



UNIVERSIDAD NACIONAL DE COLOMBIA

Revisión sistemática de literatura: Modelos de pronóstico índice Standard and Poor´s 500 (S&P500)

Héctor David Acevedo Arango

Universidad Nacional de Colombia
Facultad de Minas, Escuela Ingeniería de la Organización
Medellín, Colombia
2012

Revisión sistemática de literatura: Modelos de pronóstico índice Standard and Poor´s 500 (S&P500)

Héctor David Acevedo Arango

Trabajo final de investigación presentado como requisito parcial para optar al título de:
Magister en Ingeniería Administrativa

Director (a):
Dsc. Sergio Botero Botero

Línea de Investigación:
Econometría

Facultad de Minas
Universidad Nacional de Colombia
Medellín, Colombia
2012

The value of a college education is not the learning of many facts but the training of the mind to think.

Albert Einstein

Agradecimientos

Quiero agradecer especialmente a mi director de trabajo final el profesor Sergio Botero Botero por su permanente disposición, paciencia y confianza para enseñarme todo lo necesario para culminar este trabajo.

A mi madre que me acompañó incondicionalmente en este proceso, gracias a su ejemplo pude llevar de una manera adecuada los momentos más difíciles.

A los profesores Oscar Manco y Juan David Velásquez, por sus aportes valiosos que permitieron establecer las bases de este trabajo.

A los profesores de la asignatura Seminario de Investigación que cambiaron mi forma de ver la realidad.

A mis compañeros Néstor Raúl Ospina y Adriana Arango por su amistad y colaboración incondicional durante toda la maestría.

A mis estudiantes que me permitieron encontrar lo bello de la docencia, y adicionalmente su paciencia ante mis errores en el aprendizaje de esta.

A la Universidad Nacional de Colombia y su programa de becas del cual soy beneficiario.

Resumen

El Standard and Poor's 500 es el índice bursátil más estudiado en la literatura, ya que representa el sector industrial de Estados Unidos, el objetivo de este trabajo es hacer una Revisión Sistemática de Literatura de gran parte de los artículos que tengan como objetivo la predicción de este índice, que pretende recoger las técnicas, criterios de desempeño, pruebas de validación más utilizados con el fin de agruparlos, clasificarlos y proponer una metodología que permita un avance más rápido en esta área del conocimiento, de igual manera también se pretende establecer cuáles son los modelos de pronóstico que mejores resultados ofrecen. Se encuentra que gran parte de los criterios utilizados para medir modelos de pronóstico no son apropiados para series financieras, también se concluye que es bastante complejo la comparación entre varios autores.

Palabras clave: NYSE, mercados financieros, S&P500, pronóstico, predicción, revisión sistemática, medidas de desempeño, modelos de pronóstico.

Abstract

The Standard and Poor's 500 index is the most studied in the literature, this is because it represents the U.S. industry, the aim of this paper is to make a Systematic Literature Review of most items that have the objective prediction of this index, which aims to collect technical performance criteria, validation tests commonly used to group them, sort them and propose a methodology to faster progress in this area of knowledge, just as also seeks to establish what the forecast models that offer better results. We find that

many of the criteria used to measure forecast models are not appropriate for financial series, also concludes that it is quite complex comparison between various authors.

Keywords: Forecasting, S&P500, stock markets, financial markets, performance measures, prediction models, system review, surveying S&P500

Contenido

Pág.

Resumen	IX
Lista de tablas.....	XIV
Introducción.....	17
1. Estructuración de modelos de pronóstico.....	21
1.1 Introducción	21
1.2 Clasificación de los modelos de pronóstico	22
1.2.1 Modelos de pronóstico según variables explicativas:	23
1.2.2 Modelos de pronóstico según metodología.....	23
1.3 Medición del desempeño en modelos de pronóstico	24
1.4 Propuesta revisión sistemática de literatura del índice S&P 500.....	25
1.4.1 Índice Standard and Poor's 500	26
1.4.2 Revisión sistemática de literatura.....	27
1.5 Protocolo revisión sistemática de literatura.....	29
1.6 Objetivos de la investigación	30
1.6.1 Objetivo general	30
1.6.2 Objetivos específicos.....	30
2. Medición en modelos de pronóstico	32
2.1 Medidas de desempeño asociadas a la precisión.....	32
2.1.1 SSE Suma de los cuadrados del error	33
2.1.2 AIC Criterio de información de Akaike.....	33
2.1.3 AICC Criterio de información de Akaike modificado	33
2.1.4 BIC Criterio de información bayesiano	34
2.1.5 CD Corrección tendencia bajista.....	34
2.1.6 CU Corrección tendencia alcista.....	34
2.1.7 DA Precisión de la dirección	35
2.1.8 DM Diebold and Mariano	35
2.1.9 MDM Diebold y Mariano modificada	36
2.1.10 ES Pérdida esperada.....	36
2.1.11 Hit ratio.....	36
2.1.12 HMSE MSE heterocedástico.....	36
2.1.13 MAE Error medio absoluto.....	37
2.1.14 Max drawdown	37
2.1.15 MAPE Error absoluto porcentual medio	37
2.1.16 ME Media del error	37
2.1.17 MME(O) Media del error mixto.....	38

2.1.18	MME(U)	38
2.1.19	MPE Error porcentual medio	38
2.1.20	MSE Media cuadrada del error	38
2.1.21	NMSE Error cuadrático medio normalizado	39
2.1.22	R ² Coeficiente de determinación	39
2.1.23	Rentabilidad	39
2.1.24	RMSE Error cuadrático medio	40
2.1.25	Sharpe ratio	40
2.1.26	Sterling ratio	40
2.1.27	Theil U	40
2.1.28	VaR Valor en riesgo	41
2.2	Pruebas de validación de datos	41
2.2.1	Box Pierce.....	41
2.2.2	Chi cuadrado.....	41
2.2.3	Durbin Watson	42
2.2.4	Fisher Snedecor.....	42
2.2.5	Jarque Bera	43
2.2.6	Ljung box	43
2.2.7	LM Multiplicador de Lagrange	43
2.2.8	Q like.....	43
2.2.9	t-student.....	44
2.3	Selección de medidas de desempeño para modelos de pronóstico de índice bursátil	44
3.	Extracción de información RSL.....	49
3.1	An investigation of model selection criteria for neural network time series forecasting (Qi & Zhang, 2001).....	49
3.2	Financial Forecasting Using Support Vector Machines (SVM) (Cao & Tay, 2001)	51
3.3	Learning to trade via direct reinforcement (Moody & Saffell, 2001).....	52
3.4	Applying rough sets to market timing decisions (Shen & Loh, 2004)	53
3.5	The adaptive selection of financial and economic variables for use with artificial neural networks (Thawornwong & Enke, 2004).....	54
3.6	Stochastic volatility forecasting and risk management (Sadorsky, 2005).....	56
3.7	Trading Rules from Forecasting the Collapse of Speculative Bubbles for the S&P 500 Composite Index (Brooks & Katsaris, 2005).....	58
3.8	SV mixture models with application to S&P 500 index returns (Durham, 2007).....	59
3.9	Forecasting stock index volatility: comparing implied volatility and the intraday high-low price range (Corrado & Truong, 2007).....	60
3.10	Does implied volatility provide any information beyond that captured in model-based volatility forecasts? (Becker et al., 2007).....	61
3.11	Forecasting volatility by means of threshold models (Acosta, Muñoz, & Márquez, 2007)	62
3.12	Are combination forecasts of S&P 500 volatility statistically superior? (Becker & Clements, 2008).....	63
3.13	A hybrid derivative trading system based on volatility and return forecasting (Enke & Amornwattana, 2008).....	65
3.14	Deterministic regression model and visual basic code for optimal forecasting of financial time series (Balbás, Balbás, Galperin, & Galperin, 2008).....	66

3.15	Intra-daily information of range-based volatility for MEM-GARCH (Lam & Ng, 2009)	66
3.16	The information contents of VIX index and range-based volatility on volatility forecasting performance of S&P 500 (Hung et al., 2009)	67
3.17	Evolving least squares support vector machines for stock market trend mining (Yu et al., 2009)	68
3.18	Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic (Pooter, 2009)	69
3.19	Development and performance evaluation of FLANN based model for forecasting of stock markets(Majhi, Panda, & Sahoo, 2009)	71
3.20	Non linear prediction of the Standard & Poor’s 500 and the hang seng index under a dynamic increasing simple (Kumar, 2009)	72
3.21	Predicting the bear stock market: Macroeconomic variables as leading indicators (S. S. Chen, 2009)	73
3.22	Nonlinear dynamics in arbitrage of the S&P 500 index and futures: A threshold error-correction model (B. H. Kim, Chun, & Min, 2010)	74
3.23	Forecast accuracy and economic gains from bayesian model averaging using time-varying weights (Hoogerheide, Kleijn, Ravazzolo, Dijk, & Verbeek, 2010)	75
3.24	Predicting the equity premium with dividend ratios: Reconciling the evidence (Kellard, Nankervis, & Papadimitriou, 2010)	76
3.25	Modeling intraday volatility: A new consideration (Chu & Lam, 2011)	77
3.26	S&P500 Index Direction Forecasting from 1976 to 2010 : A Fuzzy System Approach(Duarte & Brazil, 2011)	78
3.27	Modelling and forecasting noisy realized volatility (Asai, McAleer, & Medeiros, 2012)	79
3.28	A hybrid modeling approach for forecasting the volatility of S&P 500 index return (Hajizadeh, Seifi, Fazel Zarandi, & Turksen, 2012)	81
3.29	Quantile forecasts of financial returns using realized garch models(Watanabe, 2012)	82
4.	Resultados y análisis RSL S&P500	84
4.1	Resumen de los criterios de desempeño	84
4.2	Resumen de modelos de pronóstico	87
4.3	Medición de los modelos	90
4.4	Artículos descartados en la RSL	95
5.	Metodología para formular un modelo de pronóstico	96
5.1	Objetivo del modelo de pronóstico	96
5.2	Metodología propuesta	96
6.	Conclusiones	100
	Bibliografía	105

Lista de tablas

Tabla 2-1. Escenario simulado para cálculo de diferentes medidas de desempeño.....	45
Tabla 2-2. Prueba real a criterios de desempeño más frecuentes	45
Tabla 2-3. Prueba a criterios de desempeño direccionales.....	46
Tabla 2-4. Clasificación y caracterización de criterios de desempeño en modelos de pronóstico aplicado a activos financieros.....	47
Tabla 3-1. Resultados (Qi & Zhang, 2001).....	50
Tabla 3-2. Resultados (Cao & Tay, 2001).....	51
Tabla 3-3. Resultados (Moody & Saffell, 2001).....	52
Tabla 3-4. Resultados (Shen & Loh, 2004).....	53
Tabla 3-5. Resultados (Thawornwong & Enke, 2004).....	54
Tabla 3-6. Resultados (Sadorsky, 2005).....	57
Tabla 3-7. Resultados (Brooks & Katsaris, 2005)	58
Tabla 3-8. Resultados (Corrado & Truong, 2007)	60
Tabla 3-9. Resultados (Becker et al., 2007).....	61
Tabla 3-10. Resultados (Acosta et al. 2007).....	63
Tabla 3-11. Resultados (Becker & Clements, 2008)	64
Tabla 3-12. Resultados (Enke & Amornwattana, 2008)	65
Tabla 3-13. Resultados (Hung et al., 2009)	68
Tabla 3-14. Resultados (Yu et al., 2009)	69
Tabla 3-15. Resultados (Pooter, 2009).....	70
Tabla 3-16. Resultados (Majhi et al. 2009)	71
Tabla 3-17. Resultados (Kumar, 2009).....	72
Tabla 3-18. Resultados (S. S. Chen, 2009)	74
Tabla 3-19. Resultados (Kim et al. 2010).....	75
Tabla 3-20. Resultados (Hoogerheide et al.2010).....	76
Tabla 3-21. Resultados (Kellard et al. 2010).....	77
Tabla 3-22. Resultados (Chu & Lam, 2011).....	78
Tabla 3-23. Resultados (Duarte & Brazil, 2011).....	79
Tabla 3-24. Resultados (Asai et al. 2012).....	80
Tabla 3-25. Resultados (Hajizadeh et al. 2012)	81
Tabla 3-26. Resultados 2 (Hajizadeh et al. 2012)	82
Tabla 3-27. Resultados (Watanabe, 2012)	83
Tabla 4-1. Criterios de desempeño agrupados por autor	84
Tabla 4-2. Pruebas de validación de datos reportadas en la revisión.....	86
Tabla 4-3. Modelos reportados clasificados por autor.....	87

Tabla 4-4.Comparación de modelos por rentabilidad	91
Tabla 4-5.Comparación de modelos por Hit ratio	93
Tabla 4-6.Comparación de modelos por Sharpe ratio	94
Tabla 4-7. Artículos descartados agrupados por autores	95
Tabla 5-1 Comparación MAE varios autores.....	98

Introducción

La econometría es la parte de las ciencias económicas que aplica técnicas matemáticas y estadísticas para la solución de problemas económicos mediante modelos (RAE,2012) ,esta se ha convertido en una herramienta empleada por los analistas financieros debido a que se han desarrollado varios modelos de pronóstico que arrojan buenas inferencias, lo anterior ha sido posible gracias a una creciente disponibilidad de series históricas de variables económicas y a las aplicaciones informáticas(Johnston & Dinardo, 1997).

El pronóstico de índices bursátiles es una necesidad para los inversionistas corporativos, analistas financieros, economistas, y demás agentes de los mercados financieros(Knight, Satchell, & Wang, 2003). El buen funcionamiento de las instituciones financieras es un punto fundamental en el desarrollo económico de un país, y este buen funcionamiento de la entidades depende en gran medida de su capacidad para pronosticar las condiciones del mercado en un horizonte de tiempo determinado(Bahrammirzaee, 2010). Teniendo en cuenta que los cambios en los precios de las acciones e índices manejan un alto grado de incertidumbre (Atsalakis & Valavanis, 2009), debido a que las series de tiempo son ruidosas, no estacionarias y determinísticamente caóticas (Cao & Tay, 2001), se hace necesario el estudio de modelos de pronóstico.

Atsalakis et al (2009) afirma que la medición del desempeño de los modelos de pronóstico se dividen en mediciones estadísticas (i.e. MAE, RMSE) y mediciones no estadísticas (i.e. retorno anual, tasa de aciertos), mientras que Qi y Zhang(2001) clasifican la medición a través de criterios estadísticos según su momento de medición en periodo entrenamiento y periodo de validación, finalmente en la revisión de bibliográfica se encuentra que cada autor maneja diferentes criterios para seleccionar la medida que evalúa el desempeño del modelo propuesto.

Sobre el pronóstico de índices bursátiles se han desarrollado varios estudios para las diferentes metodologías, bolsas de valores, intervalos de tiempo, y variables de entrada.

El índice bursátil mas estudiado es el Standard and Poor's 500 (S&P 500) (Thawornwong & Enke, 2004). Los índices desarrollados por la calificadora de riesgo Standard and Poors buscan crear alternativas de inversión para grandes fondos que requieren rentabilidades estables en el tiempo y que representen el mercado accionario de Estados Unidos, específicamente el S&P 500 está enfocado en las empresas de mayor capitalización bursátiles, es decir empresas líderes de sectores líderes (*S&P U.S. Indices Methodology*, 2012), de todo esto se desprende que el índice mencionado anteriormente sea el principal objeto de estudio de los modelos de pronóstico aplicados a series de tiempo financieras.

La pregunta de investigación es entonces, ¿Cuáles de los modelos de pronóstico desarrollados que han estudiado el índice S&P 500 arrojan las mejores predicciones?, sin embargo para resolver esta pregunta es necesario abordar cuestiones más específicas como: ¿Cuáles son los modelos de pronóstico que han estudiado el índice S&P 500?, ¿Cómo se mide el desempeño en los modelos de pronóstico?, ¿Cómo se mide la validez de datos en los modelos de pronóstico?

La metodología que se va a seguir para resolver la pregunta de investigación es la Revisión sistemática de literatura propuesta por Patticrew y Roberts (2006), se selecciona esta técnica porque se considera apropiada al establecer límites de manera muy explícita evitando resultados ambiguos, incompletos o exploraciones innecesarias.

El alcance de este trabajo se limita a los artículos cuyo objeto de estudio sea el pronóstico del índice S&P500 publicados en el periodo 2000-2012 reportados en la base de datos Scopus.

Como resultado de la RSL se espera obtener una metodología que recoge las mejores prácticas propuestas por los autores a lo largo del periodo de revisión, asegurando que si se sigue al pie de la letra se podrá obtener un mayor avance en el área de pronóstico de

series financieras ya que se podrán comparar todas las técnicas propuestas por diferentes autores para todo tipo de series de tiempo financieras.

1. Estructuración de modelos de pronóstico

1.1 Introducción

Las series de tiempo son un conjunto de datos organizados cronológicamente equidistantes entre sí. La econometría se encarga de recoger una o varias de estas series y modelarlas con el fin de obtener una representación matemática que permita capturar parcial o totalmente su comportamiento. Una vez construido un modelo este puede ser utilizado para realizar un pronóstico de una o varias variables estudiadas.

El estudio de las series de tiempo ha sido de particular interés en el área de las finanzas especialmente en los mercados financieros (Cao & Tay, 2001; Shen & Loh, 2004). Se ha observado que las series de tiempo financieras no tienen un comportamiento lineal y el pronóstico de estas ha resultado un proceso altamente complejo. Se ha pretendido asumir que el comportamiento de la volatilidad de estas series de tiempo se ajusta a una distribución normal estándar, pero cuando se intenta llevar este supuesto a la realidad, se obtiene que la distribución de la volatilidad presenta problemas de asimetría y kurtosis (Liljebloom & Stenius, 1997), Thawornwong y Enke (2004) por su parte sostienen que las series de tiempo financieras son influenciadas por un gran número de variables de mercado y las interrelaciones entre estas variables no son lineales dificultando así su pronóstico, Cao y Tay (2001) afirman que las series de tiempo financieras son ruidosas, no estacionarias y determinísticamente caóticas. Como consecuencia de lo descrito anteriormente la identificación y el pronóstico del comportamiento en series de tiempo financieras es un tópico de gran interés para los agentes de los mercados financieros en el cual existe una amplia discusión actualmente.

La inversión en acciones o índices bursátiles es básicamente sacrificar la posibilidad de gasto del presente a cambio de beneficios sobre el capital en el futuro (Pinches & Kinney, 1971), el objetivo del inversionista es optimizar la medida del desempeño ya sea medida en términos de rentabilidad, riesgo, o una mezcla de ambos (Moody & Saffell, 2001), además en años recientes se han incrementado los controles en las inversiones en los mercados financieros, debido a las burbujas, operaciones de especulación, incertidumbre económica, entre otros; centrando los controles en la medición del riesgo (Knight et al., 2003), dado que el futuro es incierto es importante tener modelos que pronostiquen con la información disponible los movimientos en los precios de los activos financieros. De todo esto se desprende que el pronóstico de series de tiempo financieras es una de las aplicaciones más relevantes en predicción por modelos matemáticos (Cao & Tay, 2001).

El primer objetivo que persigue este capítulo es presentar la importancia de los modelos de pronóstico en el área financiera especialmente en índices bursátiles, el segundo objetivo es dar una clasificación general a los modelos de pronóstico de acuerdo a sus principales características, el tercer aporte es identificar algunos problemas que se presentan en la comparación entre modelos y proponer una revisión sistemática de literatura como medio para obtener una guía en el modelado de series financieras, el cuarto objetivo es ilustrar el protocolo para la revisión sistemática y finalmente los objetivos de esta investigación.

1.2 Clasificación de los modelos de pronóstico

La revisión de literatura permite encontrar varias clasificaciones en los modelos de pronóstico que van desde el modelo matemático empleado para obtener la predicción hasta el número de variables explicativas que puede llegar a tener.

1.2.1 Modelos de pronóstico según variables explicativas:

Una variable explicativa es la variable que después de ser estudiada con una técnica de modelación pretende predecir el comportamiento de una variable estudiada.

Los modelos de pronóstico se clasifican en univariados y multivariados (Cao & Tay, 2001).

- Modelos univariados: Son aquellos que explican una variable estudiada con una variable explicativa.
- Modelos multivariados: Son aquellos que explican una variable estudiada con varias variables explicativas.

