VaR por Simulación Monte Carlo (MC), tuvo un mejor desempeño con la serie de AMXL que con la de AZTECA. Para las pruebas LRcc y LRuc, no se rechaza el modelo para AMXL, pero con AZTECA sí es el caso ya que los p-valores fueron de cero (con $\alpha = 0.01$) y muy cercano a cero (con $\alpha = 0.05$). Asimismo, en el caso de AMXL y con un nivel de confianza del 99% fue un poco inferior al método SHa.

El AE señala que este modelo subestimó el nivel de riesgo para todos los niveles de confianza, siendo AMXL (con un nivel de confianza del 95%), aquel punto donde mejores resultados mostró al tener un valor cercano a 1. Respecto al Avg_QL, el rendimiento de este modelo fue superior a los obtenidos de las redes LSTM por un 58.58% en promedio (tablas 12 y 13).

Tabla 12. Pruebas de Backtesting (AMXL),
período noviembre 2014 a marzo 2017

	$1 - \alpha$	LRuc (p-val)	LRcc (p-val)	AE	Avg_QL
MC		0.6568	0.7767	1.3793	0.0065
SHa	95%	0.0709	0.1002	2.7586	0.0057
LSTM_YZ		0.0001	0.0003	5.5172	0.0120
LSTM_Pk		0.0000	0.0000	6.2069	0.0131
MC		0.3002	0.5634	3.4483	0.0016
SHa	99%	0.4452	0.7472	0.0000	0.0016
LSTM_YZ		0.0000	0.0000	17.2414	0.0071
LSTM_Pk		0.0000	0.0000	20.6897	0.0083

Fuente: Elaboración propia.

Tabla 13. Pruebas de Backtesting (AZTECA),
período noviembre 2014 a marzo 2017

	$1 - \alpha$	LRuc (p-val)	LRcc (p-val)	AE	Avg_QL
MC		0.0001	0.0003	5.5172	0.0211
SHa	95%	0.6853	0.8877	0.6897	0.0106
LSTM_YZ		0.0000	0.0000	8.2759	0.0339
LSTM_Pk		0.0000	0.0000	8.2759	0.0368
MC		0.0000	0.0000	17.2414	0.0077
SHa	99%	0.4452	0.7472	0.0000	0.0026
LSTM_YZ		0.0000	0.0000	34.4828	0.0211
LSTM_Pk		0.0000	0.0000	41.3793	0.0252

Fuente: Elaboración propia.

De una manera más general, para verificar cuál de los modelos se ajusta mejor a los datos se realizó una prueba de eficiencia considerando el número de excesos (veces en que las pérdidas observadas fueron superiores al VaR) que presentó cada modelo. En la tabla 14, se muestran las pérdidas reales que fueron superiores al VaR calculado por cada modelo. Podemos observar que el modelo SHa fue el que mejor desempeño tuvo para estimar el nivel de riesgo, seguido del MC. Los pronósticos de las redes neuronales fallaron como mejor modelo de pronóstico y como modelo de gestión de riesgos al tener más excesos de los permitidos, de acuerdo con los parámetros establecidos por Kupiec y la zona verde establecida por Basilea.

Tabla 14. Número de excesos,
período noviembre 2014 a marzo 2017

	$1 - \alpha$	MC	SHa	LSTM_YZ	LSTM_Pk
AMXL	95%	3	5	13	14
AZTECA		11	3	21	21
AMXL	99%	1	0	8	10
AZTECA		7	0	16	21

Fuente: Elaboración propia.

El nivel de eficiencia de los modelos LSTM es nulo al no poder considerar cubierto el nivel de confianza deseado, requiriendo ajustar o modificar estos modelos. El máximo nivel de eficiencia alcanzado fue de un 72%, obtenido por el modelo LSTM_YZ con un α de 0.01 para la serie de AMXL. El mejor desempeño que tuvieron los modelos LSTM_YZ y LSTM_Pk fue en el caso de AMXL, situación que corresponde al nivel de pronóstico obtenido por la red neuronal, denotando que el modelo LSTM_Pk presenta los peores resultados en el caso de AZTECA al tampoco presentar una disminución en el número de excesos al aumentar el nivel de confianza.

El modelo MC tampoco resultó ser un buen modelo de estimación del riesgo, ya que su nivel de eficiencia es menor para ambos niveles de confianza, teniendo su peor nivel (62%) para un α de 0.05 en el caso de AZTECA, y su mejor nivel (96.5%) para un α de 0.01 en el caso AMXL. Debido a lo anterior es necesario calibrarlo (i.e., aumentar el número de escenarios o cambiar la dinámica de la caminata aleatoria) a modo de poder alcanzar el nivel de confianza deseado.

En lo referente al modelo de SHa, el nivel de eficiencia obtenido considerando un α de 0.01 corresponde con el nivel de confianza deseado, por lo que se puede concluir que es un buen modelo de estimación en el caso de movimientos extremos. Sin embargo, para un α de 0.05 su nivel de eficiencia fue menor al 90% y requiere un ajuste en las variables utilizadas (e.g., cambiar el valor de lambda o modificar el horizonte de análisis histórico).

