**UNIVERSIDAD DE SANTIAGO DE CHILE**

FACULTAD DE ADMINISTRACION Y ECONOMIA

Departamento de Economía

**Predicción del Value at Risk**

**Ignacio Sepúlveda Llarlluri**

**Profesor Guía: Nombre del profesor guía (sin grado académico**

**Trabajo de graduación para optar al Grado de Magíster en Economía Financiera**

**Santiago – Chile**

2020

**© Ignacio Antonio Sepulveda Llarlluri, 2022**

Esta obra está bajo una Licencia Creative Commons Atribución-Internacional 4.0 (CC BY 4.0).

**Escoja licencia del menú desplegable (letras azules)**

**Si desea que se pueda utilizar con fines académicos pero no comerciales, escoja Atribución–NoComercial o Atribución–NoComercial–Sin Deivadas**

**Si no desea que su tesis sea utilizada, o que le pidan autorización por escrito para usarla con fines académicos, escoja Todos los derechos reservados.**

**HOJA DE RESUMEN**

|  |  |
| --- | --- |
| **Título de la tesis:** Predicción del Value at Risk  **Autor(es):** Ignacio Sepúlveda Llarlluri  **Carrera:** Magister en Economía Financiera  **Profesor Guía:** Ingresar datos  **Año:** 2022 | |
| **RESUMEN**  Obligatorio.  Resumen en 300 palabras  Brevemente: planteamiento del problema, método utilizado, principales resultados y conclusiones. Debe ser: comprensible, sencillo, informativo, preciso, completo y especifico  **ESTO SE HACE AL FINAL.**  **Se sugiere considerar** (explícitamente o implícitamente)**:** Objetivo, Principales hallazgos, Diseño/metodología/Aproximación, Limitaciones del estudio/implicaciones, Implicaciones prácticas, Originalidad/aporte de valor: | |
| **Palabras claves:** | Tres (3) a cinco (5) palabras claves |

**ABSTRACT**

|  |  |
| --- | --- |
| Obligatorio. Forma objetiva, breve y específica del contenido del trabajo realizado. Su extensión no debe superar las 300 palabras. Debe incluir palabras claves. | |
| **Palabras claves:** | Tres (3) a cinco (5) palabras claves |

**DEDICATORIA**

A mi familia.

**AGRADECIMIENTOS**

A mi familia y a la facultad.

**TABLA DE CONTENIDO**

[INTRODUCCIÓN 1](#_Toc520798184)

[1 CAPÍTULO ¡Error! Marcador no definido.](#_Toc520798185)

[1.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798186)

[1.1.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798187)

[2 CAPÍTULO 2](#_Toc520798188)

[2.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798189)

[2.1.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798190)

[3 CAPÍTULO 2](#_Toc520798191)

[3.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798192)

[3.1.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798193)

[4 CAPÍTULO ¡Error! Marcador no definido.](#_Toc520798194)

[4.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798195)

[4.1.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798196)

[5 CAPÍTULO ¡Error! Marcador no definido.](#_Toc520798197)

[5.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798198)

[5.1.1 Subtítulo **¡Error! Marcador no definido.**](#_Toc520798199)

[CONCLUSIONES 12](#_Toc520798200)

[GLOSARIO ¡Error! Marcador no definido.](#_Toc520798201)

[REFERENCIAS BIBLIOGRÁFICAS 13](#_Toc520798202)

[ANEXOS 14](#_Toc520798203)

[APÉNDICES 15](#_Toc520798204)

**ÍNDICE DE TABLAS (OBLIGATORIO)**

[Tabla ‎2‑1 División Hojas de Desarrollo **¡Error! Marcador no definido.**](#_Toc320744140).

**ÍNDICE DE ILUSTRACIONES (OPTATIVO)**

[Fig. ‎2.1 Ejemplo Imagen **¡Error! Marcador no definido.**](#_Toc321173911)

[Fig. ‎2.1 Ejemplo Imagen **¡Error! Marcador no definido.**](#_Toc423611519)

INTRODUCCIÓN

La medición de la incertidumbre entorno a los retornos de los activos financieros es una labor que tanto reguladores, bancos como gestora de fondos realizan de forma rutinaria. En contexto de harta incertidumbre se hace importante poder identificar y medir este riesgo.

El principal desafío de la medición es que requiere satisfacer los tiempos actuales de la industria y la forma en que esta se comunica, por lo tanto, una medida que resuma este riesgo en una sola cifra es lo ideal.