Las variables de entrada o variables explicativas tienen diversas características, pueden ser autorregresivas que es cuando la variable de entrada es la misma variable de salida, o tener como entradas una variable exógena que es cuando se pretende explicar una variable estudiada a través de una variable explicativa diferente.

Las principales variables de entrada en los modelos de pronóstico que se encuentran en la literatura son: indicadores técnicos (Brownstone, 1996) variables macroeconómicas (Arnold & Vrugt, 2008) y modelos autorregresivos.

1.2.2 Modelos de pronóstico según metodología

Los principales modelos matemáticos empleados para el pronóstico de índices bursátiles clasificados de acuerdo a su metodología son:

- Métodos estadísticos paramétricos: ARIMA, GARCH, ARCH, regresión lineal múltiple, etc.
- Métodos no paramétricos: Árboles de decisión, frontera eficiente.
- Computación blanda: Redes neuronales, lógica difusa, algoritmos genéticos.

La clasificación anterior es un resumen de las conclusiones reportadas por Atsalakis & Valavanis (2009), y Bahrammirzaee (2010), al realizar un repaso de más de 350 artículos sobre pronóstico y aplicaciones matemáticas en finanzas.

Los experimentos demuestran que en la mayoría de los casos las técnicas de computación blanda superan a los métodos estadísticos paramétricos (Atsalakis & Valavanis, 2009), y si se realiza un comparativo del desempeño de las técnicas de computación blanda, las redes neuronales artificiales presentan mejores resultados que los algoritmos genéticos y redes neuro difusas (Bahrammirzaee, 2010)

1.3 Medición del desempeño en modelos de pronóstico

Todo modelo de pronóstico tiene una medida asociada a su capacidad de predecir, y esta medida es definida por cada autor al momento de proponer un modelo, existen múltiples medidas y criterios para evaluar el desempeño de cada modelo. Las diferentes mediciones se centran en medición de errores, riesgo o rentabilidad.

Alguno modelos de pronóstico son medidos donde se hace necesario establece un nivel de significancia que lo establece el modelador y se define como la probabilidad de rechazar la hipótesis nula cuando esta es verdadera, esta probabilidad es definida por la letra α y la mayoría de ocasiones oscila entre 0.01 y 0.1 donde 0.05 es el número más usado en la literatura, cabe aclarar que 0.05 es un nivel de significancia seleccionado de forma arbitraria.

El error del modelo es la diferencia entre el valor real y el valor pronosticado, entre menor sea el error mejor se considera el modelo, en el desarrollo del documento se confirma que esto es cierto, sin embargo este error no puede ser medido de cualquier manera.

Atsalakis et al (2009) afirma que la medición del desempeño de los modelos de pronóstico se dividen en mediciones estadísticas (i.e. MAE, RMSE) y mediciones no estadísticas (i.e. retorno anual, tasa de aciertos) entre estos dos criterios él encuentra más de 20 medidas que no son comparables entre sí, mientras que Qi y Zhang(2001) clasifican la medición a través de criterios estadísticos en el periodo de entrenamiento y periodo de validación, mencionando más de 10 mediciones para los dos periodos.

Bahrammirzaee (2010) en una revisión de más de 250 artículos sobre aplicaciones de computación blanda en las finanzas, realiza un resumen que muestra una superioridad

de los modelos no lineales frente a los modelos tradicionales, sin embargo no muestra los criterios que se tuvieron en cuenta para establecer dicha superioridad.

Una revisión preliminar de literatura sobre modelos de pronóstico encontró que los autores seleccionan diferentes medidas a la hora de evaluar la significancia estadística y el error del modelo, solo por mencionar algunas mediciones en modelos: rendimiento frente a otras estrategias (Moody & Saffell, 2001), Valor en Riesgo (Knight et al., 2003), suma cuadrados del error y desviación absoluta media (Sadorsky, 2005).

Por todo lo expresado anteriormente, se concluye que no existe una manera de comparar los modelos de pronóstico de diferentes autores, esto se presenta porque los modeladores miden los modelos desde ópticas diferentes.

1.4 Propuesta revisión sistemática de literatura del índice S&P 500

Una revisión preliminar de modelos de pronóstico para índices bursátiles arroja que:

- El modelador selecciona a su criterio la medida de desempeño del modelo propuesto, al igual que el modelo con el cual lo desea comparar. Sadorsky (2005) encuentra que el modelo más apropiado según el criterio MSE es un modelo autorregresivo con 5 entradas, mientras con el criterio MAD afirma que es el modelo media móvil de 20 periodos, y finalmente bajo el criterio MPE el modelo de suavizamiento exponencial para 20 periodos arroja los mejores resultados. Esta misma inconsistencia con la selección del modelo la presenta (Qi & Zhang, 2001), (Durham, 2007).
- Las conclusiones que arrojan los autores son muy confusas, Becker Clemens y White (2007) concluyen que la volatilidad implícita es una variable explicativa del índice S&P 500, mientras que Martens y Zein (2004) afirman que la volatilidad implícita no aporta información adicional en el modelo propuesto por ellos.
- Atsalakis y Valavanis (2009) y Bahrammirzaee (2010) hacen una revisión amplia de modelos de pronóstico y aplicaciones financieras, ambos concluyen que los modelos de computación blanda son los que ofrecen mejores resultados, sin

embargo se limitan al resultado de la revisión y no se hace una evaluación más profunda de los hechos que llevaron a esta conclusión.

- El índice bursátil objeto de estudio de la mayoría de estudios es el índice Standard and Poor's 500.

La problemática detectada en los modelos de pronóstico para índices bursátiles es que no se puede realizar un comparativo para determinar cuál es la metodología, datos de entrada y horizonte de tiempo, que arrojan mejores resultados. Existe la necesidad de comparas bajo parámetros similares, aunque se puede hacer un estudio de todos los modelos propuestos para todas las series de datos financieras, se selecciona al índice Standard and Poor's 500 por ser la serie de tiempo más estudiada en la literatura, es preferible tener un índice de referencia para que se puedan identificar comportamientos que sean atribuibles a los modelos de pronóstico y no a las series de tiempo.

1.4.1 Índice Standard and Poor's 500

El índice S&P 500 se enfoca en las 500 empresas de gran capitalización bursátil del mercado, en otras palabras empresas líderes en sectores líderes, algunas de las condiciones que deben de cumplir estas empresas son:

- Capitalización bursátil superior a los 4 billones de dólares.
- Liquidez adecuada y precio razonable, debe tranzar en bolsa con un volumen (en USD) no inferior a su capitalización bursátil cada año.
- Ser una empresa estadounidense.
- Viabilidad financiera en los últimos 4 trimestres de acuerdo a las prácticas contables aceptadas del momento.

Los pesos de cada empresa en el índice se reevalúan el último viernes de cada trimestre, la conformación del índice es un poco más compleja con el fin de evitar que este refleje información que no corresponda a una visión global del mercado (*S&P U.S. Indices Methodology, 2012*).

Teniendo en cuenta que el S&P 500 es un índice que está en capacidad de reflejar el mercado accionario estadounidense, se considera que es la mejor opción para estudiar el

desempeño de los diferentes modelos de pronóstico cuyo objeto sea predecir índices bursátiles.

1.4.2 Revisión sistemática de literatura

Es un medio que tiene como objetivo reunir grandes cantidades de información, y clasificarla según su grado de relevancia en un tópico de interés determinado previamente por el investigador (Petticrew & Roberts, 2006). Webster & Watson (2002) definen a la revisión sistemática de literatura (RSL) como el conducto para sintetizar la investigación existente en un tema específico. En conclusión la RSL es un resumen sobre la investigación disponible en un área de interés que busca responder una o varias preguntas.

En el desarrollo del texto se han encontrado dos problemas que se pretenden resolver con una RSL, el primero es determinar cuál es el modelo de pronóstico que arroja predicciones más exactas y el segundo es como se debe de medir esta exactitud de tal manera que se puedan comparar varios modelos propuestos por varios autores, enfocado al objetivo de la modelación.

Se propone entonces hacer una RSL del índice S&P 500 ya que este índice es el más estudiado en la literatura por las características que se mencionaron en el numeral anterior.

Varios autores (Budgen & Brereton, 2006; Levy & Ellis, 2006; Webster & Watson, 2002) proponen una metodología adecuada de cómo realizar una RSL, siempre que los objetivos estén previamente definidos, mientras que los mencionados al inicio del párrafo parten desde cero. Por otro lado Petticrew y Roberts (2006) proponen un proceso el cuál se basa en responder una serie de preguntas que van a definir el norte de la investigación desde cero, e incluye la definición de objetivos generales y específicos. A continuación se desarrollarán las preguntas propuestas que servirán como guía para realizar una RSL adecuada:

- ¿Qué desea responder la RSL?

Cuál es el modelo de pronóstico que arroja las mejores predicciones para el índice S&P 500, para llegar a este objetivo también es necesario tener claro cuál

es el criterio de selección de modelo, lo que lleva a responder cuales son los criterios más adecuados para evaluar modelos de pronóstico.

Se considera la necesidad de una RSL porque no se encuentra una que haga una evaluación a fondo de los resultados, simplemente se limitan a reportar el resultado del autor sin hacer un análisis, ni un contraste entre artículos, no se tiene en cuenta que el hecho seleccionar un modelo por encima de otro tiene una dependencia directa de la escogencia de la medida de desempeño.

- ¿Cuál va a ser el alcance de la revisión?

Artículos que tengan como objetivo el pronóstico de índice S&P 500 publicados en los últimos 12 años, se había planteado inicialmente un periodo de 10 años sin embargo se tuvo en cuenta que en ocasiones el periodo de publicación puede tardar hasta dos años entonces de esta manera se toma una década completa, se tomara como fuente de información las principales bases de datos que tiene acceso la Universidad Nacional de Colombia

- ¿Cuál va a ser el protocolo de búsqueda?

El protocolo de búsqueda se muestra con detalle en la siguiente sección.

- ¿Qué se va a extraer de la información revisada?

Modelos de pronóstico propuesto

Descripción de los datos de entrada: número de datos, periodicidad, procesos de suavizamiento.

Medida de desempeño propuesta por el autor.

Resultados y conclusiones

Información adicional que se pueda considerar relevante.

- ¿Cómo se va a determinar los posibles sesgos que tenga la información revisada?

Resulta imposible encontrar sesgos en los artículos estudiados sino se tiene un total acceso a la información que fue necesaria para la publicación, teniendo en cuenta que la recolección de toda esta información demanda una gran cantidad de recursos, y además todos los artículos tuvieron que pasar por un proceso de revisión de parte de jueces anónimos, se supondrá que la información reportada

es cierta. Aún cuando la información sea cierta se puede detectar ciertos sesgos en la manera como se evalúan los modelos y se hará un análisis de esto en el desarrollo de la revisión

- ¿En qué forma se van a presentar los resultados obtenidos?
Agrupación por metodología y medición de desempeño.
- ¿Qué publicaciones se pretenden obtener y cuál es el público interesado?
Se pretende publicar un artículo en una revista especializada en el área financiera, en donde el público interesado será cualquier persona dedicada a la inversión en mercados financieros y/o investigaciones econométricas.

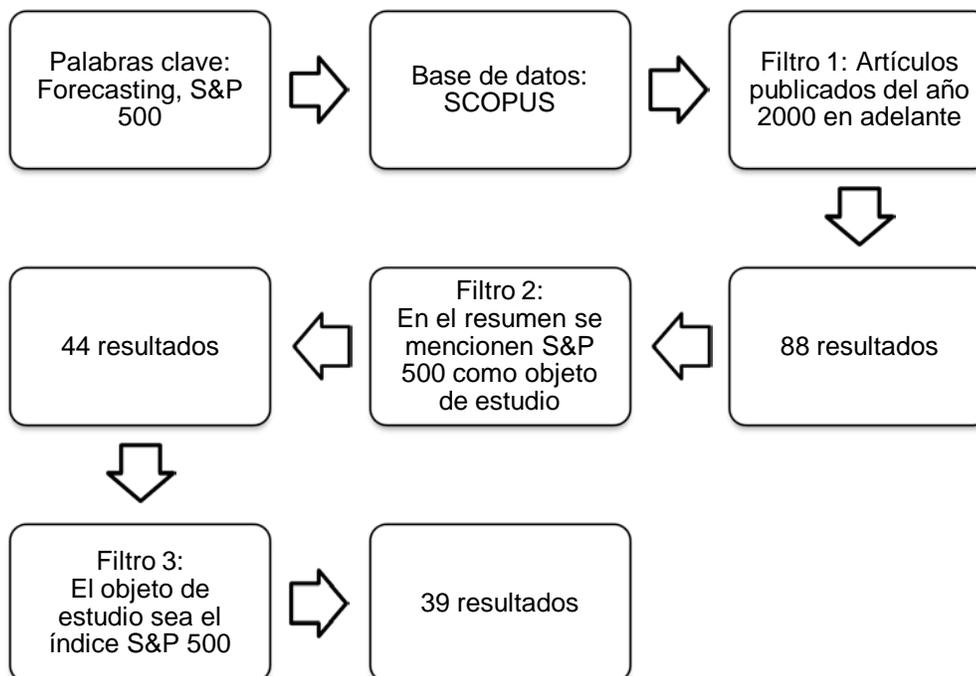
1.5 Protocolo revisión sistemática de literatura

El método de búsqueda es el primer gran problema que debe de resolver el investigador, si selecciona muchos artículos podría desviarse de su verdadero objetivo y seleccionar muy pocos empobrecería la investigación y tendría poca credibilidad.

El punto de partida de un protocolo de revisión sistemática es especificar la pregunta a la cuál va a ser orientada la investigación y los métodos empleados (Budgen & Brereton, 2006). La pregunta de investigación es: ¿Cuáles son los métodos de pronóstico que arrojan los mejores resultados para el índice S&P 500?

La estrategia de búsqueda es la selección de palabras clave en donde se utilizó el tesaurus IEEE (2009) y se obtuvieron las siguientes palabras clave: Forecasting, S&P 500. Se determinó un periodo de búsqueda de 10 años, no obstante es necesario tener en cuenta que el proceso de publicación puede llevar hasta 2 años, entonces se decide tomar un periodo de búsqueda de 12 años es decir del año 2000 en adelante. Posteriormente se realizan filtros con el fin de apuntar al objetivo general de esta investigación, el proceso de búsqueda se muestra en la Gráfico 1.1

Gráfico 1.1: Proceso de búsqueda en bases de datos



1.6 Objetivos de la investigación

1.6.1 Objetivo general

Identificar los modelos de pronóstico que arrojan mejores predicciones del índice Standard and Poor's 500

1.6.2 Objetivos específicos

Identificar los modelos de pronóstico que tienen como objetivo el índice S&P 500 a través de una revisión sistemática de literatura, para obtener una visión global de predicción de índices bursátiles.

Describir las medidas de desempeño de los modelos de pronóstico propuestos, que permitan establecer cuáles son las más apropiadas de acuerdo al objetivo del modelador.

Enumerar los criterios de validación de datos utilizados en modelos de pronóstico, con el fin de identificar las propuestas que tengan mayor credibilidad desde el punto de vista estadístico.

Comparar los resultados de la revisión bibliográfica, para determinar los modelos de pronóstico más acertados y más confiables.

Proponer un protocolo de selección de medidas de desempeño para futuras aplicaciones de predicción, dirigido hacia objetivos y que permitan comparaciones entre varios autores y metodologías.

2. Medición en modelos de pronóstico

Los modelos matemáticos que tienen como objetivo la predicción generalmente tienen dos tipos de mediciones de desempeño, la primera es asociada a la precisión del modelo comparado con el objetivo deseado, y la segunda es acerca de la confiabilidad del pronóstico del modelo, en otras palabras que el pronóstico del modelo no sea una casualidad.

Se considera que antes de entrar en los resultados que arrojan los modelos de pronóstico del S&P 500 es necesario hacer un repaso sobre las mediciones reportadas en la literatura, explicando el funcionamiento y el objetivo de estas.

Este capítulo pretende responder las siguientes preguntas:

- ¿Cuáles son las medidas de desempeño asociadas a la precisión?
- ¿Cuáles son las medidas de desempeño asociadas a la validez?

Con el fin de responder estas preguntas se hace un resumen de los resultados obtenidos en la revisión sistemática de literatura. Para el reporte de las medidas se utilizará el acrónimo en inglés con el fin de ampliar la utilidad de este capítulo en futuras investigaciones.

2.1 Medidas de desempeño asociadas a la precisión

En este capítulo se recogen todas las medidas de desempeño encontradas en la RSL con el fin de resumir en un solo capítulo la información contenida en todos los artículos en cuanto a desempeño se refiere

2.1.1 SSE Suma de los cuadrados del error

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Ecuación 2-1.SSE

Es la diferencia entre el valor pronóstico y el valor real de la variable estudiada, no presenta de manera específica si este error presenta un beneficio o una pérdida en la inversión(Devore, 2001).

2.1.2 AIC Criterio de información de Akaike

$$AIC = \log(\sigma_{MLE}^2) + 2m/n$$

Ecuación 2-2.AIC

Donde σ_{MLE}^2 corresponde al estimador de máxima verosimilitud y está definido por:

$$\sigma_{MLE}^2 = \frac{SSE}{n}$$

Este criterio de evaluación es el más común en modelos lineales y no lineales (Qi & Zhang, 2001). La primer parte de la Ecuación 2-2 mide la bondad de ajuste del modelo a los datos, mientras la segunda parte penaliza por sobreparametrización del modelo. Se considera que el modelo es mejor mientras más pequeño sea criterio de información de Akaike.

2.1.3 AICC Criterio de información de Akaike modificado

$$AICC = \log(\sigma_{MLE}^2) + \frac{2m}{n - m - 1}$$

Ecuación 2-3.AICC

Es un criterio similar al AIC la diferencia radica en la segunda parte de la ecuación que usa un término de penalización más apropiado que aminora la tendencia a la sobreparametrización del criterio AIC. Para modelos complejos el AICC penaliza más

fuerte que el AIC, por eso es importante revisar las características del modelo para seleccionar el criterio adecuado. Qi y Zhang (2001) hacen una revisión en la que encuentran otras 4 variaciones del AIC.

2.1.4 BIC Criterio de información bayesiano

$$BIC = \log(\sigma_{MLE}^2) + \frac{m \log(n)}{n}$$

Ecuación 2-4.BIC

Es muy similar al AIC y la primer parte de la ecuación es igual al AIC, la diferencia se presenta en el término de penalización, si $T > 7$ la penalización del criterio es mayor que en el AIC simplemente por ser un modelo más complejo, se debe de tener en cuenta que el criterio debe de ser usado para modelos con pocos parámetros (Qi & Zhang, 2001).

2.1.5 CD Corrección tendencia bajista

$$CD = \frac{100}{n} \sum a_i, \text{ donde } a = \begin{cases} 1, \text{ si } (\hat{y}_{i+1} - \hat{y}_i) < 0, (y_{i+1} - y_i) (\hat{y}_{i+1} - \hat{y}_i) \geq 0 \\ 0, \text{ de otra manera} \end{cases}$$

Ecuación 2-5.CD

El CD es un criterio que se encarga de informar cuantas veces el modelo identificó la tendencia bajista, junto con el CU permiten identificar el número de veces que el modelo identifico correctamente la tendencia, se puede decir que es uno de los criterios más útiles porque puede ocurrir que el modelo no capture el valor del activo, sin embargo al capturar la dirección se asegura una utilidad (Cao & Tay, 2001).

2.1.6 CU Corrección tendencia alcista

$$CU = \frac{100}{n} \sum a_i, \text{ donde } a = \begin{cases} 1, \text{ si } (\hat{y}_{i+1} - \hat{y}_i) > 0, (y_{i+1} - y_i) (\hat{y}_{i+1} - \hat{y}_i) \geq 0 \\ 0, \text{ de otra manera} \end{cases}$$

Ecuación 2-6.CU

2.1.7 DA Precisión de la dirección

$$DA = \frac{1}{n} \sum a_i, \text{ donde } a = \begin{cases} 1, \text{ si } (y_{i+1} - y_i) (\hat{y}_{i+1} - y_i) > 0 \\ 0, \text{ de otra manera} \end{cases}$$

Ecuación 2-7.DA

Esta medida el término de la sumatoria se encarga de determinar el porcentaje que el pronóstico del modelo identifica correctamente la dirección del movimiento, a mayor DA mejor es el modelo (Cao & Tay, 2001).