Es importante señalar que los pronósticos generados por los modelos tuvieron meses donde el error fue bastante amplio respecto de las pérdidas/ganancias realizadas, esto puede repercutir en la cantidad de requerimiento de capital que

deben tener los bancos en reserva, al igual que en cambios innecesarios en las posiciones de los portafolios pertenecientes a inversionistas aversos al riesgo, por lo que no serían los modelos ideales a elegir; excepto por el modelo SHa en AZTECA con un nivel de confianza de 95%, donde el pronóstico capturó en buena medida el dinamismo de la serie de Ganancias y Pérdidas (PL) (ver Anexos 3 y 4).

Conclusiones

Los modelos de redes neuronales pueden parecer modelos de pronóstico superiores dada la complejidad añadida, así como la información adicional que pueden manejar en contraste con otros modelos. Si bien, no cabe duda de que pueden presentar una mejora en cuanto a determinados problemas de estimación, no tuvieron el mismo nivel de desempeño (observado en la literatura revisada para el presente trabajo) al manejar un horizonte de pronóstico más amplio (i.e., mensual), lo que conllevó a subestimar el nivel del riesgo financiero al que se está expuesto para una acción, y siendo superados por otros modelos más simples como la Simulación Histórica con alisado y la Simulación Monte Carlo. Cabe resaltar el alto nivel de eficiencia que presentó la Simulación con alisado, ya que, aún sin buscar la calibración óptima del modelo, fue capaz de detectar el dinamismo de ambas acciones, teniendo su mejor desempeño para el caso de AZTECA con un nivel de confianza del 95%.

El error en los pronósticos generados por las redes LSTM fue más marcado en los momentos donde las noticias, tanto de contexto macroeconómico como específico de la compañía, jugaron un papel relevante en el movimiento de las acciones. Aunque, a pesar de la magnitud del error presentado, y dado el elevado nivel de incertidumbre asumido (al tratar de hacer pronósticos mensuales), las redes neuronales encontraron una función capaz de detectar cierta tendencia en los datos, así como el poder anticipar algunos de los movimientos en los picos y caídas de volatilidad, salvo aquellos donde la volatilidad fue generada por eventos que no tenían relación directa con la información fundamental-cuantitativa de la empresa. Otra situación a mencionar es que las redes neuronales tuvieron una mejor estimación de los datos futuros para el caso de la acción con alta bursatilidad respecto de la de

media bursatilidad, aunque no se puede considerar concluyente dado el comportamiento de las variables y que sólo se analizaron dos acciones de todas las presentes en el mercado bursátil.

A su vez, y de conformidad con la literatura revisada, los modelos de volatilidad que mejores estimados generaron fueron los de rango, aún para el caso de AMXL donde la desviación estándar mostró tener un mejor nivel de eficiencia en el análisis de la volatilidad histórica. Los modelos de rango fueron superiores tanto en las pruebas del coeficiente U como en la del RMSE, teniendo los mejores resultados con los modelos Yang-Zhang y Parkinson.

En cuanto a los modelos de estimación del VaR, la Simulación Histórica con alisado resultó ser el mejor modelo para el pronóstico del riesgo financiero de una acción, ya que tiene un nivel aceptable de errores. Aunque, su nivel de eficiencia va ligado a la disponibilidad de una serie histórica lo bastante amplia como para tomar en cuenta las fechas donde se incurrió en las mayores pérdidas registradas de un activo, pero considerando que para poder capturar mejor el dinamismo de la serie de tiempo, la ventana de análisis no debe ser tan amplia al generar la simulación, a modo de eliminar los eventos atípicos o eventos extremos con mucho tiempo de antigüedad que estarían generando una mayor desviación, y así, el modelo pueda adaptarse con mayor rapidez a los cambios estructurales del mercado financiero, principalmente en los análisis que consideren un nivel de confianza del 99%.

La Simulación Monte Carlo, presentó buenos resultados sólo para el activo AMXL, dado que el activo AZTECA, en la serie de pérdidas y ganancias, tuvo movimientos más bruscos lo que demostró la principal deficiencia de este modelo: supone la normalidad en la distribución de los datos. Puesto que

usualmente las series de rendimientos de las acciones no presentan distribuciones normales y, por el contrario, tienen distribuciones de colas pesadas, alto skew, y/o valores extremos, se tiene riesgo de modelo y, dado los resultados obtenidos con AZTECA, el VaR subestimar  el nivel de riesgo.

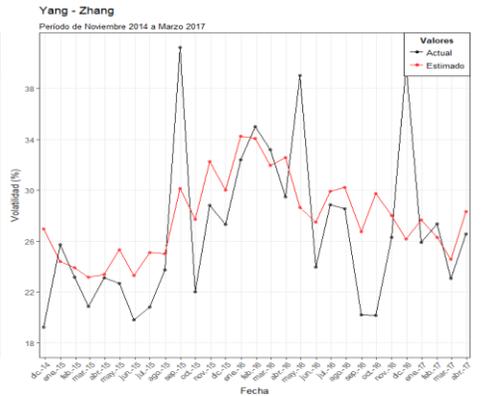
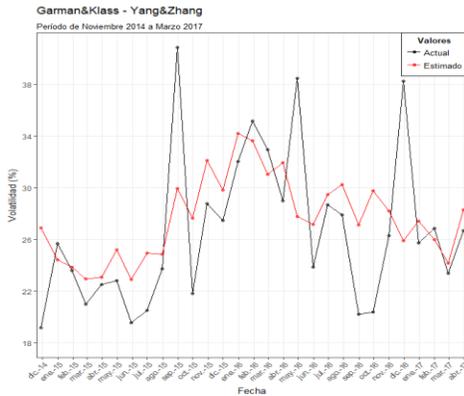
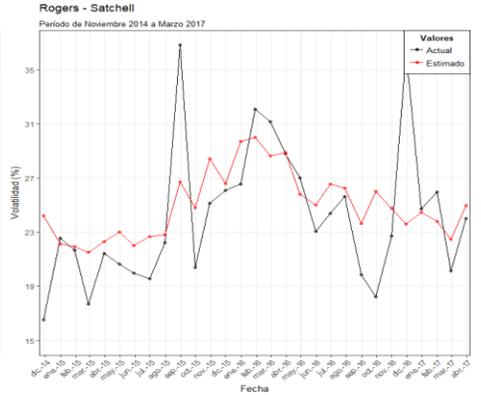
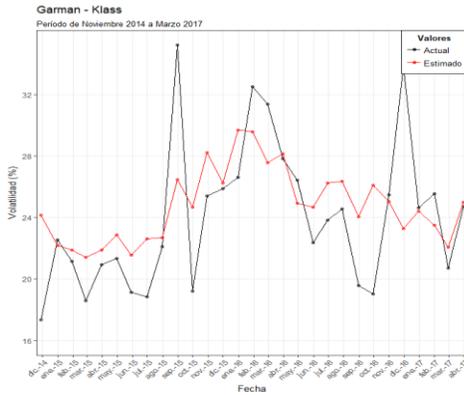
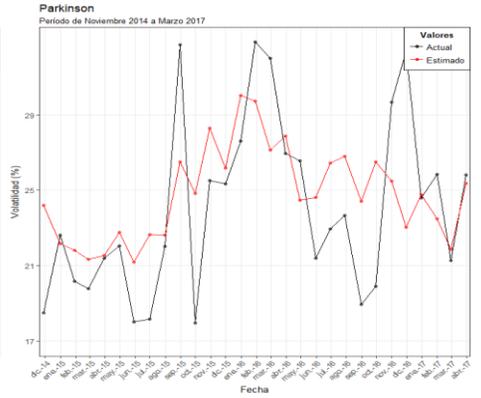
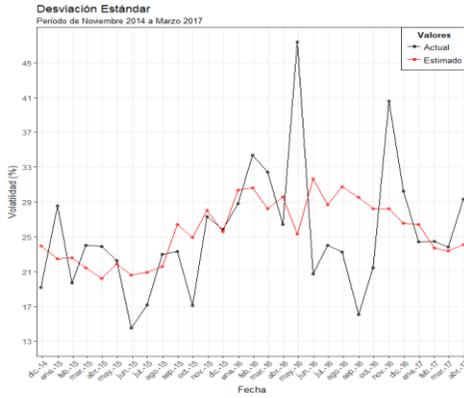
La hip tesis general planteada se rechaza dado el pobre desempe o de los modelos de redes neuronales en el c culo del VaR, sin embargo, se comprob  la hip tesis espec fica de que los modelos de rango contribuyen a la obtenci n de un mejor nivel de pron stico.

Futuras investigaciones deber an estar encaminadas a la evaluaci n de estos modelos con una serie hist rica m s amplia y con una frecuencia intermedia entre la diaria y la mensual (i.e., semanal). Ya con una menor frecuencia, se podr a considerar incluir otras variables independientes al modelo de redes para tratar de capturar los eventos at picos y/o extremos de los mercados, tales como el sentimiento del mercado y el sentimiento de las noticias, donde los estudios realizados por Chen y Lazer (2011), Goel y Mittal (2011), Corea (2016), entre otros, corroboraron los resultados positivos de Bollen, Mao y Zeng (2010) en el uso de un sentimiento del inversionista construido a base de la informaci n de la red social Twitter, para predecir los cambios diarios de los valores de cierre del Dow Jones Industrial Average (DJIA), as  como el dise o propio de Bloomberg (2016), que hizo uso de las noticias de su plataforma y la informaci n del sentimiento social (creado a partir de Twitter), con el fin de aplicar dicho  ndice a estrategias de inversi n y al pron stico de los precios y rendimientos de los instrumentos accionarios.

Tambi n se deber a evaluar el desempe o de estos modelos en diferentes escenarios, i.e., crisis financieras, momentos de baja volatilidad y momentos