2.1.8 DM Diebold and Mariano

Esta es una prueba un poco más compleja que las tradicionales, evalúa el pronóstico en dos modelos en donde se acepta o rechaza hipótesis de la información contenida entre los dos modelos.

$$d_t = L(e_t^1) - L(e_t^2)$$

Ecuación 2-8.DM

Donde 1 y 2 son los modelos de pronóstico comparados y e equivale a los errores que pueden ir descritos en medidas arbitrarias como el MSE, RMSE. La prueba Diebold Mariano es basada en la media muestral $\bar{d} = \frac{\sum d_t}{T}$ donde se asume que:

$$DM = \frac{\bar{d}}{\sqrt{\sigma_{\bar{d}}^2}} \rightarrow N(0,1)$$

Donde el valor de DM se compara con los valores críticos de una distribución t-student con T-1 grados de libertad, donde la hipótesis nula acepta que la información del modelo 2 está contenida en el modelo 1. (Diebold & Mariano, 1995)

2.1.9 MDM Diebold y Mariano modificada

$$MDM = \left[\frac{T + 1 - 2h + T^{-1}h(h-1)}{T} \right]^{1/2} DM$$

Ecuación 2-9.MDM

La diferencia de este estadístico es que mejoran el poder de la prueba cuando el número de muestras es pequeño (Qi & Zhang, 2001).

2.1.10 ES Pérdida esperada

$$ES_{\alpha} = E(r \parallel r > VaR_{\alpha})$$

Ecuación 2-10.ES α

El ES es por lo tanto un promedio de los peores 100(1- α) % casos que excedan el umbral con un nivel de confianza de α , a mayor α mayor riesgo (Watanabe, 2012).

2.1.11 Hit ratio

$$Hit\ ratio = \frac{1}{n} \sum a_i, \text{ donde } a = \begin{cases} 1, & \text{si } y_{i+1} = \hat{y}_{i+1} \\ 0, & \text{de otra manera} \end{cases}$$

Ecuación 2-11. Hit Ratio

Mide la proporción que el pronóstico del modelo es igual al dato real, resulta algo inflexible porque se puede presentar el caso que el pronóstico no sea exacto y aún así obtener utilidades (Yu, Chen, Wang, & Lai, 2009).

2.1.12 HMSE MSE heterocedástico

$$HMSE = \frac{1}{n} \sum \left(1 - \frac{y_i}{\hat{y}_i}\right)^2$$

Ecuación 2-12.HMSE

Es una variación del MSE que permite tener una mejor estimación de la capacidad para pronosticar del modelo (Martens, Van Dijk, & Pooter, 2009).

2.1.13 MAE Error medio absoluto

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

Ecuación 2-13.MAE

Es la medida más utilizada para medir desempeño en modelos de pronóstico, no identifica la dirección del error (Devore, 2001).

2.1.14 Max drawdown

$$Max\ drawdown = \frac{Valor\ máximo - valor\ mínimo}{Valor\ inicial}$$

Ecuación 2-14. Max drawdown

Es una medida que expresa la exposición que tiene una inversión en un periodo de tiempo determinado, medida similar a la volatilidad.

2.1.15 MAPE Error absoluto porcentual medio

$$MAPE = \frac{1}{n} \sum \left| \frac{(y_i - \hat{y}_i)}{y_i} \right|$$

Ecuación 2-15.MAPE

Entre el RMSE, el SSE y el MAPE es la medida más apropiada para la comparación entre modelos ya que no se presentan problemas por las unidades debido a que se expresa en porcentaje. Presenta un gran problema que no evalúa rendimientos económicos, y se puede presentar la situación que aunque el error sea pequeño genere pérdidas en las inversiones (Devore, 2001).

2.1.16 ME Media del error

$$ME = \frac{1}{n} \sum (y_i - \hat{y}_i)$$

Ecuación 2-16.ME

Muestra un promedio de las mediciones del error, presenta el problema que un error en direcciones opuestas lo anula arrojando una medida errónea (Devore, 2001).

2.1.17 MME(O) Media del error mixto

$$MME(O) = T^{-1} \left[\sum_{t=1}^o \sqrt{|y_i - \hat{y}_i|} + \sum_{t=1}^u |y_i - \hat{y}_i| \right]$$

Ecuación 2-17.MME(O)

Donde O son las sobre estimaciones y U las subestimaciones del modelo que se evalúa, T=U+O, esta medida junto con el MME(U) castigan las sobre estimaciones y sub estimaciones de manera diferente para detectar cuales de las dos estimaciones se encuentra más presente en el modelo (Hung, Ni, & Chang, 2009).

2.1.18 MME(U)

$$MME(U) = T^{-1} \left[\sum_{t=1}^o \sqrt{|y_i - \hat{y}_i|} + \sum_{t=1}^u \sqrt{|y_i - \hat{y}_i|} \right]$$

Ecuación 2-18.MME(U)

Ver 2.1.17

2.1.19 MPE Error porcentual medio

$$MPE = \frac{1}{n} \sum \left(\frac{(y_i - \hat{y}_i)}{y_i} \right)$$

Ecuación 2-19. MPE

Indica el error como un porcentaje, presenta el mismo problema de la ME (Sadorsky, 2005).

2.1.20 MSE Media cuadrada del error

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

Ecuación 2-20.MSE

Este indicador elimina el efecto de anulación que tiene el ME y MPE, a menor error mejor es el modelo (Devore, 2001).

2.1.21 NMSE Error cuadrático medio normalizado

$$NMSE = \frac{1}{\sigma^2 n} \sum (y_i - \hat{y}_i)^2$$

Ecuación 2-21. NMSE

Es la medida que permite comparar el error con la volatilidad, se espera que el NMSE sea una parte pequeña de la volatilidad para tener una buena estimación del modelo (Cao & Tay, 2001).

2.1.22 R² Coeficiente de determinación

$$R^2 = 1 - \frac{SSE/(n - m)}{\sum (y_i - \hat{y}_i)^2 / (n - 1)}$$

Ecuación 2-22 R²

Describe la proporción que el modelo explica el comportamiento de la variable estudiada, entre más alto sea este indicador mejor es el modelo (Devore, 2001).

2.1.23 Rentabilidad

$$R_t = \frac{\text{Valor del activo al final del periodo}}{\text{Valor del activo al inicio del periodo}}$$

Ecuación 2-23. R_t

La rentabilidad es el resultado de una inversión después de un periodo de tiempo determinado, se le debe de restar uno para obtener la rentabilidad total de la inversión en porcentaje.

2.1.24 RMSE Error cuadrático medio

$$RMSE = \sqrt{\frac{1}{n} SSE}$$

Ecuación 2-24.RMSE

Es una medida muy similar al MSE la diferencia es que el RMSE se parece más a las unidades estudiadas (Johnston & Dinardo, 1997).

2.1.25 Sharpe ratio

$$S = \frac{R_t}{\sigma_t}$$

Ecuación 2-25.Sharpe ratio

Es la división entre la rentabilidad y la desviación estándar en un periodo de tiempo determinado, entre más alta sea la medida se considera que el activo está rentando más bajo un mismo nivel de riesgo determinado (Sharpe, 1994)

2.1.26 Sterling ratio

$$Sterling\ ratio = \frac{Rentabilidad\ anual}{Max\ drawdown}$$

Ecuación 2-26.Sterling ratio

Se interpreta exactamente igual al Sharpe ratio, la diferencia es que toma al Max drawdown como medida de volatilidad (Moody & Saffell, 2001).

2.1.27 Theil U

$$Theil\ U = \frac{\sqrt{\sum (y_i - \hat{y}_i)^2}}{\sqrt{\sum y_i^2} + \sqrt{\sum \hat{y}_i^2}}$$

Ecuación 2-27 Theil U

Entre más cercano se encuentre a cero mejor es la predicción del modelo (Sadorsky, 2005).

2.1.28 VaR Valor en riesgo

$$VaR_{\alpha} = Z_{\alpha} \sigma_t$$

Ecuación 2-28. VaR

Se interpreta de la siguiente manera: Existe una probabilidad de $1-\alpha$ que la pérdida máxima en un tiempo t sea VaR_{α} (Watanabe, 2012)

2.2 Pruebas de validación de datos

2.2.1 Box Pierce

$$Q = n \sum_{j=i}^L r_j^2 \sim \chi^2 L, \text{ donde } r_j = \frac{\sum_{n=j+1}^n \varepsilon_t \varepsilon_{t-j}}{\sum_{t=1}^n \varepsilon_t^2}$$

Donde ε representa los residuos es una prueba para determinar si la serie de datos tiene problemas de autocorrelación, el número de rezagos L se selecciona bajo criterio del modelador y la hipótesis nula se rechaza si $Q > \chi^2 L$ (Box & Pierce, 1970).

$$\text{Prueba de hipótesis} \begin{cases} H_0 : \text{Ausencia de correlación en la muestra} \\ H_1 : \text{Presencia de correlación en la muestra} \end{cases}$$

2.2.2 Chi cuadrado

$$\chi^2 = \sum \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}$$

Donde χ^2 es el estadístico de prueba y se compara con el resultado obtenido de la distribución $\chi^2_{\alpha, k-1}$ donde k es el número de grados de libertad, y α nivel de

significancia cuando $\chi^2 > c_{\alpha, k-1}$ se rechaza la hipótesis nula afirmando que existen discrepancias entre el valores observados y valores esperados (Devore, 2001)

2.2.3 Durbin Watson

$$d = \frac{\sum_{t=2}^n ((y_t - \hat{y}_t) - (y_{t-1} - \hat{y}_{t-1}))^2}{\sum (y_t - \hat{y}_t)}$$

Es un estadístico cuyo objetivo es comprobar si existe o no autocorrelación de los errores y sus posibles valores están comprendidos entre 0 y 4, se interpreta de la siguiente manera:

$$\begin{aligned} 0 \leq d \leq dl, & \quad \text{existe autocorrelación positiva} \\ 4 - dl \leq d \leq 4, & \quad \text{existe autocorrelación negativa} \\ du \leq d \leq 4 - du, & \quad \text{no existe autocorrelación} \\ & \quad \text{de otra manera, ausencia de evidencia} \end{aligned}$$

Donde dl y du son valores obtenidos de la tabla de significancia Durbin Watson y dependen del número de datos, parámetros, y nivel de significancia (Devore, 2001).

2.2.4 Fisher Snedecor

$$F = \frac{MSTr}{MSE}$$

$$\text{Prueba de hipótesis} \begin{cases} H_0 : E(MSTr) = E(MSE) = \sigma^2 \\ H_1 : E(MSTr) > E(MSE) = \sigma^2 \end{cases}$$

Donde MSTr es la varianza poblacional, el resultado de F se compara con el estadístico de prueba de la distribución F con n como número de observaciones, k el número de grados de libertad, y α nivel de significancia cuando $F > F_{\alpha, k-1, k(n-1)}$ se rechaza la hipótesis nula. La prueba de Fisher Snedecor tiene como objetivo comparar las varianzas poblacionales y muestrales con el fin de determinar si la variabilidad de la muestra si describe la variabilidad de la población (Devore, 2001)

2.2.5 Jarque Bera

$$JB = \frac{n}{6} (\text{kurtosis}^2 + 0.25(\text{Asimetría} - 3)^2)$$

Esta prueba es para verificar normalidad, donde la hipótesis nula es que la asimetría es cero y la kurtosis es de 3 o lo mismo que decir que no se presente exceso de la misma, la región de rechazo es $JB > \chi^2_{\alpha,2}$ (Rencher, 2002).

2.2.6 Ljung box

$$Q_{LB} = n(n+2) \sum_{j=i}^L \frac{r_j^2}{n-j} \sim \chi^2 L$$

Es una prueba exactamente igual a la prueba Box Pierce solo que cambia el estadístico de prueba. Este estadístico es más reciente que la prueba Box Pierce (Ljung & Box, 1978).

2.2.7 LM Multiplicador de Lagrange

$$LM = nR^2$$

$$\text{Prueba de hipótesis} \begin{cases} H_0 : \text{Ausencia de correlación en la muestra} \\ H_1 : \text{Presencia de correlación en la muestra} \end{cases}$$

Es una prueba de autocorrelación, tiene el mismo problema que la prueba Box Pierce en el número de rezagos que se deben de elegir, si $LM > \chi^2 L$ se rechaza la hipótesis nula (Johnston & Dinardo, 1997).

2.2.8 Q like

$$Q \text{ like} = \log(\hat{y}_i) + \frac{y_i}{\hat{y}_i}$$

Es una medida similar al AIC, penaliza más la subestimación que la sobrestimación (Becker & Clements, 2008).

2.2.9 t-student

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$\text{Prueba de hipótesis} \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

El resultado de t se contrasta con el estadístico de prueba de una distribución t con nivel de significancia α , n como el número de observaciones, si $t > t_{\alpha, n-1}$ se rechaza la hipótesis nula (Devore, 2001).

Para el caso multivariado se debe de hacer una prueba Hotelling que se realiza de la misma manera en donde la hipótesis se conserva pero en vez de aplicar a una media, se utiliza para un vector de medias (Rencher, 2002).

2.3 Selección de medidas de desempeño para modelos de pronóstico de índice bursátil

Esta sección tiene como objetivo discutir la utilidad de los criterios que miden el desempeño en modelos de pronóstico enfocados a índices bursátiles.

Los modelos de pronóstico son ampliamente utilizados en diferentes áreas del conocimiento como el sector eléctrico, economía, meteorología, biología, administración, entre otras (Johnston & Dinardo, 1997). La capacidad de pronóstico de un modelo se puede medir con uno o varios de los criterios encontrados en la RSL (Sección 2.1), sin embargo para los modelos de pronóstico de índice bursátil, estos criterios arrojan información que puede ser mal interpretada.

Los modelos de pronóstico se formulan principalmente con los objetivos de obtener rentabilidad y minimizar el riesgo en inversionistas, en base a esta afirmación se hace el análisis de la utilidad de los criterios mencionados anteriormente.

Se pretende simular un escenario basado en cinco pruebas a 2 supuestos modelos, comparando el valor real de una serie con el valor pronosticado y así encontrar las medidas de desempeño SSE, MSE, RMSE, MAE y MAPE ampliamente utilizadas en la literatura. En la Tabla 2-1, se puede observar que las diferentes medidas de desempeño tienen el mismo valor para diferentes valores pronosticados por los modelos, se puede concluir que estos criterios no tienen en cuenta subestimación y sobre estimación. A continuación se demostrará como estos criterios podrían hacer que un inversionista tome decisiones erradas.

Tabla 2-1. Escenario simulado para cálculo de diferentes medidas de desempeño

Pruebas de desempeño n=5					
	1	2	3	4	5
Modelo 1	107	108	106	105	109
Modelo 2	93	94	92	95	91
Real	100	100	100	100	100
SSE					
	SSE	MSE	RMSE	MAE	MAPE
Modelo 1	255	51	7.141	7	7.0%
Modelo 2	255	51	7.141	7	7.0%

En la Tabla 2-2 se puede observar una simulación de lo que podría ocurrir al utilizar los modelos propuestos, en donde se tiene un activo S con un precio inicial de 102.5 y se simula 5 veces el precio un periodo adelante para observar los resultados con ambos modelos.

Tabla 2-2. Prueba real a criterios de desempeño más frecuentes

Pruebas reales S=102.5					
	Prueba1	Prueba 2	Prueba 3	Prueba 4	Prueba 5
Real t+1	105	103	104	103	104
Modelo 1	110	112	108	111	110
Operación	Compra	Compra	Compra	Compra	Compra
Pérdida/ Ganancia	2.5	0.5	1.5	0.5	1.5

Modelo 2	98	95	95	96	97
Operación	Venta	Venta	Venta	Venta	Venta
Pérdida/ Ganancia	-4.5	-7.5	-7.5	-6.5	-5.5

El caso propuesto en la Tabla 2-2 arroja resultados muy diferentes para modelos que manejan los mismos errores, este problema se puede presentar al ignorar la subestimación o sobrestimación de un modelo en particular, porque si el activo hubiera presentado resultados por debajo del precio al inicio del periodo, el modelo 1 presentaría pérdidas, por todo lo anterior se considera inapropiado tomar como medida de desempeño medidas como el RMSE, MAE, MAPE, SSE y MSE. Existen otros criterios de desempeño en modelos más complejos en el cálculo que también tienen en cuenta las medidas tomadas en la Tabla 2-1 lo cual supone que también son inapropiadas para medir la precisión de un modelo de pronóstico de índice bursátil. Estas medidas son: AIC, AICC, BIC, DM, MDM, HMSE, NMSE, R^2 , Theil U, MME(O) y MME (U), estas últimas dos si toman en cuenta la subestimación y sobrestimación, sin embargo los resultados pueden ser ambiguos.

Descartadas dieciséis de las medidas para medir el desempeño de modelos de pronóstico utilizadas que se encontraron en la RSL, se prosigue a evaluar las medidas direccionales, propuestas en los modelos de varios autores.

Tabla 2-3. Prueba a criterios de desempeño direccionales

	DA	Hit ratio	CU	CD
Modelo 1	100%	0%	100%	0%
Modelo 2	0%	0%	0%	0%

En la Tabla 2-3, se evalúan cada uno de los criterios con los datos de la Tabla 2-2, y se encuentra que todos los criterios son apropiados para medir precisión en los modelos de pronóstico, es importante hacer algunas precisiones para el uso de estos, Hit ratio es un indicador muy riguroso debido a que exige un acierto del 100%, sin embargo un modelo de pronóstico puede generar utilidades sin tener este grado de acierto, DA es el indicador direccional más apropiado de los 4 evaluados en la Tabla 2-3 siempre y cuando se esté

prediciendo en un mercado y un activo que permita ventas en corto, porque el modelo puede ser mejor prediciendo momentos de venta que de compra o viceversa, finalmente CU y CD en conjunto eliminan la restricción de los mercados y activos en corto que presenta el DA, para obtener resultados óptimos con CD y CU es obligatorio tener en cuenta los dos siempre.

Las otras mediciones son bastante simples y son útiles para medir desempeño de modelos de pronóstico en índice bursátil. Clasificación y caracterización de criterios de desempeño en modelos de pronóstico aplicado a activos financieros

Tabla 2-4. Clasificación y caracterización de criterios de desempeño en modelos de pronóstico aplicado a activos financieros

Criterio	Medición	Utilidad		Problemas
		SI	NO	
SSE	Error			Sobrestimación y subestimación
AIC	Error			Sobrestimación y subestimación
AICC	Error			Sobrestimación y subestimación
BIC	Error			Sobrestimación y subestimación
CD	Dirección			Ninguno
CU	Dirección			Ninguno
DA	Dirección			Mercados sin cortos
DM	Error			Sobrestimación y subestimación
MDM	Error			Sobrestimación y subestimación
ES	Riesgo			Ninguno
Hit ratio	Error			Altamente riguroso
HMSE	Error			Sobrestimación y subestimación
MAE	Error			Sobrestimación y subestimación
Max drawdown	Volatilidad			Ninguno
MAPE	Error			Sobrestimación y subestimación
ME	Error			Sobrestimación y subestimación
MME(O)	Error			Sobrestimación y subestimación

Criterio	Medición	Utilidad		Problemas
		SI	NO	
MME(U)	Error			Sobrestimación y subestimación
MPE	Error			Sobrestimación y subestimación
MSE	Error			Sobrestimación y subestimación
NMSE	Error			Sobrestimación y subestimación
R2	Aproximación			Sobrestimación y subestimación
Rentabilidad	Beneficio			Ninguno
RMSE	Error			Sobrestimación y subestimación
Sharpe ratio	Rt y riesgo			Ninguno
Sterling ratio	Rt y riesgo			Ninguno
Theil U	Error			Sobrestimación y subestimación
VaR	Riesgo			Ninguno

3. Extracción de información RSL

Este capítulo contiene la información extraída en cada uno de los artículos recuperados de la búsqueda en SCOPUS. La información que se extrae es la siguiente:

- Modelos de pronóstico propuesto
- Descripción de los datos de entrada: número de datos, periodicidad, procesos de suavizamiento.
- Medida de desempeño propuesta por el autor.
- Resultados y conclusiones
- Información adicional que se pueda considerar relevante.

La información extraída se expondrá en orden cronológico debido a que los modelos recientes se presentan dos situaciones la primera es que se apoyan en las propuestas anteriores y la segunda es que el nivel de complejidad de las metodologías empleadas incrementa.

3.1 An investigation of model selection criteria for neural network time series forecasting (Qi & Zhang, 2001)

Descripción de los datos: Mensual de 01/1954 a 12/1992, se toma como dato el reporte del índice para el último día del mes. Los datos se suavizan mediante la función logaritmo natural, los autores afirman que este suavizamiento permite obtener mejores resultados. El periodo de entrenamiento es desde enero de 1954 hasta diciembre de 1984, el periodo de prueba comprende desde enero de 1985 hasta diciembre de 1992.

Metodología de pronóstico: Red neuronal multicapa hacia adelante

Procedimiento y resultados: Se corrieron diferentes conjuntos de redes neuronales variando el número de entradas y neuronas, el artículo no reporta todos los resultados,

sin embargo se reportan las redes neuronales que obtuvieron los mejores resultados de acuerdo a los diferentes criterios de desempeño. Además se hace un estudio de diferentes medidas de desempeño para el periodo de entrenamiento, sin embargo los resultados de los modelos arrojan que estas medidas no son confiables porque los modelos de mejor desempeño en periodo de entrenamiento no eran los mismos que en el periodo de prueba, algunas de estas medidas son AIC, AICC, BIC.

Tabla 3-1. Resultados (Qi & Zhang, 2001)

	Horizonte de tiempo					
	1 año		4 años		8 años	
	Modelo	x 100	Modelo	x 100	Modelo	x 100
RMSE	(5,5)	3.149	(5,1)	3.898	(5,1)	3.514
MAE	(5,5)	2.629	(5,3)	3.814	(5,3)	1.986
MAPE	(2,3)	0.847	(4,2)	0.927	(4,2)	55.208
ME	(5,5)	*	(5,3)	*	(1,4)	*
DA	(2,3)	*	(1,2)	*	(4,3)	*
Sign	(4,2)	*	*	*	(4,1)	*

En la Tabla 3-1 se puede observar que la red que arrojó resultados más precisos según los criterios RMSE es la red neuronal (5,5) para un año, según MAE es la red (5,3) para 8 años, y según el MAPE la red (2,3) para un año.

Conclusiones: Los criterios de desempeño en redes neuronales no deben tomarse a partir del periodo de entrenamiento, debido a que los resultados que arrojan en este periodo difieren totalmente de los resultados en el periodo de validación y este último es el que mide el desempeño del modelo en situaciones reales. Además se observa claramente que dependiendo del criterio de desempeño la selección modelo cambia.

3.2 Financial Forecasting Using Support Vector Machines (SVM) (Cao & Tay, 2001)

Descripción de los datos: Diario, desde 04/01/1993 a 29/12/1995, el periodo comprendido entre 04/01/1993 y 30/12/1994 se toma como periodo de entrenamiento, mientras que el periodo entre 03/01/1995 y 29/12/1995 periodo de validación del modelo obtenido en el periodo de entrenamiento. Un primer grupo de datos (G-1) resulta de la transformación de la serie original en diferencias relativas porcentuales en periodos de 5 días. También se estudia un segundo grupo (G-2) de datos constituido con los indicadores técnicos media móvil de convergencia divergencia (MACD), volumen de equilibrio (OBV), y volatilidad.

Metodología de pronóstico: Maquinas de vectores de soporte (SVM), Red neuronal multicapa hacia atrás (BP).

La metodología SVM es una generalización de las redes neuronales utilizadas en la resolución de problemas no lineales, se diferencia de las redes en la minimización del error, ya que en su planteamiento se disminuyen los problemas de sobre estimación, que es uno de los principales inconvenientes que se encuentran cuando se trabaja con redes neuronales. Por otro lado una desventaja de SVM es que no existe una metodología establecida para obtener resultados óptimos a través de esta técnica.

Procedimiento y resultados: El artículo desea comparar dos metodologías las redes neuronales hacia atrás y las maquinas de vectores de soporte, se proponen dos modelos de redes neuronales BP-1, BP-2 y el modelo de SVM.

Tabla 3-2. Resultados (Cao & Tay, 2001)

	BP-1		BP-2		SVM	
	G-1	G-2	G-1	G-2	G-1	G-2
NMSE	1.4446	2.1592	1.6937	4.42	1.4383	1.6762
MAE	0.3496	0.4353	0.3726	0.6347	0.3403	0.3706
DA	49.24	41.23	49.75	39.2	47.72	46.23
CU	51.08	41.3	50	40.21	46.21	47.82
CD	47.66	40.47	49.53	38.31	48.6	44.86

Conclusiones: No existen aportes adicionales en el desempeño del modelo al tomar las variables consideradas por el grupo G-2, en el artículo se menciona que la metodología que arroja mejores resultados es SMV si se tiene en cuenta los datos de entrenamiento, sin embargo en los datos de prueba los 3 modelos arrojan resultados similares para el grupo de datos G-1.

3.3 Learning to trade via direct reinforcement (Moody & Saffell, 2001)

Descripción de los datos: Mensual, estudia el periodo comprendido entre 01/1950 y 12/1994, los primeros 20 años se consideran periodo de entrenamiento, y los 25 siguientes como periodo de prueba o validación.

Metodología de pronóstico: Métodos de refuerzo directo

Esta metodología recurre a la programación para mejorar las decisiones de inversión, la metodología propuesta es la aplicación del algoritmo de aprendizaje recurrente, y la ventaja que ofrece frente a las otras metodologías es que elimina la necesidad de formulación de modelos matemáticos complejos, también se utiliza el algoritmo Q-learning para comparar con otras propuestas de programación dinámica.

Procedimiento y resultados: El artículo propone dos modelos uno con el algoritmo de aprendizaje recurrente (RRL) y el otro con el algoritmo Q-learning, compara estos dos modelos con la estrategia Buy and hold que significa comprar un activo y conservarlo. Estas comparaciones se realizan con el objetivo de demostrar si los modelos propuestos son capaces de batir al mercado.

Tabla 3-3. Resultados (Moody & Saffell, 2001)

	Rentabilidad	Sharpe ratio
Buy and hold	1348%	0.34
Q trader	3359%	0.63
RRL trader	5860%	0.83

Conclusiones: El modelo RRL trader es la mejor alternativa comparado con las opciones propuestas. Esto deja a la programación dinámica como una alternativa más simple que los modelos de pronóstico tradicionales y sugieren al algoritmo RRL como una herramienta apropiada para operaciones en bolsa.

3.4 Applying rough sets to market timing decisions (Shen & Loh, 2004)

Descripción de los datos: Periodo comprendido entre 04/01/1988 y 26/07/1999. Donde el periodo de entrenamiento comprende entre 04/01/88 y 30/12/98, y el periodo de validación entre 04/01/99 y 26/07/99. El conjunto de datos de entrenamiento es discretizado utilizando un algoritmo de la distribución chi cuadrado.

Metodología utilizada: Minería de datos Rough Sets

Procedimiento y resultados: El modelo propuesto recoge como datos de entrada únicamente el índice S&P 500, con esta serie de datos hace un cálculo de los indicadores técnicos MACD, tasa de cambio (ROC), oscilador estocástico, índice de fuerza relativa (RSI), regresión lineal. La teoría DOW sugiere que existen momentos de sobreventa y sobrecompra, que es cuando los indicadores técnicos se encuentran en sus niveles superiores e inferiores, a partir de la programación de un conjunto de reglas, el modelo genera órdenes de compra y venta de activos. Se compara la estrategia Buy and hold con el modelo propuesto y se encuentra que el modelo SMV presenta mayor.

Tabla 3-4. Resultados (Shen & Loh, 2004)

	Utilidad neta	Sharpe ratio
Buy and hold	163	
SMV	539	1.31

Conclusiones: Los resultados del artículo no son muy concluyentes pero muestran a la programación de reglas utilizando los Rough Sets como una alternativa a los modelos de regresión lineal tradicionales y a las redes neuronales, en donde los Rough Sets permiten

cumplir el objetivo de un modelo de pronóstico de índice bursátil que es la rentabilidad y gestión del riesgo.

3.5 The adaptive selection of financial and economic variables for use with artificial neural networks (Thawornwong & Enke, 2004)

Descripción de los datos: Mensual, periodo comprendido entre 03/1976 y 12/1999, de un conjunto de variables macroeconómicas, estas variables son seleccionadas mediante un proceso que involucra un algoritmo de árboles de decisiones. Las variables seleccionadas por el algoritmo y que actuarán como variables explicativas del modelo son: serie histórica índice S&P500, bonos y letras del tesoro, calificación de bonos Aaa y Baa, producción industrial, inflación, índice de precios al productor y oferta monetaria.

Metodología: Red neuronal hacia adelante, modelo de regresión lineal.

Procedimiento y resultados: Se toman 3 modelos de redes neuronales todas hacia adelante la diferencia entre un modelo y otro es el cambio en las funciones de optimización que van a variar la decisión del modelo, el primero utiliza el RMSE promedio, el segundo el mínimo RMSE, y el tercero una red neuronal probabilística, además se compara con otros modelos y propuestas como caminata aleatoria (RW) que consiste en comprar acciones unas veces y comprar bonos otra.

Tabla 3-5. Resultados (Thawornwong & Enke, 2004)

	Sign	Rt Mensual	Volatilidad	Sharpe ratio
11/1992–08/1994				
Red RMSE promedio	0.6818	0.8	1.7	0.32
Red RMSE mínimo	0.6818	0.91	1.22	0.54
Red probabilística	0.8182	1.1	1.29	0.66
Regresión lineal	0.5455	0.69	1.76	0.25
Buy and hold	*	0.61	2.32	0.16

	Sign	Rt Mensual	Volatilidad	Sharpe ratio
RW	*	0.34	1.76	0.05
T-bill	*	0.25	*	*
09/1994–06/1996				
Red RMSE promedio	0.7727	1.67	1.72	0.73
Red RMSE mínimo	0.6818	1.57	1.74	0.66
Red probabilística	0.7727	1.6	2.07	0.57
Regresión lineal	0.5	1.04	1.51	0.41
Buy and hold	*	1.6	2.07	0.57
RW	*	1.15	1.93	0.38
T-bill	*	0.42	*	*
07/1996–04/1998				
Red RMSE promedio	0.7273	2.61	3.62	0.6
Red RMSE mínimo	0.8182	2.84	3.6	0.67
Red probabilística	0.7727	2.41	4.19	0.47
Regresión lineal	0.5455	1.6	3.72	0.32
Buy and hold	*	2.41	4.19	0.47
RW	*	1.65	3.73	0.33
T-bill	*	0.42	*	*
05/1998–12/1999				
Red RMSE promedio	0.65	1.53	5	0.23
Red RMSE mínimo	0.7	1.78	5.1	0.28
Red probabilística	0.6	1.54	5.35	0.22
Regresión lineal	0.5	0.82	4.45	0.1
Buy and hold	*	1.54	5.35	0.22
RW	*	1.31	3.23	0.29
T-bill	*	0.37	*	*
11/1992–12/1999				
Red RMSE promedio	0.708	1.65	3.27	0.39
Red RMSE mínimo	0.7205	1.78	3.26	0.43
Red probabilística	0.7409	1.66	3.52	0.37
Regresión lineal	0.5228	1.04	3.05	0.22

	Sign	Rt Mensual	Volatilidad	Sharpe ratio
Buy and hold	*	1.54	3.68	0.32
RW	*	1.11	2.77	0.27
T-bill	*	0.37	*	*

Las redes neuronales siempre ofrecen una tasa de acierto y rentabilidad mayor que los otros modelos, el Sharpe ratio arroja resultados mixtos, sin embargo estas diferencias no son tan grandes como para considerar las redes un modelo con riesgo excesivo.

Conclusiones: Las variables macroeconómicas como datos de entrada si proporcionan información adicional para el pronóstico del índice S&P500, las redes neuronales se convierten entonces en una metodología apropiada porque para todos los periodos de tiempo evaluados supera en rentabilidad la estrategia Buy and Hold.

3.6 Stochastic volatility forecasting and risk management (Sadorsky, 2005)

Descripción de los datos: Mensual, un periodo de veinte años que comprenden entre 01/1984 a 12/2003. Los datos se transforman en rendimientos diarios utilizando la función logaritmo natural y multiplicando el resultado por cien.

Metodología: Caminata aleatoria (RW), media histórica (HM), media móvil 20-60-180 periodos (MA 20-60-180), suavizamiento exponencial (ES), media móvil exponencial 20-60-180 (ES 20-60-180), regresión lineal (RL), modelo autorregresivo de orden 5 (AR 5), modelo de volatilidad estocástica 1 y 2 (SV 1-2).

Procedimiento y resultados: Este es un artículo que recoge las principales metodologías con el fin de determinar cuál de estas ofrece resultados más precisos, tiene modelos simples desde medias móviles, caminatas aleatorias.

Tabla 3-6. Resultados (Sadorsky, 2005)

	MSE	MAE	MPE	MAPE	Theil U
RW	0.287	0.432	-0.008	-0.099	
HM	0.268	0.413	-0.019	-0.097	0.932
MA 20	0.174	0.33	-0.009	-0.076	0.607
MA 60	0.189	0.345	-0.009	-0.08	0.658
MA 180	0.205	0.359	-0.01	-0.084	0.713
ES	0.17	0.325	-0.009	-0.075	0.593
ES 20	0.178	0.334	-0.007	-0.077	0.619
ES 60	0.173	0.33	-0.009	-0.076	0.603
ES 180	0.171	0.327	-0.008	-0.076	0.595
LR	0.214	0.37	-0.014	-0.086	0.744
AR 5	0.175	0.332	-0.012	-0.077	0.609
SV 1	0.226	0.38	-0.017	-0.088	0.787
SV 2	0.195	0.339	-0.01	-0.078	0.679

La Tabla 3-6 muestra que los modelos de mejor desempeño según los diferentes criterios son: para MSE ES, para MAE ES, para MPE RW, para MAPE ES 20, y finalmente para Theil U ES, los criterios Diebold y Mariano y su modificada no se reportan debido a que no aportan información relevante sobre los modelos evaluados.

Conclusiones: La conclusión más relevante que arroja este artículo es que el modelo ES arroja los pronósticos más precisos, se debe de tener en cuenta que este es un modelo muy simple que supera en resultado a modelos más estructurados matemáticamente y complejos como lo son AR5 o SV 1-2.

3.7 Trading Rules from Forecasting the Collapse of Speculative Bubbles for the S&P 500 Composite Index (Brooks & Katsaris, 2005)

Descripción de los datos: Mensual, periodo comprendido entre 01/1888 y 01/2003, el periodo utilizado para desarrollar el modelo va desde 01/1888 hasta 12/1945, el periodo comprendido entre 1946 y 2003 se utiliza como periodo de prueba o validación.

Metodología: Cambio de régimen (RS). Este es un método que es utilizado para detectar cambios importantes en las series de tiempo, donde se afirma que después de estos cambios la serie pierde memoria. Los modelos de RS se utilizan en finanzas para detectar cambios atribuidos a las burbujas financieras.

Procedimiento y resultados: Se utilizan varias propuestas que aplican el modelo de RS,

Tabla 3-7.Resultados (Brooks & Katsaris, 2005)

	Rt mensual	Riqueza	Sharpe ratio
Buy and hold	0.60%	38.36	0.1541
Libre de riesgo	0.05%	1.44	
RS aumentado	0.51%	24.79	0.2172
RS Van Norden Schaller	0.36%	9.7	0.1711
RS EC B V	0.38%	10.97	0.1306
RS EC	0.28%	5.36	0.0869

En la Tabla 3-7 se muestra que no resulta útil utilizar el modelos de cambio de régimen ya que la estrategia Buy and hold que no requiere ningún cálculo supera en rentabilidad y riqueza a los modelos de propuestos, el Sharpe Ratio resulta superior en el modelo RS aumentado y en RS Van Norden Schaller, sin embargo esto ocurre porque el dinero en los modelos de RS no se encuentran en el mercado todo el tiempo.

Conclusiones: Los modelos de cambio de régimen no son útiles para predecir el índice S&P500 ya que no pueden superar la estrategia pasiva Buy and hold.

3.8 SV mixture models with application to S&P 500 index returns (Durham, 2007)

Descripción de los datos: Diario, la serie de datos estudiada inicia el 23/06/1980 finaliza el 02/09/2002, para un total de 5.616 datos, los cuales se someten a un suavizamiento mediante la función logaritmo natural

Metodología: Modelos estándar de volatilidad estocástica (SV)

Se describen varios modelos de volatilidad estocástica en donde se varía la componente del error en cada uno de los modelos propuestos, con el fin de determinar cuál de los modelos captura mejor el comportamiento del índice S&P 500.

Procedimiento y resultados:

No es rechazada la hipótesis nula de independencia de datos mediante la prueba Box Pierce para los niveles de significancia convencionales con $L=20$ y $L=25$, se rechaza la hipótesis nula de normalidad mediante la prueba Jarque Bera.

Los resultados del pronóstico se encuentran en gráficos (del artículo original) y no se pueden determinar valores exactos, sin embargo gráficamente se encuentra que los modelos de volatilidad estocástica más sencillos capturan el comportamiento de la volatilidad del índice de la misma manera que los modelos más complejos.

Conclusiones: Los modelos son útiles para describir más no para predecir el comportamiento del índice, ninguno de los modelos captura adecuadamente el comportamiento de la volatilidad.

3.9 Forecasting stock index volatility: comparing implied volatility and the intraday high-low price range (Corrado & Truong, 2007)

Descripción de los datos: Diario, periodo comprendido entre 01/1990 y 01/2006, en donde se toma los retornos diarios, además se toman las medidas del cuadrado de logaritmo natural de la división entre el máximo y el mínimo diario, finalmente se toma la volatilidad implícita de las opciones del S&P 500.

Metodología: Modelo generalizado autorregresivo heterocedástico (GARCH)

Se toman 4 modelos, el primero toma la serie de datos del S&P 500, el segundo adiciona la volatilidad implícita de las opciones, el tercero el rango entre el máximo y mínimo diario, finalmente el cuarto toma todas las series de datos disponibles. Se desea pronosticar para un intervalo de 1, 10, y 20 días.

Procedimiento y resultados:

Tabla 3-8. Resultados (Corrado & Truong, 2007)

	1 día			10 días			20 días		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
Modelo 1	2.522	1.166	0.148	11.133	6.006	0.331	20.15	11.491	0.332
Modelo 2	2.485	1.153	0.171	11.281	6.019	0.348	19.981	11.361	0.343
Modelo 3	2.481	1.153	0.174	10.577	5.717	0.412	19.302	11.111	0.387
Modelo 4	2.471	1.149	0.181	10.624	5.674	0.416	19.041	10.776	0.403

El modelo que arroja los mejores resultados es el modelo 4 para todos los horizontes, bajo todos los criterios excepto para el RMSE que el modelo 3 lo supera por un pequeño margen. El rango entre el máximo y mínimo aporta más información para pronosticar que la volatilidad implícita, esto se puede ver en el criterio R² que para los tres horizontes de tiempo es mayor para el modelo 3 que para el modelo 2.

Conclusiones: La volatilidad implícita, el rango entre el máximo y mínimo diario si aportan información adicional para la predicción del índice S&P 500.

3.10 Does implied volatility provide any information beyond that captured in model-based volatility forecasts? (Becker et al., 2007)

Descripción de los datos: Diario, desde 02/01/1990 hasta 17/10/2003, para un total de 3481 observaciones. Se tiene en cuenta la volatilidad implícita de las opciones put y call sobre el índice S&P 500 este índice se denomina VIX, dicho índice muestra la volatilidad implícita de las opciones sobre el índice para un periodo de 22 días de operación. Debido a la manera como está compuesto el índice VIX el periodo de pronóstico solo puede ser de máximo 22 días.

Metodología: GARCH, modelo autorregresivo de media móvil ARMA, modelo autorregresivo fraccionalmente integrado (ARFIMA), modelo de volatilidad estocástica (SV). Se toman los cuatro modelos y se compara cual de los modelos es el que tiene más relación con la serie de datos de la volatilidad realizada del índice (RV) y el índice VIX.