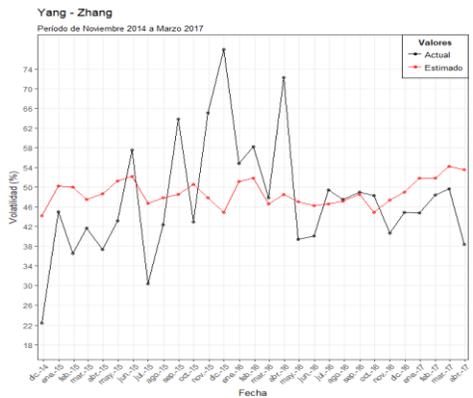
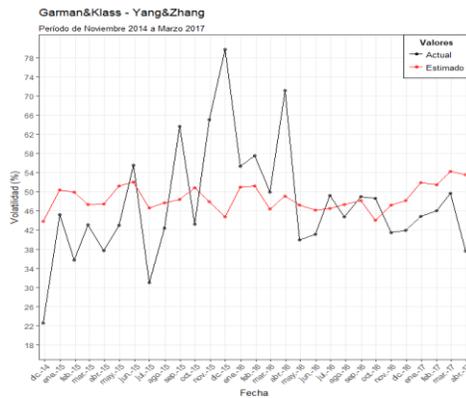
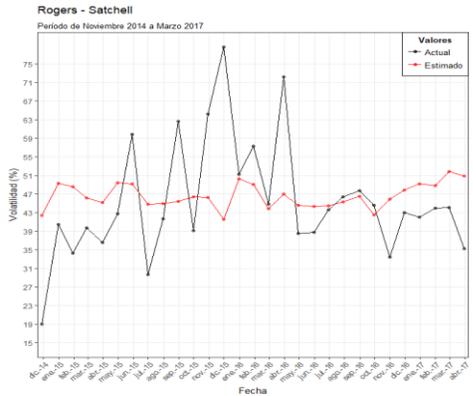
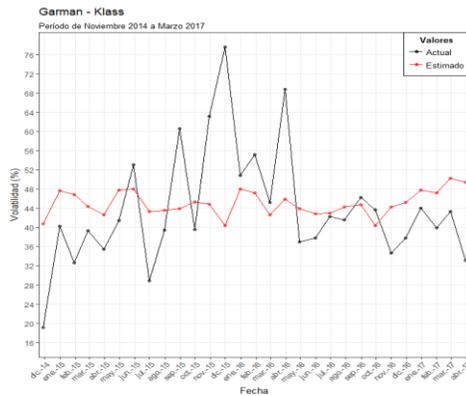
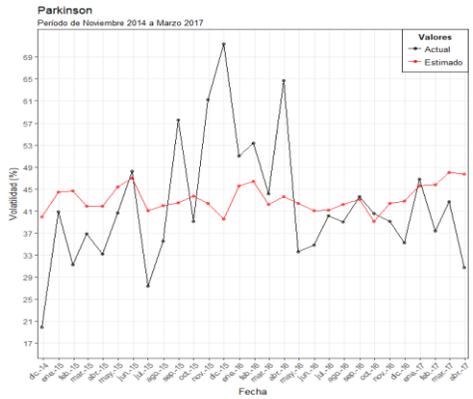
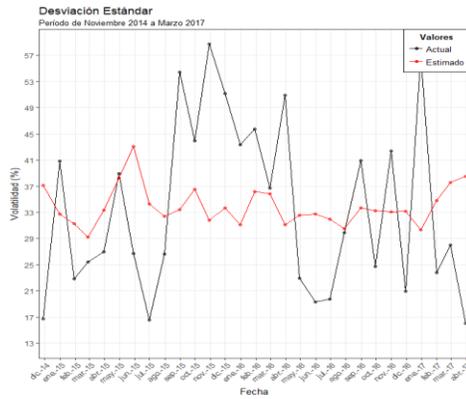
de una elevada volatilidad por escenarios atípicos (negociaciones políticas, opiniones de figuras públicas y/o líderes de la industria, anuncio de inicio o cancelación de proyectos específicos de la empresa, entre otros).

Anexo 1. Pronósticos de Volatilidad un mes adelante (AMXL)



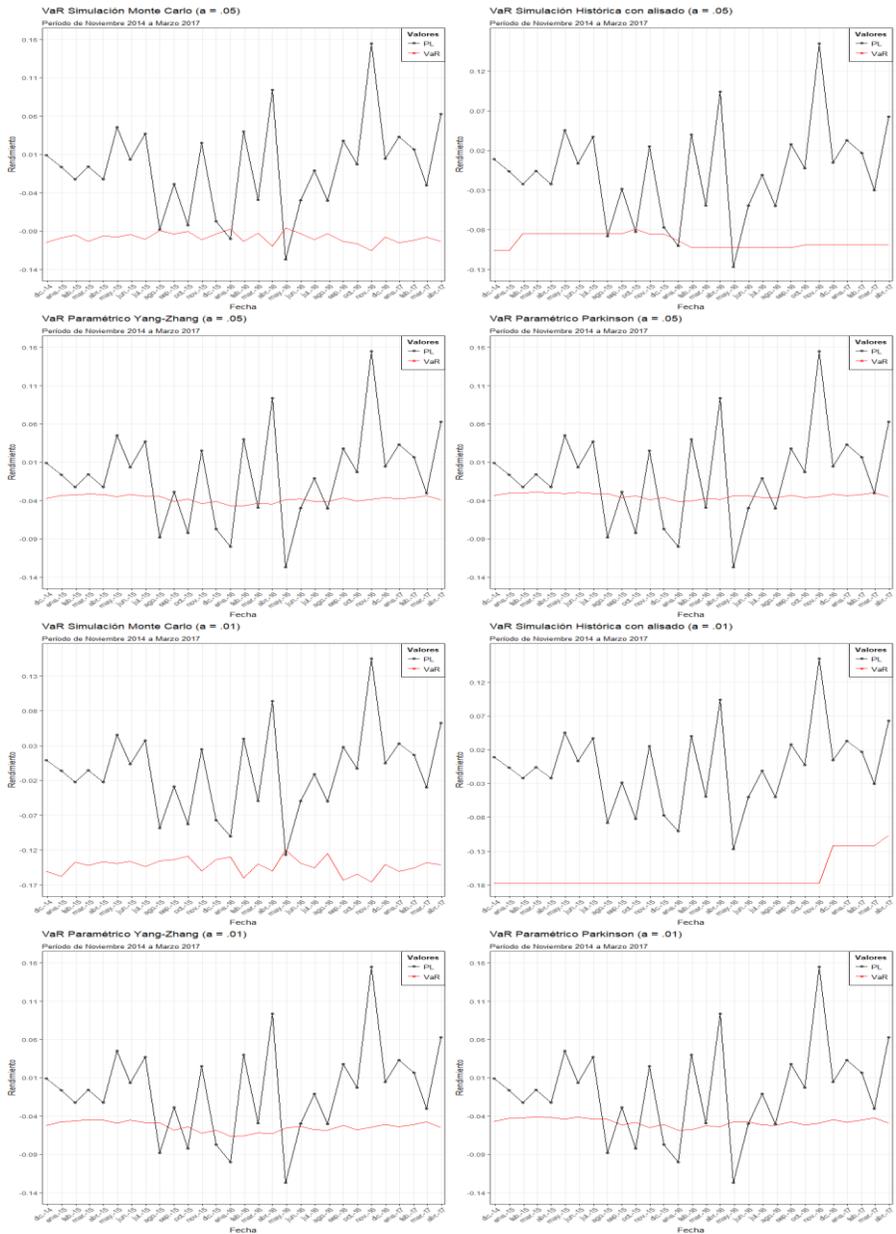
Fuente: Elaboración propia.