Procedimiento y resultados: De los cuatro modelos el que predice con mayor exactitud la volatilidad del índice S&P 500 es el modelo ARMA, y el que predice mejor el índice VIX es el modelo ARFIMA (Ver Tabla 3-9).

Tabla 3-9. Resultados (Becker et al., 2007)

	GARCH	ARMA	ARFIMA	SV
Corr (.,RV 22)	0.601	0.683	0.6751	0.6214
Corr (.,VIX)	0.8309	0.8529	0.8539	0.7615

Además de los modelos se hace una prueba de la relación existente entre el índice S&P 500 y el índice VIX, con una prueba F y una prueba Hotelling, para diferentes horizontes de tiempo de 1, 5, 10, 15 y 22 días, donde la hipótesis nula es que el VIX no aporta

información adicional a los modelos de pronóstico, hipótesis que según los estadísticos de prueba no puede ser rechazada.

Conclusiones: El índice VIX o la volatilidad implícita de las opciones sobre el índice S&P 500 no ofrece información adicional que permita predecir el comportamiento del índice con mayor precisión que con los modelos que no tienen en cuenta esta medida.

3.11 Forecasting volatility by means of threshold models

(Acosta, Muñoz, & Márquez, 2007)

Descripción de los datos: Diario, periodo comprendido entre 10/10/1990 y 03/12/2003, para un total de 3085 observaciones, las primera 2785 son utilizadas para estimar los parámetros de los modelos y las siguientes 300 para validar el modelo.

Metodología: GARCH, SV y una variación de cada uno de los modelos SETAR-TGARCH y SETAR-THSV, estos últimos son modelos de frontera.

Las variaciones de los modelos permiten capturar las asimetrías que ocurren en la volatilidad si se presenta una tendencia alcista o bajista, las propuestas de las variaciones a los modelos GARCH y SV, las variaciones en los modelo asumen que la volatilidad se comporta de diferente manera si presenta signo positivo o negativo.

Procedimiento y resultados: Inicialmente se hace una simulación monte carlo con un número de modelos mayor, con los resultados obtenidos se seleccionan los cuatro modelos mencionados en la metodología.

La prueba Diebold y Mariano (Tabla 3-10) arroja que los modelos tienen capacidad de pronóstico diferente, esta capacidad se comprueba con el estadístico MSE, que indica al modelo SV como el más preciso.

Conclusiones: El modelo que arroja mejores resultados es el más simple, el modelo SV, además resulta no ser cierto el supuesto que la volatilidad se comporta de una manera diferente dependiendo la dirección.

	Diebold y Mariano			MSE
	SV	SETAR-T GARCH	SETAR-THSV	
GARCH	7.28	-3.76	2.82	3.364
SV	*	8.57	-2.11	2.216
SETAR-TGARCH	*	*	6.34	3.308
SETAR-THSC	*	*	*	3.47

Tabla 3-10. Resultados (Acosta et al., 2007)

3.12 Are combination forecasts of S&P 500 volatility statistically superior? (Becker & Clements, 2008)

Descripción de los datos: Diario, desde 02/01/90 hasta 17/10/03, para un total de 3481 observaciones. Los datos de entrada corresponden al índice VIX que es compuesto por la volatilidad implícita de las opciones put y call sobre el índice S&P 500, este índice muestra la volatilidad implícita de las opciones para un periodo de 22 días de operación. Debido a la manera como está compuesto el índice VIX el periodo de pronóstico solo puede ser de máximo 22 días.

Metodología: GARCH, SV, RV, ARMA, ARFIMA

Procedimiento y resultados: El principal objetivo del artículo es determinar si los modelos propuestos operan mejor como una combinación de estos u operan mejor de manera independiente, también se cambia la serie de datos del modelo para comparar la serie de los rendimientos del S&P500 por una serie de promedios (U) y una serie resultado de una regresión lineal (R).

La Tabla 3-11 muestra el ranking según el MSE los modelos de mayor a menor precisión, según ambos criterios los modelos que ofrecen mejores resultados son el modelo ARMA y ARFIMA en los modelos sencillos, cuando se hace una mezcla de modelos ya sea por promedios o regresión el desempeño de los modelos mejora un poco, sin embargo los dos mejores modelos sencillos combinados por medio de una regresión lineal también ocupan el primer lugar en cuanto a precisión este modelo es el ARMA+ ARFIMA (R).

Tabla 3-11. Resultados (Becker & Clements, 2008)

	MSE	QLIKE
ARMA+ARFIMA (R)	1.477	3043.8
(ARMA+ARFIMA+SVRV+VIX)		
R	1.502	3042.5
ALL R	1.509	5796.9
ALLMBF R	1.549	3266.1
ARMA+ARFIMA U	1.648	3161.6
ARMA	1.659	3174.3
ARFIMA	1.667	3159.2
GARCHRV	1.952	3266.6
GJRRV	2.161	3700.7
GJRRVG	2.404	3278.7
VIX	2.525	3253
GARCH	2.575	3472.5
SV	2.73	3923
GJR	2.857	3575.7
SVRV	4.543	3222.5

Conclusiones: Nuevamente como en el artículo publicado en el 2007 por el mismo autor, los modelos más sencillos resultan ser los más precisos, el hallazgo interesante en este artículo resulta ser que los dos modelos en conjunto generan una sinergia mejorando la capacidad de pronóstico cuando se combinan.

3.13 A hybrid derivative trading system based on volatility and return forecasting (Enke & Amornwattana, 2008)

Descripción de los datos: Diario, periodo comprendido entre 01/01/2000 a 31/12/2004, el periodo de entrenamiento corresponde a los primeros tres años y el periodo de prueba o validación los últimos dos, el estudio inicialmente toma 22 variables para determinar cuáles de estas entrarán al modelo. Las variables que resultaron más relevantes son la volatilidad diaria, volatilidad implícita y el los retornos diarios del índice, posteriormente los datos son normalizados antes de ingresarlos a la red neuronal.

Metodología: Red neuronal de regresión generalizada (GRNN), modelo de regresión lineal (MLR), C4.5

La red neuronal es un modelo que pretende operar en el mercado con reglas aprendidas en el periodo de entrenamiento que describen el comportamiento del índice con el premisa de lo que ocurre en el pasado volverá a ocurrir en el futuro.

El modelo de regresión lineal y el C4.5 son modelos para comparar pero no son el objeto de estudio del artículo, el modelo C4.5 es un modelo de arboles binomiales que simplemente da una sugerencia de compra o venta y no arroja un número exacto de pronóstico.

Procedimientos y resultados: El horizonte de tiempo que se desea pronosticar es de 5 días, debido a que las diferencias entre el bid y al ask, pueden llevar a errores en las decisiones tomadas por los resultados arrojados por el modelo pronosticando un horizonte de un día.

Tabla 3-12. Resultados (Enke & Amornwattana, 2008)

	GRNN		MLR		C 4.5	
	Diario	Semanal	Diario	Semanal	Diario	Semanal
MSE	1.2556	1.541	1.707	5.3053	*	*
MAE	0.9941	0.9606	0.9652	1.6828	*	*
%	55.65	52.31	53.98	57.58	94.1	96.2

La red neuronal resulta mejor en promedio para pronosticar que una regresión lineal, esto quiere decir que el índice tiene un comportamiento no lineal y se describe mejor por medio de las redes neuronales, cuando se compara su precisión con el modelo C4.5 la red neuronal se queda corta.

Conclusiones: El modelo de red neuronal propuesto capta mejor el comportamiento del índice S&P500 que un modelo de regresión tradicional, sin embargo no es lo suficientemente preciso para afirmar que es un modelo con alta capacidad predictiva.

3.14 Deterministic regression model and visual basic code for optimal forecasting of financial time series (Balbás, Balbás, Galperin, & Galperin, 2008)

Descripción de los datos: Diario, periodo comprendido entre 06/01/2006 y 27/03/2008,

Metodología: Regresión lineal ajustada con visual basic.

Se trata de un método de regresión lineal que por medio de un algoritmo en visual basic mejora la función por iteraciones con el objetivo de obtener un grado de precisión mayor que una regresión lineal simple.

Procedimiento y resultados: Los resultados se reportan de una manera gráfica y no es posible determinar con exactitud la precisión del modelo.

Conclusiones: Mediante este procedimiento aplicado a la regresión lineal se pueden obtener mejores resultados que con los modelos lineales tradicionales.

3.15 Intra-daily information of range-based volatility for MEM-GARCH (Lam & Ng, 2009)

Descripción de los datos: Diario, periodo comprendido entre 02/01/98 y 19/02/06, para un total de 2045 datos, los primeros 1800 datos son utilizados para la formulación del modelo y los 245 restantes para probar los resultados de este.

Metodología: GARCH y MEM-GARCH

El primero es el modelo GARCH tradicional mientras el segundo es una variación del modelo que consiste en un cambio en la volatilidad, el objetivo es demostrar que el cambio genera un aumento en la capacidad predictiva comparado con el modelo tradicional.

Procedimiento y resultados: Solo se reportan resultados gráficos los cuales no se puede determinar con exactitud los valores obtenidos.

Conclusiones: Se formula un modelo exitoso que incorpora las volatilidades dentro del día y por la inclusión de estas volatilidades se obtiene un grado de precisión mayor con un modelo MEM-GARCH que con un modelo GARCH tradicional.

3.16 The information contents of VIX index and range-based volatility on volatility forecasting performance of S&P 500 (Hung et al., 2009)

Descripción de los datos: Diario, periodo comprendido entre 02/01/2001 y 31/12/2007, para un total de 1758 observaciones, donde los 1250 observaciones se utilizan en el planteamiento del modelo y las restantes para probar los resultados del modelo, se realiza un test Jarque Bera en donde se comprueba la no normalidad de la serie de los retornos, adicionalmente las pruebas Q like y LM muestran una fuerte dependencia lineal. También se encuentra que la serie de los retornos del índice S&P500 exhibe un comportamiento asimétrico apto para los modelos GARCH.

Metodología: Varios modelos GARCH en donde se varia la componente de la volatilidad de acuerdo a propuestas de diferentes autores, para un total de 5 modelos.

Procedimiento y resultados:

El modelo GRJ-VIX ocupa el primer lugar según los criterios MSE, MAE, MME(O), y VAR, el segundo lugar según el criterio MME(U) lo que lo convierte en el modelo con más precisión de los propuestos, el segundo puesto lo ocupa el modelo GJR-GK ocupando el

segundo lugar en todos los criterios menos en el criterio MME(U) donde el primer lugar(Ver Tabla 3-13).

Tabla 3-13. Resultados (Hung et al., 2009)

	MSE	MAE	MME (O)	MME (U)	VAR
GJR	0.378	0.396	0.142	0.074	0.157
GJR-VIX	0.351	0.369	0.133	0.072	0.154
GJR-PK	0.383	0.379	0.137	0.072	0.158
GJR-GK	0.374	0.374	0.137	0.071	0.157
GJR-RS	0.387	0.379	0.137	0.072	0.156

Conclusiones: El modelo que incluye el índice VIX es el más preciso lo que quiere decir que la volatilidad implícita si aporta información adicional, cuando se busca predecir el índice S&P 500.

3.17 Evolving least squares support vector machines for stock market trend mining (Yu et al., 2009)

Descripción de los datos: Mensual, periodo comprendido entre 01/1926 y 12/2005, en donde los primeros 64 años se utilizan para desarrollar el modelo y los restantes 16 para hacer pruebas y validación de resultados. A partir de esta serie de datos, se calculan los siguientes indicadores técnicos: Oscilador estocástico, media del oscilador estocástico, oscilador estocástico lento, tasa de cambio, momentum, varianza móvil, varianza promedio, media exponencial, radio de varianza media, media móvil de convergencia y divergencia, oscilador de acumulación y distribución, entre otros.

Metodología: LSSMV utilizando el algoritmo de mínimos cuadrados evolucionado, LSSMV utilizando otros algoritmos de mínimos cuadrados, SMV tradicional, red neuronal hacia atrás (BP), ARIMA y análisis discriminante (LDA).

Tabla 3-14. Resultados (Yu et al., 2009)

Modelo	Hit ratio
ARIMA	55.78%
LDA	61.43%
BP	67.46%
SVM	72.61%
LSSVM poly	70.43%
LSSVM rbf	75.45%
LSSVMsig	74.78%
LSSVMmix	77.86%
LSSVMgafs	82.43%
LSSVMgapo	82.58%
LSSVMevol	82.66%

Procedimiento y resultados: En la Tabla 3-14 se puede observar que el modelo de maquinas de soporte de vectores utilizando el algoritmo de mínimos cuadrados evolucionado resulta ser el mejor modelo para predecir el comportamiento del índice S&P 500 entre los evaluados. Se puede decir que es considerablemente más asertivo que los modelos tradicionales como ARIMA y LDA, y supera también a las redes neuronales que son los modelos que preceden a las maquinas de soporte de vectores.

Conclusiones: El modelo de SVM que utiliza el algoritmo de mínimos cuadrados evolucionados resulta ser un modelo apropiado para predecir series de tiempo financieras, sin recurrir a otras fuentes de información diferentes a la serie.

3.18 Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic (Pooter, 2009)

Descripción de los datos: Diario, periodo comprendido entre 03/12/1994 y 29/12/2006, utilizando la volatilidad del día del precio de los futuros del índice S&P 500, donde el periodo de evaluación del modelo es el periodo comprendido entre 01/01/98 y

29/12/2006. También involucra en los datos, los cambios irregulares de nivel del índice, nivel de apalancamiento que es la relación entre el anuncio de las noticias y la volatilidad del índice y finalmente las noticias .

Metodología: ARIMA, ARFIMA, modelo heterogéneo autorregresivo (HAR).

Procedimiento y resultados: Cada modelo tendrá en cuenta las variables mencionadas en la descripción de los datos, con el fin de determinar cual serie de datos, y cual modelo es el que tiene mayor capacidad predictiva en el índice S&P500

Tabla 3-15. Resultados (Pooter, 2009)

	ME	MSE	MAE	HMSE	R ²
GJR (1,1) DA	-0.047	0.133	0.243	0.096	0.575
ARFI	0.002	0.122	0.222	0.094	0.6
ARFI-D	0.001	0.121	0.22	0.8	0.603
ARFI-DA	0.001	0.12	0.22	0.078	0.606
ARFI-DAR	0	0.12	0.219	0.075	0.606
ARFI-DARL	-0.001	0.117	0.217	0.075	0.606
ARFI-DAXRL	-0.018	0.115	0.216	0.074	0.617
SPARFI-DAXRL	-0.026	0.116	0.22	0.069	0.626
HAR-DAXRL	0.01	0.116	0.213	0.069	0.622
AR(1)-DAXRL	0.031	0.142	0.241	0.075	0.536
AR(5)-DAXRL	0.004	0.115	0.213	0.107	0.625
AR(22)-DAXRL	-0.005	0.114	0.214	0.074	0.629

Nota: El modelos DAXRL se refiere al modelo con todas las variables, excepto en el modelo ARFIMA que no involucra los cambios irregulares de nivel, sin embargo el modelo SPARFIMA-DAXRL estudia todas las variables, DAR a los modelos que involucran el día y los anuncios de noticias, los modelos DA se refiere que involucran a los anuncios de noticias, D involucra la variable día de la semana.

En promedio el modelo que arroja mejores resultados según los diferentes criterios es el modelo ARIMA de orden 22 que involucra todas las variables.

Conclusiones: Los efectos de apalancamiento, el día de la semana, las noticias, y la información histórica del índice contienen información relevante que permite el pronóstico del mismo, donde el modelo AR (22) resulta ser el más preciso de los modelos evaluados.

3.19 Development and performance evaluation of FLANN based model for forecasting of stock markets(Majhi, Panda, & Sahoo, 2009)

Descripción de los datos: Diario, periodo comprendido entre 01/1994 y 10/2006, la serie incluye los precios de apertura y cierre de la jornada. De la serie de datos se obtienen los siguiente indicadores técnicos: media móvil, media móvil exponencial, oscilador estocástico, oscilador de acumulación/distribución, volumen de equilibrio, Williams ratio, índice de fuerza relativo, tasa de cambio del precio, aceleración del precio de cierre, aceleración de alza de precios.

Metodología: Red neuronal FLANN, consiste en una red neuronal de conexión funcional y selecciona un algoritmo de mínimos cuadrados de recursivos.

Procedimiento y resultados: Se corre el modelo con diferentes combinaciones de indicadores técnicos de los cuales las combinaciones más sobresalientes excluyen la media móvil, balance en el volumen.

Tabla 3-16. Resultados (Majhi et al. 2009)

	MAPE	Tamaño de muestra	Otros indicadores fuera
FLANN 1	2.95	658	*
FLANN 2	2.66	658	Oscilador estocástico
FLANN 3	2.22	60	Oscilador estocástico
FLANN 4	2.09	60	Williams ratio

De la Tabla 3-16 se puede observar que la red neuronal presenta una mayor precisión para predecir el índice S&P 500 cuando toma 60 datos y excluye los indicadores técnicos, media móvil, balance en el volumen, oscilador estocástico y Williams ratio.

Conclusiones: El modelo FLANN que predice mejor el índice S&P 500 es el que toma una muestra de solo 60 días y excluye cuatro indicadores técnicos.

3.20 Non linear prediction of the Standard & Poor's 500 and the hang seng index under a dynamic increasing simple (Kumar, 2009)

Descripción de los datos: Diario, periodo comprendido entre 06/11/1928 y 31/12/2008, el periodo entre 1928 y 2005, es utilizado para plantear el modelo y el periodo entre 2006 y 2008 para realizar las pruebas de desempeño. El horizonte de tiempo de predicción es de un mes.

Metodología: ARIMA, Red neuronal hacia adelante que utiliza mínimos cuadrados residuales como algoritmo de desempeño para evaluar el error.

Procedimiento y resultados: Se evalúa el modelo bajo diferentes criterios de desempeño como lo son DA, Sharpe ratio; se hace una evaluación de 36 meses la cual se resume en datos anuales con el fin de facilitar la lectura de los resultados, también se reporta el Max drawdown del índice S&P500 que pretende comparar si los niveles de riesgo del modelo son superiores a los del mercado.

Tabla 3-17. Resultados (Kumar, 2009)

	2006		2007		2008	
	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
Sign	56.91%	53.87%	56.59%	45.40%	53.76%	49.23%
Rentabilidad	18.42%	4.03%	12.21%	-35.71%	55.48%	17.40%
Sharpe ratio	2.4	1.88	3.8	1.25	2.55	1.9
Volatilidad anual	9.90%	12.00%	15.60%	16.20%	41.10%	47.80%

Max drawdown	-5.20%	-7.13%	-7.90%	9.20%	32.10%	-40.53%
Max drawdown S&P 500	-7.70%		-10.10%		-48.80%	

La Red neuronal gana en todos los criterios resultando ser un modelo superior al modelo ARIMA (Ver Tabla 3-17), es muy importante resaltar que el modelo de Red neuronal hacia adelante aunque ofrece mayor rentabilidad que comprar el índice y dejar pasar el tiempo, los niveles de riesgo son menores.

3.21 Predicting the bear stock market: Macroeconomic variables as leading indicators (S. S. Chen, 2009)

Descripción de los datos: Mensual, periodo comprendido entre 02/1957 y 12/2007, también para el mismo periodo se tienen la serie de datos de diferencia en tasas (Yield spread) para varios periodos, tasa de inflación, producción industrial, M1,M2, cambios en la tasa de empleo, cambios en las tasas de bonos federales, cambios en la tasa de cambio, tasas de los fondos federales y cambios en la deuda pública.

Metodología: SV Markov Switching (SVMS), es un modelo SV en donde la volatilidad se mantiene constante y es acompañada de un término que cambia en el tiempo.

Procedimiento y resultados: Se realiza un estudio de las variables que acompañan al modelo, se hace una comparación con la estrategia Buy and hold para probar la utilidad del modelo y si es capaz de superar al mercado.

En la Tabla 3-18 se puede observar que la mayoría de las variables macroeconómicas tenidas en cuenta en cada uno de los modelos SVMS, tienen información relevante que permite predecir el comportamiento del índice S&P500 en donde las variables macroeconómicas más relevantes son la inflación y los diferenciales en los títulos del tesoro de Estados Unidos.