Anexo 2. Pronósticos de Volatilidad un mes adelante (AZTECA)



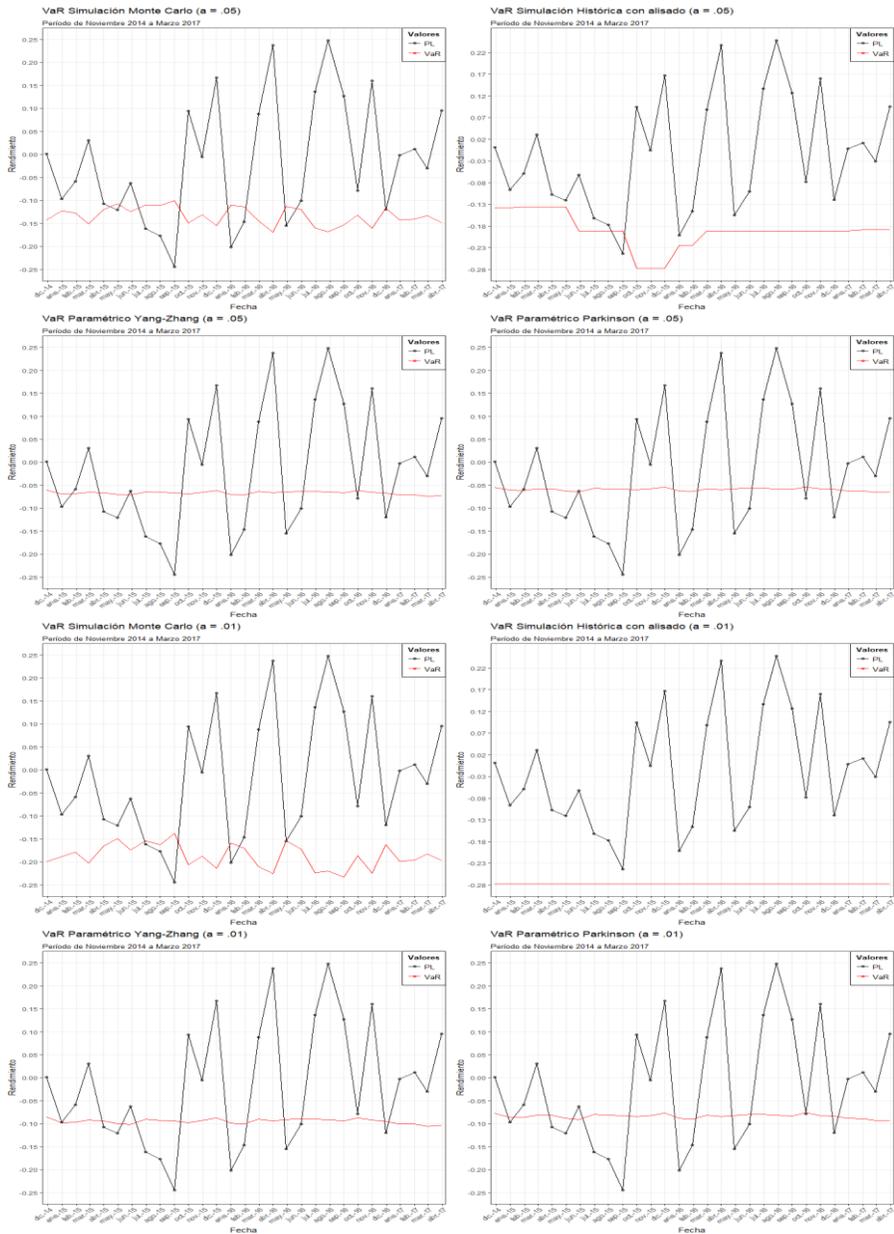
Fuente: Elaboración propia.

Anexo 3. Pronósticos de VaR un mes adelante (AMXL)



Fuente: Elaboración propia.

Anexo 4. Pronósticos de VaR un mes adelante (AZTECA)



Fuente: Elaboración propia.

Anexo 5. Rutina red neuronal LSTM en R (2018)

```
##### LSTM Model #####  
  
#Libraries  
library(xts)  
library(DMwR)  
library(keras)  
install_keras()  
  
# Parameters -----  
iter <- 100  
lunit <- 100 # 100 for AMXL and 15 for AZTECA  
tsteps <- 116  
vsplit <- .1  
nv <- 1  
  
x <- lag.xts(x, 1)  
x <- x[-(1:2)] # Ratio training/testing must be an entire number  
x_t <- x_t[-30,]  
y <- vol_tr_norm[-(1:2)]  
  
# LSTM Keras (2018) -----  
  
# Data defined as (samples/length, timesteps/lag, features/variables) for inputs, for  
outputs is defined as (samples/length, timesteps/lags)  
x <- array(x, dim = c(nrow(x), tsteps, ncol(x)))  
x_t <- array(x_t, dim = c(nrow(x_t), tsteps, ncol(x_t)))  
y <- array(y, dim = c(nrow(y)))
```

```

# LSTM Model

use_session_with_seed(4)

model <- keras_model_sequential() %>%

  layer_lstm(units = lunit,
             input_shape = c(tsteps, 10), # Input defined as (timesteps, features)
             use_bias = T, return_sequences = F) %>%

  layer_dense(units = 1) %>%

  compile(optimizer = 'sgd', loss = 'mean_squared_error')

history <- model %>% fit(x, y, epochs = iter,
                       validation_split = vsplit,
                       verbose = 1, view_metrics = F)

plot(history)

# Forecasting -----
# Volatility Forecast

preds <- predict(model, x_t)

preds <- unscale(preds, vol_t_norm)

```

Referencias

- Alatorre, J. (2015). Trading volatility in the real world. *Risk Management and Trading Conference 2015*. México: RiskMathics.
- Allaire, J., & Chollet, F. (2018). *keras: R Interface to 'Keras'*. R package version 2.1.6. Recuperado de <https://CRAN.R-project.org/package=keras>
- Arbib, M. (2003). *The Handbook of Brain Theory and Neural Networks*. Wisconsin: The MIT Press.
- Ardia, D., Boudt, K., & Catania, L. (2016, Noviembre 18). *Value-at-Risk Prediction in R with the GAS Package*. Recuperado de Cornell University Library: <https://arxiv.org/abs/1611.06010>
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166. doi:10.1109/72.279181
- Bennčík, D. (2014). Range-based Volatility Estimation and Forecasting. *IES Working Paper*, 34. Recuperado de <http://ies.fsv.cuni.cz>
- Bennet, C., & Gil, M. (2012). *Measuring historical volatility*. Madrid: Santander Investment Bolsa, Sociedad de Valores, S.A. Recuperado de https://dynamiproject.files.wordpress.com/2016/01/measuring_historic_volatility.pdf
- Bliemel, F., & MacKay, D. (1973). Theil's Forecast Accuracy Coefficient: A Clarification Spatial Measurement of Retail Store Demand. *Journal of Marketing Research*, X, 444-447.
- Bollen, J., Mao, H., & Zeng, X. (2010). Twitter mood predicts the stock market. *Journal of Computational Science*., 2(1), 1-8. doi:10.1016/j.jocs.2010.12.007
- Boz, G., Menéndez, C., Orgaz, N., & Pior, D. (2015). ¿Influyen las Normas Internacionales de Información Financiera en el riesgo de las acciones? *Contaduría y Administración*, 60(3), 556-577. doi:10.1016/j.cya.2015.05.010

-
- Carmona, E. (2013). Ajuste a la calificación del riesgo de mercado de las acciones más volátiles que conforman el Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores, con la implementación de una Red Neuronal Artificial clasificadora. *Revista Mexicana de Economía y Finanzas*, 8(1), 25-51. Recuperado de <http://www.redalyc.org/articulo.oa?id=423739493003>
- CFA Institute. (2015). *CFA® Program Curriculum 2016 level I* (Vol. 5). USA: Wiley.
- CFA Institute. (2016). *CFA® Program Curriculum 2017 level II* (Vol. 1). USA: Wiley.
- CFA Institute. (2016). *CFA® Program Curriculum 2017 level II* (Vol. 6). USA: Wiley.
- CFA Institute. (2016). *CFA® Program Curriculum 2017 level II* (Vol. 4). USA: Wiley.
- Chaudhuri, T., & Ghosh, I. (2015). Forecasting Volatility in Indian Stock Market using Artificial Neural Networks with Multiple Inputs and Outputs. *International Journal of Computer Applications*, 120(8). doi:10.5120/21245-4034
- Chen, N., Roll, R., & Ross, S. (1986). Economic Forces and the Stock Market. *The Journal of Business*, 59(3), 383-403. Recuperado de <http://www.jstor.org/stable/2352710>
- Chen, R., & Lazer, M. (2011). Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement. Recuperado de <http://cs229.stanford.edu/proj2011/>
- Chen, X., Lai, K., & Yen, J. (2009). A statistical neural network approach for value-at-risk analysis. *2009 International Joint Conference on Computational Sciences and Optimization*, 17-21. doi:10.1109/CSO.2009.350
- Chicago Board Options Exchange. (n.d.). The CBOE Volatility Index-VIX. *White paper*. USA. Recuperado de http://www.cboe.com/framed/pdf/framed?content=/micro/vix/vixwhite.pdf§ion=SEC_OPTIONS_PRODUCTS&title=VIX+White+Paper
-