Conclusiones: Los modelos SVMS que tienen en cuenta variables macroeconómicas captan el comportamiento del índice S&P 500 permitiendo obtener rentabilidades superiores a una estrategia pasiva de Buy and hold.

	Riqueza	Rt mensual
Buy and hold	18.98	0.6
Spread (3M-10Y)	109.47	0.96
Spread (3M-5Y)	75.77	0.88
M1	27.92	0.68
M2	42.84	0.77
Tasa de inflación	264.95	1.14
Cambio tasa de desempleo	33.09	0.71
Crecimiento producción industrial	30.05	0.6
Cambio tasa de cambio	16.76	0.58
Tasas fondos federales	24.46	0.6
Cambio en niveles deuda pública	19.34	0.58

Tabla 3-18. Resultados (S. S. Chen, 2009)

3.22 Nonlinear dynamics in arbitrage of the S&P 500 index and futures: A threshold error-correction model (B. H. Kim, Chun, & Min, 2010)

Descripción de los datos: Frecuencia 1 minuto, datos de dos meses que corresponden mayo y noviembre de 1993 de los futuros y el índice S&P500, se toman estos meses en específico para permitir la comparación entre el índice y los futuros próximos a vencerse en junio y diciembre respectivamente, además cuentan con una base de datos de 7060 operaciones bursátiles efectuadas en mayo de 1993.

Metodología: Modelo de corrección de error de frontera, modelo de corrección de error.

Ambas metodologías formulan un modelo original con un error dato y este error se le somete a un proceso de reparametrización con el objetivo de obtener un modelo más preciso que el inicial.

Procedimiento y resultados: Se formulan dos modelos el modelo de corrección de error de frontera (TRTECM), y el modelo de corrección de error (ECM), además se hace la comparación del TRTECM con una caminata aleatoria (RW) para observar si es

justificable el uso de este modelo. Como se puede observar en la Tabla 3-19, el modelo TRTECM tiene errores más pequeños que el modelo ECM, resultado que se corrobora en la prueba Diebold y Mariano, también se afirma que el modelo pronostica mejor que el modelo de caminata aleatoria.

Tabla 3-19. Resultados (Kim et al. 2010)

	RMSE		Diebold y Mariano	
	TRTECM	ECM	TRTTECM/RW	TRTTECM/ECM
1 día	0.5894	0.9794	-1.7143	-4.3026
2 días	0.4544	0.9759	-1.9219	-4.5331
3 días	0.3879	0.9592	-3.1071	-4.7958
4 días	0.348	0.9272	-4.5162	-5.0992
5 días	0.3222	0.8884	-5.9437	-5.3937

Conclusiones: Los modelos de corrección de error de frontera pronostican mejor que sus similar que no tiene en cuenta una frontera, igualmente supera al modelo de caminata aleatoria.

3.23 Forecast accuracy and economic gains from bayesian model averaging using time-varying weights (Hoogerheide, Kleijn, Ravazzolo, Dijk, & Verbeek, 2010)

Descripción de los datos: Mensual, periodo comprendido entre 01/66 y 12/08, de las series de tiempo del índice S&P 500 y letras del tesoro de un mes, el diferencial entre los rendimientos del índice y las letras, será la serie utilizada en la modelación, uno de los modelos tiene como entrada las siguientes variables macroeconómicas: dividend yield, letras del tesoro de 3 meses, cambio mensual de las letras de 3 meses, diferencial entre bonos Baa y Aaa, diferencial entre los fondos federales y las letras de 3 meses, inflación bimestral, tasa bimestral de crecimiento en la producción industrial, tasa de crecimiento de M1.

Metodología: Regresión lineal con variables macro, regresión lineal con Indicador de Halloween que consiste comprar en noviembre y vender en mayo, modelo de volatilidad estocástica, modelo de volatilidad estocástica robusto, promedio de modelos bayesianos (BMA), modelo bayesiano por pesos por mínimos cuadrados (LIN), modelo bayesiano por pesos variables en el tiempo (TVW), modelo bayesiano robusto por pesos variables en el tiempo (RTVW), los cálculos de los modelos son para un horizonte de tiempo de 1 mes.

Procedimiento y resultados:

Tabla 3-20. Resultados (Hoogerheide et al.2010)

	LI	HI	SV	RSV	BMA	LIN	TVW	RTVW
RMSE	4.618	4.478	4.509	4.47	4.5	4.514	4.484	4.485
Sign ratio	0.527	0.549	0.614	0.598	0.587	0.61	0.602	0.598

Por el número de predicciones correctas o sign ratio el modelo más preciso es el modelo RSV, mientras que por el criterio RMSE es SV (Ver Tabla 3-20).

Conclusiones: Los modelos probados en este artículo demuestran que predecir el mercado resulta muy difícil y ninguna de las propuestas se puede considerar un modelo que captura de una manera adecuada el comportamiento del índice S&P 500.

3.24 Predicting the equity premium with dividend ratios:

Reconciling the evidence (Kellard, Nankervis, & Papadimitriou, 2010)

Descripción de los datos: Mensual, periodo comprendido entre 04/1975 y 11/2009, además de utilizar los rendimientos del índice, también se recolecta la información del dividend yield (DY), relación precio dividendo (DP), letras del tesoro de 3 meses.

Metodología: Modelo de promedios móviles con el S&P500 (MM), modelo de regresión lineal univariado con el DY, y modelo de regresión lineal con la DP.

Procedimiento y resultados: Se corrió el modelo para 5 periodos de tiempo diferentes, buscando identificar si para diversos periodos el comportamiento de los modelos era similar, en la Tabla 3-21 se puede observar que el desempeño del modelo varía notablemente de un periodo a otro. El modelo que arroja mejores resultados según el RMSE y el MAE es el modelo de medias móviles en el periodo 02/1995-02/2000

Tabla 3-21. Resultados (Kellard et al. 2010)

	01/1990-11/2009			01/1990-01/1995			02/1995-02/2000			03/2000-03/2005		
	MM	DP	DY									
RMSE	7.5	7.35	7.35	7.2	6.99	7	5.78	5.82	5.82	9.75	9.63	9.64
MAE	5.74	5.55	5.55	5.57	5.36	5.38	4.42	4.67	4.67	7.72	7.26	7.28
	04/2005-11/2009			03/2000-03/2005			04/2005-11/2009					
	MM	DP	DY	MM	DP	DY	MM	DP	DY			
RMSE	6.61	6.28	6.27	9.75	9.63	9.64	6.61	6.28	6.27			
MAE	5.18	4.84	4.83	7.72	7.26	7.28	5.18	4.84	4.83			

Conclusiones: No es confiable un modelo que arroje diferentes resultados para diferentes periodos de tiempo, por lo cual se concluye que un modelo de regresión lineal con las variables DY y DP como entradas no predicen el comportamiento del índice S&P500.

3.25 Modeling intraday volatility: A new consideration (Chu & Lam, 2011)

Descripción de los datos: Frecuencia 10 min, retornos multiplicados por cien del periodo comprendido entre 13/04/2006 y 12/09/2008, para un total de 22893 observaciones (387 días).

Metodología: GARCH, FIGARCH, HYGARCH.

Procedimiento y resultados: Se hace una evaluación de la capacidad de pronóstico de estos los modelos con el objetivo de determinar dos incognitas, la primera es encontrar cual método de filtrado es más apropiado para desestacionalizar la serie de datos, la

forma flexible de fourier (FFF) o promedios simples, la segunda es responder si es más apropiado tomar la distribución de los retornos diarios como una distribución normal, o t-student, para un horizonte de tiempo de un día.

En la Tabla 3-22 el MSE nos dice que el modelo que mejor se comporta es el FIGARCH normal y el que peor desempeño tiene es el modelo es el GARCH t-distribuido, según el MAE el mejor modelo es el HYARCH normal y el de peor desempeño es el GARCH t-distribuido.

Tabla 3-22. Resultados (Chu & Lam, 2011)

	Desestacionalización por promedios				Desestacionalización por método FFF			
	Normalmente distribuidos		t- distribuidos		Normalmente distribuidos		t- distribuidos	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
GARCH	0.092	0.208	0.095	0.214	0.099	0.216	0.102	0.221
FIGARCH	0.080	0.196	0.084	0.204	0.085	0.202	0.097	0.218
HYGARCH	0.081	0.195	0.089	0.209	0.085	0.201	0.094	0.215

Conclusiones: La manera más apropiada de desestacionalizar los datos es por promedios simples, y asumir que los retornos se comportan como una distribución normal.

3.26 S&P500 Index Direction Forecasting from 1976 to 2010 : A Fuzzy System Approach(Duarte & Brazil, 2011)

Descripción de los datos: Diario, periodo comprendido entre 1970 y 2010, en donde el periodo 1970-1975 se utiliza para elaborar el modelo, en adelante el modelo continúa modificándose pero a la vez se utiliza como prueba.

Metodología: Sistema de lógica difusa.

Procedimiento y resultados: Se procede a correr el modelo con las reglas establecidas en el periodo de entrenamiento, solo se mostrarán los resultados más relevantes (máximo y mínimo de la estrategia Buy and hold y el modelo difuso) con el objetivo de sintetizar ya que el autor muestra los resultados para los 25 años estudiados.

Tabla 3-23. Resultados (Duarte & Brazil, 2011)

	Modelo difuso	Buy and hold	Diferencia
1981	-9.73%	-9.73%	0.00%
1995	24.08%	34.11%	-10.03%
2008	26.70%	-38.49%	65.18%
2009	34.05%	23.45%	10.59%
Promedio	13.04%	7.82%	5.22%

La pérdida máxima del modelo es igual a la de la estrategia Buy and hold, lo que nos muestra que el modelo tiene un riesgo máximo igual que no correr ningún modelo, en 2008 se presenta una situación particular entre el modelo y Buy and hold, esto es debido a que el modelo permite ventas en corto, por esto no se considera relevante este resultados. En promedio el modelo arroja mejores resultados que la estrategia Buy and hold, especialmente en los últimos años.

Conclusiones: Un modelo de lógica difusa arroja es capaz de arrojar una rentabilidad mayor que la del mercado lo que lo convierte en una buena alternativa para predecir el índice S&P 500.

3.27 Modelling and forecasting noisy realized volatility (Asai, McAleer, & Medeiros, 2012)

Descripción de los datos: Diario, periodo comprendido entre 03/01/1996 y 29/04/2007. El modelo utiliza como datos de entrada los retornos diarios.

Metodología: Modelos de dos componentes (MA) que consta de un modelo AR(1) y un ARMA(1,1), ARFIMA (1,d,1), y un modelo HARMA.

Procedimiento y resultados: El artículo pretende resolver si los errores incluidos en la volatilidad realizada que es la tomada por la mayoría de los autores afecta la capacidad de pronóstico de los modelos propuestos.

Para el horizonte de tiempo de 1,5 días el modelo que mejor arroja resultados es el ARFIMA, para el periodo de tiempo de 10 días se presentan diferentes candidatos dependiendo si se tuvo en cuenta o no el error (Ver Tabla 3-24). Sin embargo la prueba no se trata de comparar el mejor modelo sino cómo influye el error presente cuando se toma como entrada al modelo la volatilidad realizada comparada con la volatilidad integrada, y los resultados arrojan que no existen diferencias importantes en la capacidad de pronóstico de los modelos si se toma en cuenta o no la medida del error presente en la volatilidad realizada.

Tabla 3-24. Resultados (Asai et al. 2012)

	Dos componentes		ARFIMA		HAR	
	R ² CP	R ² CT	R ² CP	R ² CT	R ² CP	R ² CT
1 día						
ME	0.308	0.384	0.306	0.386	0.276	0.347
Sin ME	0.305	0.38	0.307	0.386	0.301	0.377
MA	0.261	0.33	0.263	0.343	0.302	0.382
5 días						
ME	0.296	0.366	0.287	0.366	0.299	0.378
Sin ME	0.302	0.37	0.289	0.369	0.281	0.358
MA	0.277	0.351	0.287	0.372	0.285	0.367
10 días						
ME	0.289	0.357	0.261	0.338	0.296	0.373
Sin ME	0.301	0.366	0.263	0.339	0.267	0.339
MA	0.26	0.332	0.268	0.351	0.269	0.346
20 días						
ME	0.28	0.343	0.226	0.297	0.294	0.368
Sin ME	0.291	0.352	0.225	0.295	0.272	0.343
MA	0.248	0.312	0.235	0.312	0.239	0.313

3.28 A hybrid modeling approach for forecasting the volatility of S&P 500 index return (Hajizadeh, Seifi, Fazel Zarandi, & Turksen, 2012)

Descripción de los datos: Diario, periodo comprendido entre 02/01/1998 y 31/08/2009, en donde se hace un proceso de selección entre un conjunto de variables exógenas y endógenas, el conjunto de variables explicativas es el siguiente: S&P500, índice Nasdaq, índice Dow Jones, volatilidad del día anterior, notas del tesoro de 3 y 6 meses, retornos cuadrados del índice S&P500, volatilidad, volumen tranzado del índice S&P500. Los datos utilizados en el modelo se distribuyen un 70% como periodo de entrenamiento, 20% validación y 10% de prueba.

Metodología: GARCH, EGARCH, GJR, híbrido entre red neuronal y EGARCH.

Procedimiento y resultados: Inicialmente se realiza una prueba AICC y BIC para comprobar que configuración de modelos es la más precisa (Ver Tabla 3-25), posteriormente se evalúan estos modelos GARCH y se selecciona el modelo más preciso de los propuestos y que va a ser también el que se incluya en el modelo híbrido.

Tabla 3-25. Resultados (Hajizadeh et al. 2012)

	(p,q)	AIC	BIC
EGARCH	(3,3)*	18150.6	18072.9
EGARCH	(1,1)	18067.1	18025.3
GARCH	(2,2)*	17961.5	17913.7
GARCH	(1,1)	17943.1	17907.3
GJR-GARCH	(3,3)*	18072.1	17994.4
GJR-GARCH	(1,1)	18061.3	18019.4

En el artículo se muestran las evaluaciones del pronóstico de cada uno de los modelos en donde el modelo EGARCH arroja los mejores resultados, con el modelo de la familia GARCH seleccionado que hará conjunto con la red neuronal, se compara el modelo con dos modelos híbridos, la diferencia entre ambos modelos híbridos es que el modelo 2 tiene un proceso adicional en la serie de datos que busca reducción en el error.

Tabla 3-26. Resultados 2 (Hajizadeh et al. 2012)

	10 días			15 días		
	EGARCH (3,3)	Hybrid I	Hybrid II	EGARCH (3,3)	Hybrid I	Hybrid II
RMSE	0.005426	0.003156	0.002558	0.004374	0.0026593	0.002049
MAE	0.005305	0.00268	0.002069	0.004176	0.0021505	0.001522
MAPE	0.841136	0.433799	0.338739	0.58192	0.304631	0.227798
ME	0.005305	0.001906	0.001614	0.004176	0.001471	0.001207

Según la Tabla 3-26, el modelo que mejor arroja resultados según todos los criterios para ambos horizontes de tiempo es el modelo Hybrid II.

Conclusiones: Los modelos de redes neuronales ayudados con los modelos GARCH, arrojan mejores resultados que los modelos GARCH.

3.29 Quantile forecasts of financial returns using realized garch models(Watanabe, 2012)

Descripción de los datos: Diario, periodo comprendido entre 03/01/1996 y 27/02/2009, se obtienen dos series de tiempo una corresponde a la volatilidad realizada y la otra a la volatilidad Kernel.

Metodología: GARCH, EGARCH.

Procedimiento y resultados: El artículo tiene como objetivo encontrar que distribución de los errores es más apropiada para los errores en un modelo GARCH, entre una normal (n), una t-student (t) y una t-student asimétrica (skt), también estudia medidas de volatilidad, la volatilidad realizada (RV) que es la medida más tradicional en los modelos de pronóstico y volatilidad kernel (KR).

En general los resultados para $\alpha=5\%$ y 10% son similares en todos los modelos (Ver . Las grandes diferencias se presentan en $\alpha=1\%$ donde el mejor desempeño lo tiene el modelo GARCH(RV)skt que reporta el menor VaR y ES.

Tabla 3-27. Resultados (Watanabe, 2012)

	VaR α			ES α		
	10%	5%	1%	10%	5%	1%
GARCH(RV)-n	10.323	5.332	1.645	0.092	0.168	0.437
GARCH(RV)-t	10.72	5.672	1.134	0.07	0.089	0.237
GARCH (RV)-skt	10.267	4.594	0.908	0.077	0.03	0.087
GARCH (RK)-n	10.38	5.615	1.645	0.11	0.174	0.443
GARCH (RK)-t	11.004	5.729	1.248	0.072	0.091	0.223
GARCH (RK)-skt	10.323	4.651	0.908	0.076	0.031	0.12
EGARCH-n	10.04	5.445	1.872	0.136	0.24	0.456
EGARCH -t	10.777	5.559	1.531	0.082	0.14	0.287
EGARCH -skt	10.607	5.218	1.248	0.071	0.064	0.211

Conclusiones:

El modelo que mejor reporta resultados es el que tenga en cuenta una distribución t-student asimétrica, para todos los modelos.

4. Resultados y análisis RSL S&P500

Este capítulo presenta los resultados encontrados en los dos capítulos anteriores

4.1 Resumen de los criterios de desempeño

A continuación se realiza una clasificación de los criterios de desempeño encontrados en la RSL del índice S&P500 donde se reportan más de 20 criterios.

Tabla 4-1. Criterios de desempeño agrupados por autor

Criterio		Frecuencia
AIC	(Qi & Zhang, 2001),(Hajizadeh et al., 2012),	2
AICC	(Qi & Zhang, 2001),	1
BIC	(Qi & Zhang, 2001), (Hajizadeh et al., 2012),	2
CD	(Cao & Tay, 2001),	1
Correlación	(Becker et al., 2007)	1
CU	(Cao & Tay, 2001),	1
DA	(Qi & Zhang, 2001), (Cao & Tay, 2001)	2
DM	(Qi & Zhang, 2001),(Acosta et al., 2007), (B. H. Kim et al., 2010)	3
MDM	(Qi & Zhang, 2001), (Sadorsky, 2005),	2
ES	(Watanabe, 2012)	1
HMSE	(Martens et al., 2009)	1
Max drowdown	(Kumar, 2009)	1

Criterio		Frecuencia
MAE	(Qi & Zhang, 2001), (Cao & Tay, 2001), (Sadorsky, 2005), (Corrado & Truong, 2007), (Enke & Amornwattana, 2008), (Hung et al., 2009), (Yu et al., 2009), (Martens et al., 2009), (Kellard et al., 2010), (Chu & Lam, 2011)	10
MAPE	(Qi & Zhang, 2001), (Sadorsky, 2005), (Majhi et al., 2009), (Hajizadeh et al., 2012)	4
ME	(Qi & Zhang, 2001), (Yu et al., 2009), (Martens et al., 2009), (Hajizadeh et al., 2012)	4
MME(O)	(Hung et al., 2009)	1
MME(U)	(Hung et al., 2009)	1
MPE	(Sadorsky, 2005)	1
MSE	(Sadorsky, 2005), (Acosta et al., 2007), (Becker & Clements, 2008), (Enke & Amornwattana, 2008), (Hung et al., 2009), (Yu et al., 2009), (Martens et al., 2009), (Chu & Lam, 2011)	8
NMSE	(Cao & Tay, 2001)	1
R ²	(Qi & Zhang, 2001), (Corrado & Truong, 2007), (Martens et al., 2009), (Asai et al., 2012), (S. S. Chen, 2009), (Kellard et al., 2010), (Chu & Lam, 2011)	7
Rentabilidad	(Moody & Saffell, 2001), (Shen & Loh, 2004) (Brooks & Katsaris, 2005), (Kumar, 2009), (S. S. Chen, 2009), (Duarte & Brazil, 2011)	5
RMSE	(Qi & Zhang, 2001), (Corrado & Truong, 2007), (B. H. Kim et al., 2010), (Hoogerheide et al., 2010), (Kellard et al., 2010), (Hajizadeh et al., 2012)	6
SSE	(Qi & Zhang, 2001)	1
Hit ratio	(Qi & Zhang, 2001), (Yu et al., 2009), (Kumar, 2009), (Hoogerheide et al., 2010)	3
Sharpe ratio	(Moody & Saffell, 2001), (Thawornwong & Enke, 2004) (Shen & Loh, 2004), (Brooks & Katsaris, 2005), (Kumar, 2009)	4
Sterling ratio	(Moody & Saffell, 2001)	1
% precisión	(Enke & Amornwattana, 2008)	1

Criterio		Frecuencia
Theil U	(Sadorsky, 2005)	1
VaR	(Hung et al., 2009), (Watanabe, 2012)	2

Tabla 4-2. Pruebas de validación de datos reportadas en la revisión

Prueba	
Box pierce	(Durham, 2007)
Durbin Watson	(Corrado & Truong, 2007)
Fisher Sneadecor	(Becker et al., 2007),
Jarque Bera	(Durham, 2007), (Acosta et al., 2007), (Hung et al., 2009)
Ljung Box	(Acosta et al., 2007)
LM	(Hung et al., 2009)
Q like	(Becker & Clements, 2008), (Acosta et al., 2007), (Hung et al., 2009)
t-student	(Chu & Lam, 2011)

La Tabla 4-1 resume todas las medidas encontradas en los artículos revisados en donde el objetivo fuera pronosticar el índice bursátil S&P500, se encuentra que la mayoría de autores seleccionan los criterios MAE, MSE y RMSE como principales criterios para medir el desempeño en todo tipo de modelos, como se demostró al final del capítulo 2 (Ver Tabla 2-4) estas medidas no son apropiadas para medir este tipo de modelos, ya que no están enfocadas a los principales objetivos de este tipo de modelos de pronóstico que son, buscan rentabilidad y manejar el riesgo. Se puede afirmar que se alcanza a llenar un vacío en el área de pronóstico de series financieras, en donde se comprueba que la gran parte de los autores no miden correctamente el desempeño de sus modelos.