-
- Chou, R. (2005). Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range (CARR) Model. *Journal of Money, Credit and Banking*, 37, 561-582. Recuperado de <http://www.jstor.org/stable/3839168>
- Chou, R., Chou, H., & Liu, N. (2009). Range Volatility Models and Their Applications in Finance. In C. Lee, & A. Lee, *Handbook of Quantitative Finance and Risk Management* (pp. 1273-1281). Boston, MA: Springer. doi:10.1007/978-0-387-77117-5_83
- Ciaburro, G., & Venkateswaran, B. (2017). *Neural Networks with R*. Birmingham, UK: Packt Publishing.
- CINIF. (2015). *NIF A-3 Necesidades de los Usuarios y Objetivos de los Estados Financieros*. México: IMCP (Instituto Mexicano de Contadores Públicos). Recuperado de http://fcaenlinea1.unam.mx/anexos/1165/1165_u2_a3.pdf
- Connor, G. (1995). The three types of Factor Models: A comparison of their explanatory power. *Financial Analysts Journal*, 51(3), 42-46. Recuperado de <http://www.jstor.org/stable/4479845>
- Corea, F. (2016). Can Twitter Proxy the Investors' Sentiment? The Case for the Technology Sector. *Big Data Research*, 4, 70-74. doi:10.1016/j.bdr.2016.05.001
- Cui, X., Lam, D., Verma, A., & Bloomberg. (2016). Embedded value in Bloomberg news & social sentiment data. Recuperado de <http://b.bloomberg.com/reg-EDFSentimentWhitepaper?src=blog>
- De Lara, A. (2013). *Medición y control de riesgos financieros*. México: Limusa.
- Disposiciones de Carácter General Aplicables a las Instituciones de Crédito (Circular Única de Bancos). Publicado en el Diario Oficial de la Federación, México, 2 de diciembre de 2005, última reforma 25 de octubre de 2017, por la Comisión Nacional Bancaria y de Valores.
-

-
- Donkin, T. C. (1864). *An Etymological Dictionary of the Romance Languages: Chiefly From the German of Friedrich Diez*. London: Williams and Norgate.
- dos Santos Maciel, L., & Ballini, R. (2017). Value-at-risk modeling and forecasting with range-based volatility models: empirical evidence. *Revista Contabilidade & Finanças*, 28(75), 361-376. doi:10.1590/1808-057x201704140
- Du, K., & Swamy, M. (2014). *Neural Networks and Statistical Learning*. London: Springer-Verlag. doi:10.1007/978-1-4471-5571-3_2
- Fama, E. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 21(5), 55-59. doi:10.2469/faj.v51.n1.1861
- Fama, E. (1970). Efficient Capital Markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. doi:10.2307/2325486
- Fine, T. (1999). *Feedforward Neural Network Methodology*. Nueva York: Springer-Verlag.
- García, C. (2016). *Impacto de los factores macroeconómicos en el riesgo sistemático del mercado bursatil mexicano (2010-2015)*. México: Universidad Nacional Autónoma de México. Recuperado de <http://tesis.unam.mx/F>
- Garman, M., & Klass, M. (1980). On the Estimation of Security Price Volatilities from Historical Data. *The Journal of Business*, 53, 67-78. Recuperado de <http://www.jstor.org/stable/2352358>
- GARP. (2016). *Financial Risk Manager (FRM) Exam Part I: Quantitative analysis*. New York: Pearson.
- GARP. (2016). *Financial Risk Manager (FRM) Exam Part I: Valuation and Risk Models*. New York: Pearson.
- GARP. (2016). *Financial Risk Manager (FRM) Exam Part II: Market Risk Measurement and Management*. Nueva York: Pearson.
- Goel, A., & Mittal, A. (2011). Stock prediction using Twitter Sentiment Analysis. Recuperado de <http://cs229.stanford.edu/proj2011/>
-

-
-
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. USA: The MIT Press.
- Greff, K., Srivastava, R., Koutnik, J., Steunebrink, B., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232. doi:10.1109/TNNLS.2016.2582924
- Grupo Banco Bilbao Vizcaya Argentaria. (2017). *Informe con Relevancia Prudencial 2016 - Pilar III de Basilea*. España: BBVA. Recuperado de <https://accionistaseinversores.bbva.com/informacion-financiera/informes-financieros/#2016>
- Gujarati, D. (2010). *Econometría*. México: McGraw-Hill.
- Hagan, M., Demuth, H., Beale, M., & De Jesús, O. (2014). *Neural Networks Design*. USA: Martin Hagan.
- Hall, R. (2001). Struggling to understand the stock market. *American Economic Review*, 91(2), 1-11. doi:10.1257/aer.91.2.1
- Hamid, S., & Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research*, 57(10), 1116-1125. doi:10.1016/S0148-2963(03)00043-2
- Haykin, S. (2009). *Neural Networks and Learning Machines*. Nueva Jersey: Pearson.
- Hellström, T., & Holmström, K. (1998). *Predicting the Stock Market*. Center of Mathematical Modeling, Mathematics and Physics. Suecia: Mälardalen University. doi:10.1.1.12.9843
- Hinton, G. (2012). The Perceptron learning procedure. *Neural Networks for Machine Learning*. Recuperado de <https://www.coursera.org/learn/neural-networks/home/>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. doi:10.1162/neco.1997.9.8.1735
- Hull, J. (2012). *Risk management and financial institutions*. New Jersey: Wiley.
- Jorion, P. (2004). *Valor en Riesgo*. México: Limusa.
-
-