Se recomienda utilizar como medida de desempeño medidas como rentabilidad, Sharpe Ratio, DA, CU, CD que miden realmente el desempeño de un modelo de pronóstico en

series de tiempo financieras enfocados en los objetivos de rentabilidad y administración del riesgo.

La Tabla 4-2 muestra alguna de las pruebas empleadas para aceptar o rechazar hipótesis sobre características de la serie de tiempo empleada, debido a que pocos autores las emplean y parece que no fuese relevante su uso, no se hace un estudio diferente a identificar la presencia de ellas en algunos artículos.

4.2 Resumen de modelos de pronóstico

A continuación se presenta un resumen con todos los modelos reportados en la RSL, para obtener detalles de cada autor remitirse al capítulo 3 en donde se encuentran una descripción del modelo, resultados y conclusiones relevantes para este estudio.

Tabla 4-3. Modelos reportados clasificados por autor

	Modelos propuestos	Mejor modelo
(Qi & Zhang, 2001)	Red neuronal multicapa hacia adelante	RN (5,5)
(Cao & Tay, 2001)	SVM, BP-1, BP-2	Similares
(Moody & Saffell, 2001)	Buy and hold, Q trader, RRL trader	RRL trader
(Shen & Loh, 2004)	SVM, Buy and hold	SVM
(Thawornwong & Enke, 2004)	Red RMSE promedio, red RMSE mínimo, red probabilística, regresión lineal, Buy and hold, RW, T-bill	Red RMSE mínimo
(Sadorsky, 2005)	RW, HM, MA 20, MA 60, MA 180, ES, ES 20, ES 60, ES 180, LR, AR 5, SV1, SV2	ES
(Brooks & Katsaris, 2005)	Buy and hold, T-bill, RS aumentado, RS Van, Norden Schaller, RS EC B V, RS EC	Buy and hold
(Durham, 2007)	SV, SV mix	SV
(Corrado & Truong, 2007)	GARCH con diferentes entradas, solo S&P500, volatilidad implícita, máximo y mínimo diario, todas las entradas	Garch todas entradas

	Modelos propuestos	Mejor modelo
(Becker et al., 2007)	GARCH, ARMA, ARFIMA, SV	ARMA
(Acosta et al., 2007)	GARCH, SV, SETAR-TGARCH, SETAR-THSV	SV
(Becker & Clements, 2008)	ARMA+ARFIMA (R), (ARMA+ARFIMA+SVRV+VIX) R, ALL R, ALLMBF R, ARMA+ARFIMA U, ARMA, ARFIMA, GARCHRV, GJRRV, GJRRVG, VIX, GARCH, SV, GJR, SVRV	ARMA+ARFIMA (R)
(Enke & Amornwattana, 2008)	GRNN, MRL, C4.5	C 4.5
(Balbás et al., 2008)	Regresión lineal, regresión lineal con algoritmo de corrección	Algoritmo corrección
(Lam & Ng, 2009)	GARCH, MEM-GARCH	MEM-GARCH
(Hung et al., 2009)	GJR, GJR-VIX, GJR-PK, GJR-GK, GJR-RS	GJR-VIX
(Yu et al., 2009)	ARIMA, LDA, BPNN, SVM, LSSVM poly, LSSVM rbf, LSSVMsig, LSSVMmix, LSSVMgafs, LSSVMgapo, LSSVMevol	LSSVMevol
(Pooter, 2009)	GJR (1,1) DA, ARFI, ARFI-D, ARFI-DA, ARFI-DAR, ARFI-DARL, ARFI-DAXRL, SPARFI-DAXRL, HAR-DAXRL, AR(1)-DAXRL, AR(5)-DAXRL, AR(22)-DAXRL	AR(22)-DAXRL
(Majhi et al., 2009)	FLANN 1, FLANN 2, FLANN 3, FLANN 4	FLANN 4
(Kumar, 2009)	ANN, ARIMA	ANN
(S. S. Chen, 2009)	Buy and hold, spread (3M-10Y), spread (3M-5Y), M1, M2, tasa de inflación, cambio tasa de desempleo, crecimiento producción industrial,	Tasa de inflación

	Modelos propuestos	Mejor modelo
	cambio tasa de cambio, tasas fondos federales, cambio en niveles deuda pública	
(B. H. Kim et al., 2010)	TRTECM, ECM	TRTECM
(Hoogerheide et al., 2010)	LI, HI, SV, RSV, BMA, LIN, TVW, RTWW	RSV
(Kellard et al., 2010)	MM, DP, DY	MM
(Chu & Lam, 2011)	GARCH, FIGARCH, HYGARCH	HYGARCH
(Duarte & Brazil, 2011)	Modelo lógica difusa, buy and hold	Lógica difusa
(Asai et al., 2012)	AR(1)+ARMA(1,1), ARFIMA, HAR	ARFIMA
(Hajizadeh et al., 2012)	EGARCH (3,3), Hybrid I, Hybrid II.	Hybrid II
(Watanabe, 2012)	GARCH(RV)-n, GARCH(RV)-t, GARCH (RV)-skt, GARCH (RK)-n, GARCH (RK)-t, GARCH (RK)-skt, EGARCH-n, EGARCH -t, EGARCH -skt	GARCH (RV)-skt

En la Tabla 4-3 se identifica un comportamiento interesante en los artículos de Becker y Clements (2008), Becker et al. (2007), Pooter (2009), concluyen que los modelos que ofrecen mejores resultados pertenecen a la familia de modelos AR, que además resultan ser los más simples, este mismo fenómeno del modelo más simple es el que mejor resultados ofrece se presenta en Watanabe (2012), Hoogerheide et al. (2010), Sadorsky (2005), Durham (2007). Brooks y Katsaris (2005) arrojan un resultado inesperado en donde un modelo tan complejo como lo es cambio de régimen arroja peores resultado que la estrategia Buy and hold.

Los modelos más utilizados en el pronóstico de índice bursátil son modelos GARCH, modelos ARMA (Ver Tabla 4-3), redes neuronales, maquinas de soporte de vectores, volatilidad estocástica, considerando estos modelos como familias, cada autor utiliza

variaciones de estos según su criterio o incluso llegan a utilizar mezclas de modelos buscando mejores resultados.

Del capítulo 3 se puede concluir que las variables explicativas de los modelos cumplen lo planteado en el capítulo 1. Estas variables son:

- Variables macroeconómicas: Todo tipo de variables macroeconómicas como desempleo, PIB, inflación, producción industrial, tasas de los bonos del tesoro, entre otras (Thawornwong & Enke, 2004), (Pooter, 2009),.
- Indicadores técnicos: Todos estos se basan en la información que surge día a día en los mercados bursátiles, como lo es el máximo, mínimo, volumen, volatilidad, máximos y mínimos anteriores, y relaciones entre estos, conformando este tipo de indicadores (Majhi et al., 2009), (S. S. Chen, 2009).
- Series históricas índices: La variable que se encuentra en la mayoría de artículos examinados por la RSL toman como variable explicativa el comportamiento histórico del índice S&P500, lo cual afirma que los autores consideran que el comportamiento histórico del índice contiene información que permite describirlo y pronosticarlo (Becker et al., 2007), (Acosta et al., 2007).
- Volatilidad implícita: Algunos autores incluyen la volatilidad implícita de las opciones sobre futuros del S&P500 (Becker et al., 2007), (Becker et al., 2007), (Hung et al., 2009),

4.3 Medición de los modelos

Uno de los objetivos de esta investigación es establecer cuál de los modelos propuestos ofrece mejores resultados, en la sección 4.1 se puede corroborar que gran parte de los modelos de pronóstico propuestos para índice bursátil utilizan medidas de desempeño inapropiadas para medir rentabilidad y riesgo. En el capítulo 2 se recomienda utilizar un conjunto de criterios que resultan apropiados para los objetivos que podría tener un modelo de pronóstico en índice bursátil y cualquier activo financiero en general (Ver Tabla 2-4).

Este punto busca comparar el desempeño de los modelos que toman en cuenta como medida los indicadores recomendados al final del capítulo 2 y seleccionar los modelos que arrojan mejores resultados de la RSL.

Tabla 4-4. Comparación de modelos por rentabilidad

	Rentabilidad reportada	Rentabilidad anual
(Moody & Saffell, 2001),		
Buy and hold	1348%	11.5%
Q trader	3359%	16.2%
RRL trader	5860%	18.5%
(Shen & Loh, 2004)		
Buy and hold	163	14.2%
SMV	539	12%
(Brooks & Katsaris, 2005)		
Buy and hold	0.60%	7.4%
Libre de riesgo	0.05%	0.6%
RS aumentado	0.51%	6.3%
RS Van Norden Schaller	0.36%	4.4%
RS EC B V	0.38%	4.7%
RS EC	0.28%	3.4%
(S. S. Chen, 2009)		
Buy and hold	0.6	7.4%
Spread (3M-10Y)	0.96	12.1%
Spread (3M-5Y)	0.88	11.1%
M1	0.68	8.5%
M2	0.77	9.6%
Tasa de inflación	1.14	14.6%
Cambio tasa de desempleo	0.71	8.9%
Crecimiento producción industrial	0.6	7.4%
Cambio tasa de cambio	0.58	7.2%

	Rentabilidad reportada	Rentabilidad anual
Tasas fondos federales	0.6	7.4%
Cambio en niveles deuda pública	0.58	7.2%
(Kumar, 2009)		
ANN	55.48%	55.48%
ARIMA	17.40%	17.40%
(Duarte & Brazil, 2011)		
Modelo difuso	13.04%	13.04%
Buy and hold	7.82%	7.82%

El modelo que resulta más rentable es la red neuronal propuesta por Kumar (2009) (Ver Tabla 4-4), al revisar este resultado se observa que corresponde al año 2008 y fue el inicio de la crisis financiera con la quiebra de Lehman Brothers, en donde cualquier modelo inmediatamente ordena ventas en corto obteniendo ganancias atípicas, por esto se concluye que este resultado no es fiable, a este modelo le sigue el modelo RRL trader y Q trader propuesto por Moody y Saffell (2001), en este caso el indicador de rentabilidad presenta confiabilidad ya que recoge el desempeño del modelo en 25 años, los modelos GARCH propuestos por Chen (2009) arrojan resultados interesantes involucrando variables macroeconómicas como explicativa, los modelos que resultan ser un fiasco son los propuestos por Brooks y Katsaris 2005 como se menciona en la tabla anterior, por lo anterior se concluye que los modelos propuestos para el índice S&P 500 son capaces de superar al mercado, sin embargo es importante anotar que la estrategia Buy and hold debería ser revisada ya que un inversionista que observe cambios muy grandes en su portafolio podría cambiar de parecer y querer vender cuando su portafolio alcance ciertos niveles.

Tabla 4-5. Comparación de modelos por Hit ratio

	Hit ratio
(Qi & Zhang, 2001)	
Red neuronal (4,2) (Yu et al., 2009)	No reporta
ARIMA	55.78%
LDA	61.43%
BPNN	67.46%
SVM	72.61%
LSSVM poly	70.43%
LSSVM rbf	75.45%
LSSVMsig	74.78%
LSSVMmix	77.86%
LSSVMgafs	82.43%
LSSVMgapo	82.58%
LSSVMevol	82.66%
(Hoogerheide et al., 2010)	
LI	52.7%
HI	54.9%
SV	61.4%
RSV	59.8%
BMA	58.7%
LIN	61%
TVW	60.2%
RTVW	59.8%
(Kumar, 2009)	
ANN	53.76%
ARIMA	49.23%

Tabla 4-6.Comparación de modelos por Sharpe ratio

	Sharpe ratio
(Moody & Saffell, 2001),	
Buy and hold	0.34
Q trader	0.63
RRL trader	0.83
(Shen & Loh, 2004)	
SMV	1.31
(Thawornwong & Enke, 2004)	
Red RMSE promedio	0.39
Red RMSE mínimo	0.43
Red probabilística	0.37
Regresión lineal	0.22
Buy and hold	0.32
RW	0.27
(Brooks & Katsaris, 2005)	
Buy and hold	0.1541
Libre de riesgo	
RS aumentado	0.2172
RS Van Norden Schaller	0.1711
RS EC B V	0.1306
RS EC	0.0869
(Kumar, 2009)	
ANN	2.55
ARIMA	1.9

Los modelo que presentan mejores Sharpe Ratio son el modelo ANN y ARIMA de Kumar(2009), y el modelo SMV de Shen y Loh (2004). El indicador Hit Ratio es un indicador útil pero debe de ser utilizado con cuidado porque solo mide la capacidad del modelo para predecir un número, sin embargo la predicción puede estar mal en el número y obtener ganancias.

DA,CD y CU solo los toma en cuenta dos autores y uno de estos no reporta resultados, estos criterios resultan muy útiles porque reportan el número de veces que el modelo genera utilidades, estos criterios sumados a la rentabilidad representan un conjunto de medidas apropiadas para evaluar la capacidad predictiva de un modelo.

4.4 Artículos descartados en la RSL

Es importante mencionar los artículos que fueron revisados parcialmente y se descartaron porque su objetivo no era predecir el índice S&P500. Algunos reportaban los resultados gráficamente (Geweke & Amisano, 2011), (A. S. Chen, Cheng, & Cheng, 2009). La principal causa de rechazo fue que el objetivo era pronosticar los futuros del índice S&P500, activos que no eran objeto de investigación en este estudio.

Tabla 4-7. Artículos descartados agrupados por autores

Artículos rechazados en la RSL
(Eakins & Stansell, 2003),(Martens et al., 2009),(Xekalaki & Degiannakis, 2005),(Wilkins & Röder, 2006),(Casas, 2008),(Y. S. Kim, Rachev, Bianchi, & Fabozzi, 2008),(Chavarnakul & Enke, 2008),(Chou, Wu, & Liu, 2009),(Zhang & Wu, 2009),(Capelli, Mariani, Recchioni, Spinelli, & Zirilli, 2010), (Lu & Perron, 2010),(Aretz, Bartram, & Pope, 2011), (Geweke & Amisano, 2011),(Wu & Lee, 2011).

5. Metodología para formular un modelo de pronóstico

Este capítulo tiene como objetivo recoger la información que se considera más relevante en cuanto a metodología se refiere y proponer una que permita la comparación entre varios autores, para así obtener un avance más rápido en esta área del conocimiento. Actualmente cada autor selecciona varios de los parámetros en la formulación del modelo en forma arbitraria, siendo el criterio de desempeño el que puede presentar mayores problemas a la hora de interpretar un modelo de pronóstico de series de tiempo financieras.

5.1 Objetivo del modelo de pronóstico

Proponer un modelo de pronóstico en series de tiempo financieras puede tener varios objetivos, para la metodología propuesta se asumirá que la razón de ser de un modelo de pronóstico en series de tiempo financieras es obtener rentabilidad, disminuir el riesgo o una mezcla de ambas.

5.2 Metodología propuesta

El método que se propone a continuación es un resultado de este trabajo final de maestría y consta de 7 pasos, cubre desde la selección y recolección de información hasta la publicación de resultados.

1. Selección de la serie de tiempo

Cuando se va a realizar un modelo de pronóstico en series de tiempo financieras se puede presentar el problema que sea una serie poco estudiada, en este caso se sugiere tomar mínimo dos series de tiempo financieras, una que es el objeto de estudio y otra(s) ampliamente estudiada(s) como lo son: S&P500, DJIA, Nasdaq 100, DAX 30, Notas del tesoro de los Estados Unidos de 10 años. Esto se hace con dos objetivos, el primero es establecer que los pronósticos del modelo no se deban a las características de la serie sino a la propuesta del autor, y el segundo es aprovechar la formulación del modelo como aporte al área del conocimiento permitiendo un avance más rápido en esta, ya que va a permitir compararse con mayor cantidad de modelos en la serie más estudiada.

2. Selección del número de datos y periodos de validación

El autor puede seleccionar el número de datos a recolectar como quiera 20 años como lo propone Sadorsky (2005), toda la serie disponible 115 años como Brooks & Katsaris (2005), 13 años como Becker & Clements (2008) o solo 4 años como Enke & Amornwattana 2008.

La propuesta metodológica recomendada independiente del periodo de tiempo a tomar es validar el modelo en diferentes periodos de tiempo, es decir ingresar al modelo diferente número de datos y probarlo para fechas diferentes como lo hace Thawornwong & Enke (2004) (Ver 3.5), Kellard et al. (2010) (Ver 3.24), Kumar (2009) (Ver 3.20).

3. Horizonte de tiempo

Respecto al horizonte de tiempo es importante que el modelador identifique la necesidad de pronóstico que desea satisfacer si es corto o largo plazo, después de identificada esta necesidad la recomendación es realizar pronósticos con el periodo de tiempo deseado y periodos cercanos con el fin de identificar para que horizonte de tiempo cumple mejor su función, como lo hace Qi & Zhan (2001) (Ver 3.1), Corrado & Truong (2007)(Ver 3.9), Enke & Amornwattana (2008) (Ver 3.13).

4. Descripción de la metodología

Este punto es abordado de una manera apropiada por gran parte de los autores, unos son más extensos que otros, sin embargo esto puede depender de las condiciones de las revistas académicas y también de la complejidad del modelo. Como recomendación se quiere resaltar dos artículos de diferente extensión que sobresalen por la facilidad en su lectura y comprensión debido a su formato, gráficos y tablas, estos son: Thawornwong & Enke (2004), Hajizadeh et al (2012).

5. Selección del criterio de desempeño

En el capítulo 2 se hace una revisión de los criterios de desempeño utilizados en modelos de pronóstico de índice bursátil para el índice S&P500 utilizados por los últimos 12 años, al final del capítulo se recoge una tabla con un resumen y recomendaciones respecto a la selección de criterios de desempeño del modelo propuesto (Ver Tabla 2-4).

6. Medición del modelo

Medir la precisión de la metodología propuesta siguiendo los parámetros establecidos en el numeral 2 y 4, es de suma importancia ser coherente con las medidas y que estas puedan ser comparadas con otros modelos, las medidas propuestas en la Tabla 2-4 apuntan a este objetivo sin embargo el modelador debe de confirmar esto y evitar situaciones como la propuesta en la Tabla 5-1 en donde los resultados son incoherentes si se comparan entre sí, porque se supone que a mayor horizonte de tiempo mayor es el error en el pronóstico, claramente se identifica que los autores tomaron la medida desde una perspectiva diferente.

Tabla 5-1 Comparación MAE varios autores

Autor	MAE	Horizonte
(Qi & Zhang, 2001)	0.026	Anual
(Cao & Tay, 2001)	0.4353	Diario
(Sadorsky, 2005)	0.33	Mensual

7. Estrategia Buy and hold

La estrategia Buy and hold o comprar y mantener es ampliamente utilizada para comparar el desempeño en los modelos de pronóstico, cuando el modelo supera esta estrategia se considera útil porque está en capacidad de superar al mercado. Esta comparación no tiene en cuenta dos grandes consideraciones, la primera es que comparar una estrategia que solo permite compras en largo y no ventas en corto, se pone en desventaja inmediatamente y es altamente probable que la mayoría de los modelos la superen, en otras palabras en tiempos de crisis el inversionista mantiene su inversión todo el tiempo, mientras que un modelo de pronóstico inmediatamente arroja hacer ventas en corto, la segunda es que lo anterior no necesariamente es cierto puesto que el inversionista al observar ciertos niveles de retroceso en el activo en el que se encuentre su inversión, puede verse tentado a vender y cuando identifique estabilidad en el mercado volver a ingresar.

Para evitar este problema se recomienda comparar los modelos de pronóstico con una estrategia Buy and hold de umbral, esto quiere decir que el inversionista mantendrá su inversión en el mercado mientras no observe un retroceso determinado previamente, e inmediatamente el activo reporte una caída equivalente al umbral previamente establecido vender, después de que el mercado supere el periodo de bajista el inversionista identifica cuando el mercado se encuentra nuevamente al alza con el mismo nivel tenido en cuenta anteriormente al realizar la venta y vuelve a ingresar en el activo. Esta estrategia se asemeja más al comportamiento de un inversionista, que la estrategia Buy and hold que se utiliza regularmente en la literatura.

6. Conclusiones

Los resultados de la Revisión Sistemática de Literatura realizada al índice S&P500 ha resultado en un trabajo difícil porque no existe una metodología que permita clasificar las propuestas de diferentes autores. La gran mayoría de artículos siguen el mismo orden en la formulación de los modelos, empezando por la necesidad de pronóstico, explicando las series de datos y los modelos con los que se piensa predecir, formulando el modelo a partir de un periodo de entrenamiento y validando los resultados en un periodo de prueba. La problemática se presenta a la hora de evaluar la capacidad predictiva del modelo, en donde cada autor evalúa de la manera que considere apropiado, sin tener en cuenta el objetivo real que tiene un modelo de pronóstico en series financieras.

En esta investigación se utilizó la metodología de Revisión Sistemática de Literatura en donde los parámetros de búsqueda tienen que ser decididos por el investigador en donde se puede crear un sesgo que ocasione pérdida de información relevante, se intentó por establecer un protocolo claramente acotado que permitiera responder las preguntas de investigación planteadas al inicio del documento.

A continuación se presenta un resumen de las principales conclusiones a las que se llegaron por medio de la RSL:

1. El primer aporte de este trabajo en el campo de la predicción es identificar los problemas que se tienen al momento de medir el desempeño en un modelo de pronóstico para series financieras, en donde se encontraron 28 maneras diferentes de evaluar un modelo, en donde la mayoría de ellas y las más utilizadas no guardan relación con el objetivo real que tiene un modelo de pronóstico de esta clase.

2. En el campo de la predicción, se obtiene un resumen, que recoge gran parte de las técnicas y resultados, reportados por autores que hayan propuesto modelos de pronóstico para el índice S&P500 en el periodo 2000-2012. Este resumen servirá como punto de partida para investigaciones en series financieras de cualquier tipo en cualquier lugar del mundo.
3. Se propone una metodología que con el tiempo permitirá avanzar más rápido en el área de pronóstico de series financieras, permitiendo comparar las diversas técnicas de modelación de diferentes autores, asegurando que la medida este enfocada al objeto real del modelo.

Logros obtenidos en base a los objetivos específicos planteados en el capítulo 1 de este documento:

1. Identificar los modelos de pronóstico que tienen como objetivo el índice S&P 500 a través de una revisión sistemática de literatura, para obtener una visión global de predicción de índices bursátiles.
Este trabajo encontró que los modelos de pronóstico se dividen en grandes grupos principales:
 - GARCH: Modelos que toman la volatilidad como una variable heterocedástica
 - SV: Modelo que toman la volatilidad como una variable estocástica
 - ARMA: Modelos autorregresivos de medias móviles tienen variaciones como lo son los modelos ARIMA y ARFIMA.
 - Redes neuronales: Modelos que permiten retroalimentación que tienen gran capacidad de capturar comportamientos no lineales característicos de las series de tiempo financieras.
 - SMV: Modelos de maquinas de soporte de vectores, son modelos similares a las redes neuronales, su diferencia radica en que mediante algoritmos evitan los problemas de sobre estimación presentes en los modelos de redes neuronales.
 - Modelos programación: Dentro de este tipo de modelos caben los encargados de ajustar los modelos tradicionales por medio de algoritmos adaptando los

parámetros del modelo en el tiempo con el fin de optimizar los resultados comparándolos con los modelos tradicionales.

Estos son los grandes grupos de modelos, dentro de estos grupos existen gran cantidad de propuestas, e incluso un modelo de pronóstico puede llevar en su proceso dos técnicas pertenecientes a diferentes grupos.

2. Describir las medidas de desempeño de los modelos de pronóstico propuestos, que permitan establecer cuáles son las más apropiadas de acuerdo al objetivo del modelador.

Esta investigación obtuvo los hallazgos más importantes en este objetivo, al encontrar 28 medidas de desempeño diferentes para medir la precisión de los modelos de pronóstico propuesto para el índice S&P500 en el periodo 2000-2012. Se encontró que gran parte de los modelos propuestos miden el desempeño de estos, sin tener en cuenta el objetivo real de la formulación de modelos para pronóstico de series financieras que principalmente son la rentabilidad y gestión. Además se hizo una demostración porque los criterios tradicionalmente utilizados en predicción no son útiles, al final del capítulo 2 se propone un conjunto de medidas de desempeño que miden el desempeño en los modelos de pronóstico enfocados a los objetivos reales de estos.

3. Enumerar los criterios de validación de datos utilizados en modelos de pronóstico, con el fin de identificar las propuestas que tengan mayor credibilidad desde el punto de vista estadístico.

Se encontró un conjunto de pruebas con el objetivo de encontrar características en las series de tiempo, esto se hace de clasificar adecuadamente y seleccionar el modelo apropiado. Este fue el punto que se considera menos relevante ya que pocos autores realizan este tipo de pruebas, y parece que el hecho de

caracterizar o no los datos no afectan el desempeño de los modelos de pronóstico.

4. Comparar los resultados de la revisión bibliográfica, para determinar los modelos de pronóstico más acertados y más confiables.

Se encontró que existe gran cantidad de medidas, te la totalidad de los autores. Es posible comparar modelos que tengan la misma medida, sin embargo los hallazgos del capítulo 2 muestran que realizar estas comparaciones podría no ser útil debido a que la medición en los modelos no es la apropiada, sin embargo es importante mencionar que varios autores reportan al modelo más simple como el más preciso especialmente los modelos de medias móviles.

Sigas propuestas que miden los modelos de forma apropiada y se encontró ANN y ARIMA, para medidas Hit ratio y Sharpe ratio, el modelo LSSVMevol perteneciente el grupo de los modelo SV, en cuanto a rentabilidad, y una mención especial de los modelo ARIMA y GARCH simples con buenas rentabilidades superiores a la estrategia Buy and hold y modelo más complejos.

5. Proponer un protocolo de selección de medidas de desempeño para futuras aplicaciones de predicción, dirigido hacia objetivos y que permitan comparaciones entre varios autores y metodologías.

Como resultado de la RSL se propone una metodología que recoge las mejores prácticas propuestas por los autores a lo largo del periodo de revisión, asegurando que si se sigue al pie de la letra se podrá obtener un mayor avance en el área de pronóstico de series financieras ya que se podrán comparar todas las técnicas propuestas por diferentes autores para todo tipo de series de tiempo financieras.

Finalmente se propone para investigaciones futuras, seguir la metodología recomendada en el capítulo 5, también se espera que esta metodología siga un proceso de mejora

continua enfocada al objetivo de generar utilidades y minimizar el riesgo por parte de los especialistas en pronóstico de series financieras. También se recomienda por parte de alguna entidad especializada la creación de una página libre estilo wiki en la que se permitan publicar resultados y metodologías de cada publicación, esto con el fin de recoger en un solo lugar información descriptiva suficiente para no redundar en estudios y permitir un avance más rápido en esta área del conocimiento.

Bibliografía

- Acosta, L. M., Muñoz, P., & Márquez, D. (2007). Forecasting volatility by means of threshold models. *Journal of Forecasting*, 36(3), 343–363. doi:10.1002/for
- Aretz, K., Bartram, S. M., & Pope, P. F. (2011). Asymmetric loss functions and the rationality of expected stock returns. *International Journal of Forecasting*, 27(2), 413–437. doi:10.1016/j.ijforecast.2009.10.008
- Arnold, I. J. M., & Vrugt, E. B. (2008). Fundamental uncertainty and stock market volatility. *Applied Financial Economics*, 18, 1425–1440. doi:10.1080/09603100701857922
- Asai, M., McAleer, M., & Medeiros, M. C. (2012). Modelling and forecasting noisy realized volatility. *Computational Statistics & Data Analysis*, 56(1), 217–230. doi:10.1016/j.csda.2011.06.024
- Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques - Part II: Soft computing methods. *Expert Systems with Applications*, 36(3 PART 2), 5932–5941. Retrieved from <http://www.bases.unal.edu.co:2066/inward/record.url?eid=2-s2.0-58749094998&partnerID=40&md5=8050c168a72345a0003906226dcbee2e>
- Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, 19, 1165–1195. doi:10.1007/s00521-010-0362-z
- Balbás, A., Balbás, B., Galperin, I., & Galperin, E. (2008). Deterministic regression model and visual basic code for optimal forecasting of financial time series. *Computers & Mathematics with Applications*, 56(10), 2757–2771. doi:10.1016/j.camwa.2008.07.032
- Becker, R., & Clements, A. E. (2008). Are combination forecasts of S&P 500 volatility statistically superior? *International Journal of Forecasting*, 24(1), 122–133. doi:10.1016/j.ijforecast.2007.09.001

- Becker, R., Clements, A. E., & White, S. I. (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking & Finance*, 31(8), 2535–2549. doi:10.1016/j.jbankfin.2006.11.013
- Box, G., & Pierce, D. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, 65, 1509–1526.
- Brooks, C., & Katsaris, A. (2005). Trading Rules from Forecasting the Collapse of Speculative Bubbles for the S&P 500 Composite Index*. *The Journal of Business*, 78(5), 2003–2036. doi:10.1086/431450
- Brownstone, D. (1996). Using percentage accuracy to measure neural network predictions in Stock Market movements. *Neurocomputing*, 10, 237–250. doi:10.1016/0925-2312(95)00052-6
- Budgen, D., & Brereton, P. (2006). Performing systematic literature reviews in software engineering. *Proceeding of the 28th international conference on Software engineering - ICSE*, 1051. doi:10.1145/1134285.1134500
- Cao, L., & Tay, F. E. . (2001). Financial Forecasting Using Support Vector Machines. *Neural Computing & Applications*, 10(2), 184–192. doi:10.1007/s005210170010
- Capelli, P., Mariani, F., Recchioni, M. C., Spinelli, F., & Zirilli, F. (2010). Determining a stable relationship between hedge fund index HFRI-Equity and S&P 500 behaviour, using filtering and maximum likelihood. *Inverse Problems in Science and Engineering*, 18(1), 83–109. doi:10.1080/17415970903234398
- Casas, I. (2008). Estimation of stochastic volatility with LRD. *Mathematics and Computers in Simulation*, 78(2-3), 335–340. doi:10.1016/j.matcom.2008.01.040
- Chavarnakul, T., & Enke, D. (2008). Intelligent technical analysis based equivolume charting for stock trading using neural networks. *Expert Systems with Applications*, 34(2), 1004–1017. doi:10.1016/j.eswa.2006.10.028
- Chen, A. S., Cheng, L. Y., & Cheng, K. F. (2009). Intrinsic bubbles and Granger causality in the S&P 500: Evidence from long-term data. *Journal of Banking & Finance*, 33(12), 2275–2281. doi:10.1016/j.jbankfin.2009.06.005
- Chen, S. S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance*, 33(2), 211–223. doi:10.1016/j.jbankfin.2008.07.013
- Chou, R. Y., Wu, C.-C., & Liu, N. (2009). Forecasting time-varying covariance with a range-based dynamic conditional correlation model. *Review of Quantitative Finance and Accounting*, 33(4), 327–345. doi:10.1007/s11156-009-0113-3

- Chu, C. C. F., & Lam, K. P. (2011). Modeling intraday volatility: A new consideration. *Journal of International Financial Markets, Institutions and Money*, 21(3), 388–418. doi:10.1016/j.intfin.2011.01.002
- Corrado, C., & Truong, C. (2007). Forecasting stock index volatility: comparing implied volatility and the intraday high-low price range. *The Journal of Financial Research*, XXX(2), 201–215.
- Devore, J. (2001). *Probabilidad y estadística para ingeniería y ciencias* (Quinta edi.). Thomson Learning.
- Diebold, F. X., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 263–65.
- Duarte, C., & Brazil, C. (2011). S & P 500 Index Direction Forecasting from 1976 to 2010 : A Fuzzy System Approach. *The International Journal of Digital Accounting Research*, 11(November), 111–134. doi:10.4192/1577-8517-v11
- Durham, G. B. (2007). SV mixture models with application to S&P 500 index returns. *Journal of Financial Economics*, 85(3), 822–856. doi:10.1016/j.jfineco.2006.06.005
- Eakins, S. G., & Stansell, S. R. (2003). Can value-based stock selection criteria yield superior risk-adjusted returns: an application of neural networks. *International Review of Financial Analysis*, 12(1), 83–97. doi:10.1016/S1057-5219(02)00124-2
- Enke, D., & Amornwattana, S. (2008). A hybrid derivative trading system based on volatility and return forecasting. *The Engineering Economist*, 53(3), 259–292. doi:10.1080/00137910802263174
- Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1), 130–141. doi:10.1016/j.jeconom.2011.02.017
- Hajizadeh, E., Seifi, a., Fazel Zarandi, M. H., & Turksen, I. B. (2012). A hybrid modeling approach for forecasting the volatility of S&P 500 index return. *Expert Systems with Applications*, 39(1), 431–436. doi:10.1016/j.eswa.2011.07.033
- Hoogerheide, L., Kleijn, R., Ravazzolo, F., Dijk, H. K. V. A. N., & Verbeek, M. (2010). Forecast accuracy and economic gains from bayesian model averaging using time-varying weights. *Journal of Forecasting*, 269(December 2009), 251–269. doi:10.1002/for
- Hung, J., Ni, R., & Chang, M. C. (2009). The information contents of VIX index and range-based volatility on volatility forecasting performance of S&P 500. *Economics Bulletin*, 29(4), 2592–2604.
- IEEE. (2009). *2009 IEEE Thesaurus*.

- Johnston, J., & Dinardo, J. (1997). *Econometric methods*. (McGraw-Hill, Ed.). New York, New York, USA.
- Kellard, N. M., Nankervis, J. C., & Papadimitriou, F. I. (2010). Predicting the equity premium with dividend ratios: Reconciling the evidence. *Journal of Empirical Finance*, 17(4), 539–551. doi:10.1016/j.jempfin.2010.04.002
- Kim, B. H., Chun, S. E., & Min, H. G. (2010). Nonlinear dynamics in arbitrage of the S&P 500 index and futures: A threshold error-correction model. *Economic Modelling*, 27(2), 566–573. doi:10.1016/j.econmod.2009.11.011
- Kim, Y. S., Rachev, S. T., Bianchi, M. L., & Fabozzi, F. J. (2008). Financial market models with Lévy processes and time-varying volatility. *Journal of Banking & Finance*, 32(7), 1363–1378. doi:10.1016/j.jbankfin.2007.11.004
- Knight, J., Satchell, S., & Wang, G. (2003). Value at risk linear exponent (VARLINEX) forecasts. *Quantitative Finance*, 3(4), 332–344. doi:10.1088/1469-7688/3/4/309
- Kumar, M. (2009). Non linear prediction of the Standard & Poor's 500 and the hang seng index under a dynamic increasing sample. *Asian Academy of Management Journal of Accounting and Finance*, 5(2), 101–118.
- Lam, K. P., & Ng, H. S. (2009). Intra-daily information of range-based volatility for MEM-GARCH. *Mathematics and Computers in Simulation*, 79(8), 2625–2632. doi:10.1016/j.matcom.2008.12.007
- Levy, Y., & Ellis, T. J. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research, 9.
- Liljeblom, E., & Stenius, M. (1997). Macroeconomic volatility and stock market volatility : empirical evidence on Finnish data Macroeconomic v olatility and stock market v olatility : empirical e v idence on Finnish data. *Applied Financial Economics*, 7, 419–426.
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. doi:10.1093/biomet/65.2.297
- Lu, Y. K., & Perron, P. (2010). Modeling and forecasting stock return volatility using a random level shift model. *Journal of Empirical Finance*, 17(1), 138–156. doi:10.1016/j.jempfin.2009.10.001
- Majhi, R., Panda, G., & Sahoo, G. (2009). Development and performance evaluation of FLANN based model for forecasting of stock markets. *Expert Systems with Applications*, 36(3), 6800–6808. doi:10.1016/j.eswa.2008.08.008
- Martens, M., Van Dijk, D., & Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and

- macroeconomic announcements. *International Journal of Forecasting*, 25(2), 282–303. doi:10.1016/j.ijforecast.2009.01.010
- Martens, M., & Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*, 24(11), 1005–1028. doi:10.1002/fut.20126
- Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4), 875–89. doi:10.1109/72.935097
- Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences*. Blackwell Publishing Ltd.
- Pinches, G. E., & Kinney, J. (1971). THE MEASUREMENT OF THE VOLATILITY OF COMMON STOCK PRICES. *Journal of Finance*, 26(1), 119–125. doi:Article
- Qi, M., & Zhang, G. P. (2001). An investigation of model selection criteria for neural network time series forecasting. *European Journal of Operational Research*, 132(3), 666–680. doi:10.1016/S0377-2217(00)00171-5
- Rencher, A. (2002). *Methods of Multivariate Analysis* (Second edi.). John Wiley & Sons, Inc.
- S&P U.S. Indices Methodology. (2012). Standard and Poors. Retrieved from <http://www.spindices.com/indices/equity/sp-500>
- Sadorsky, P. (2005). Stochastic volatility forecasting and risk management. *Applied Financial Economics*, 15(2), 121–135. doi:10.1080/0960310042000299926
- Sharpe, W. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, Fall.
- Shen, L., & Loh, H. T. (2004). Applying rough sets to market timing decisions. *Decision Support Systems*, 37(4), 583–597. doi:10.1016/S0167-9236(03)00089-7
- Thawornwong, S., & Enke, D. (2004). The adaptive selection of financial and economic variables for use with artificial neural networks. *Neurocomputing*, 56, 205–232. doi:10.1016/j.neucom.2003.05.001
- Watanabe, T. (2012). Quantile forecasts of financial returns using realized garch models. *Japanese Economic Review*, 63(1), 68–80. doi:10.1111/j.1468-5876.2011.00548.x
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii.
- Wilkins, S., & Röder, K. (2006). The informational content of option-implied distributions: Evidence from the Eurex index and interest rate futures options market. *Global Finance Journal*, 17(1), 50–74. doi:10.1016/j.gfj.2006.06.006

- Wu, C., & Lee, J. C. (2011). Forecasting Time-Varying Covariance with a Robust Bayesian Threshold Model. *Journal of Forecasting*, 468(May 2010), 451–468. doi:10.1002/for
- Xekalaki, E., & Degiannakis, S. (2005). Evaluating volatility forecasts in option pricing in the context of a simulated options market. *Computational Statistics & Data Analysis*, 49(2), 611–629. doi:10.1016/j.csda.2004.05.030
- Yu, L., Chen, H., Wang, S., & Lai, K. (2009). Evolving least squares support vector machines for stock market trend mining. *IEEE Transaction on Evolutionary Computation*, 13(1), 87–102.
- Zhang, Y., & Wu, L. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications*, 36(5), 8849–8854. doi:10.1016/j.eswa.2008.11.028