-
- Jorion, P. (2007). *Value at Risk: The new benchmark for managing financial risk*. USA: McGraw-Hill.
- Jorion, P. (2009). *Financial risk manager handbook*. USA: Wiley.
- Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). An Empirical Exploration of Recurrent Neural Network Architectures. *32nd International Conference on Machine Learning*, 37, pp. 2342-2350. Lille: Journal of Machine Learning Research.
- JP Morgan & Co. (1996). *RiskMetrics*. New York: Reuters.
- Karpathy, A. (2017). Redes Neuronales. *CS231n: Convolutional Neural Networks for Visual Recognition*. USA. Recuperado de <http://cs231n.github.io/>
- Kim, H., & Won, C. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems With Applications*, 103, 25-37. doi:10.1016/j.eswa.2018.03.002
- Konar, A. (2005). *Computational Intelligence. Principles, techniques and application*. Berlín: Springer-Verlag.
- Kriesel, D. (2007). *A brief introduction to Neural Networks*. Recuperado de <http://www.dkriesel.com>
- Locarek-Junge, H., & Prinzler, R. (1998). Estimating value-at-risk using neural networks. *Informationssysteme in der Finanzwirtschaft*, 385-397. doi:2443/10.1007/978-3-642-60327-3_28
- Martín, B., & Sanz, A. (2007). *Redes Neuronales y Sistemas Borrosos*. México: Alfaomega.
- Miazhynskaia, T., Dorffner, G., & Dockner, E. (2003). Risk Management Application of the Recurrent Mixture Density Network Models. *Artificial Neural Networks and Neural Information Processing*, 589-596. doi:10.1007/3-540-44989-2_70
- Mitchell, T. (1997). *Machine Learning*. Nueva York: McGraw-Hill & MIT Press.
-

-
- Nilsson, N. (1998). *Introduction to Machine Learning*. Recuperado de <http://robotics.stanford.edu/people/nilsson/mlbook.html>
- Olah, C. (2015, Agosto 27). *Understanding LSTM Networks*. Retrieved Diciembre 3, 2017, from Colah's blog: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Parkinson, M. (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. *The Journal of Business*, 53, 61-65. Recuperado de <http://www.jstor.org/stable/2352357>
- Petneházi, G., & Gáll, J. (2018, Marzo 19). *Exploring the predictability of range-based volatility estimators using RNNs*. Recuperado de Cornell University Library: <https://arxiv.org/abs/1803.07152v1>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Recuperado de <https://www.R-project.org/>
- Real Academia Española. (2014). *Diccionario de la lengua española* (23a ed.). Recuperado de <http://dle.rae.es/?id=WT8tAMI>
- Reilly, F., & Brown, K. (2011). *Investment analysis & portfolio management*. USA: Cengage.
- Reyes, J. (2015). *Administración de Riesgos: Una aplicación del Valor en Riesgo Alisado en el Mercado Cambiario 2005 - 2010*. México: Universidad Nacional Autónoma de México. Recuperado de <http://tesis.unam.mx/F>
- RiskMetrics Group. (1999). *Risk management: A practical guide*. USA: RMG.
- Rogers, L., & S., S. (1991). Estimating Variance From High, Low and Closing Prices. *The Annals of Applied Probability*, 1, 504-512. Recuperado de <http://www.jstor.org/stable/2959703>
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536. doi:10.1038/323533a0
- Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A modern approach*. Nueva Jersey: Prentice Hall.
-

-
- Shu, J., & Zhang, J. (2006). Testing Range Estimators of Historical Volatility. *The Journal of Futures Markets*, 26(3), 297-313. doi:10.1002/fut.20197
- Siddique, N., & Adeli, H. (2013). *Computational intelligence: Synergies of fuzzy logic, neural networks, and evolutionary computing*. West Sussex: Wiley.
- Silva, M. (2017). *Propuesta metodológica para el cálculo de la volatilidad implícita y pronóstico mediante modelo de redes neuronales aplicado a las opciones en el MexDer: América Móvil (AMX), México 2014*. México: Universidad Nacional Autónoma de México. Recuperado de <http://tesis.unam.mx/F>
- Slepaczuk, R., & Zakrzewski, G. (2009). High-Frequency and Model-Free Volatility Estimator. *Working Papers*, 13(23). doi:10.2139/ssrn.2508648
- Sumathi, S., & Paneerselvam, S. (2010). *Computational Intelligence paradigms: Theory and applications using Matlab*. Florida: CRC Press.
- The Volatility Institute NYU. (2013). GARCH model. USA. Recuperado de <https://vlab.stern.nyu.edu/doc/2?topic=mdls>
- Tino, P., Schittenkopf, C., & Dorffner, G. (2001). Financial Volatility Trading Using Recurrent Neural. *IEEE Transactions on Neural Networks*, 12(4), 865-874. doi:10.1109/72.935096
- Trejo, C., Noguera, M., & White, S. (2015). Financial ratios used by equity analysts in Mexico and stock returns. *Contaduría y Administración*, 60(3), 578-592. doi:10.1016/j.cya.2015.02.001
- Wallisch, P., Lusignan, M., Benayoun, M., Baker, T., Dickey, A., & Hatsopoulos, N. (2014). *Matlab for Neuroscientists*. Amsterdam: Elsevier & Academic Press.
- Yang, D., & Zhang, Q. (2000). Drift-Independent Volatility Estimation Based on High, Low, Open, and Close Prices. *The Journal of Business*, 73, 477-492. Recuperado de <http://www.jstor.org/stable/10.1086/209650>
-