Bajo este argumento surge el Value at Risk, el cual cumple la característica de ser solo un numero y además que es fácil de interpretar, dado que no es más que un cuantil de la distribución de retornos, de particular relevancia practica los cuantiles asociados a la cola izquierda de la distribución.

El problema surge en el cálculo de esta distribución de retornos, dado que en general los retornos tienen presencia de clúster de volatilidad, excesos de curtosis y asimetría entre otras cosas que dificultan una correcta predicción, al menos por métodos tradicionales.

Una primera aproximación es ocupar la historia, o una muestra de ella. Otra forma es asumir una ley de probabilidad para los retornos, lo que implica estimar los parámetros para obtener su función de densidad. Para ser menos restrictivos, podemos modelar directamente la función de densidad de retornos, o un cuantil en particular, esto se conoce como modelo semi paramétricos. Finalmente, están los modelos de aprendizaje supervisado que busca generalizar a partir de una muestra de datos el futuro comportamiento de la variable y se caracterizan por su alto poder predictivo.

**Objetivo general**

La investigación tiene por objetivo general ver cual familia de modelos, o algoritmo, tiene un mejor poder predictivo en el contexto del Value at Risk.

**Objetivos específicos**

Corresponde a evaluar el uso de modelos actuales, como los es la aplicación de una red neuronal recurrente en la predicción de series de tempo, versus técnicas que ya tienen una historia de resultados y literatura en estadística como econometría.

**Hipótesis**

Modelos más complejos debiesen poder generalizar de mejor manera, sobre todo en periodo de alta volatilidad. Por lo tanto, modelos de aprendizaje supervisado, en particular en este caso una red neuronal recurrente, debiesen tener un mejor poder predictivo de forma sistemática a través de los activos testeados y test hechos.

**Metodología**

Se usarán series de precios de activos financieros de cierta relevancia práctica, además que son de libre acceso dado que son descargados de la plataforma Investing. Toda la programación de los modelos fue hecha en Python[[1]](#footnote-1). La serie corresponde al periodo de XXXXX al XXXXX, y evaluaremos el poder predictivo de una cantidad relevante de modelos contra una red neuronal de arquitectura relativamente sencilla. La evaluación se hará mediante métricas recogidas de la literatura. Evidentemente buscamos al modelo con la mejor performance a través de los activos y métricas. El flujo de trabajo en la figura 1 ejemplifica como se desarrolla el proceso aplicado, el cual es un estándar a través de la literatura.

**Figura 1:** Resumen flujo de trabajo.

El contenido restante de este trabajo se divide en 4 capítulos restantes, el primero corresponde a una definición forma del riesgo que nos permite poder obtener una definición más precisa y situar al VaR en un contexto con una cierta coherencia, definimos el VaR en el siguiente capitulo, que implica predecir el VaR y también incorpora la literatura. El próximo desarrolla los modelos que vamos a ocupar para predecir, el siguiente capitulo es la aplicación. Finalizamos con una conclusión, discusión y potencial extensiones.

# RIESGO

## ¿Qué es el riesgo?

En general, podemos definir el riesgo como el cambio probable en una posición entre dos fechas. Pero según Artzner et al. (1999) una mejor definición requiere incorporar el concepto del futuro. Por eso define el riesgo como el relacionado a la variabilidad del valor futuro de una posición debido a eventos incierto. Entonces podemos tratarlo como una variable aleatoria que proviene de un set estados de la naturaleza para un determinado periodo de tiempo.

Continuando con Artzner et al. (1999) el desarrollo de una primera aproximación seria calcular un subconjunto del set de riesgo, que lo definimos como el riesgo aceptable y nuestro interés es saber cuan probable es que mi medida de riesgo caiga dentro de esa zona. Que es una forma general de lo que intentan la mayoría de las medidas de riesgo.

## Características de una medida coherente de riesgo.

Una definición formal se encuentra en Artzner et al. (1999), aquí solo ocupamos sus propiedades deseables.

Supongamos que medimos el riesgo por , son todos los posibles eventos futuros y corresponde a un resultado aleatorio de un activo o portafolio en el estado de la naturaleza i, según el autor las propiedades para que esta medida sea coherente son:

* Primero, . El riesgo de algo que no vale nada es 0. A esto se le conoce como normalización.
* Segundo, si es que entonces . Se le conoce como monotocidad e implica que, por ejemplo, los activos que siempre han presentado un mayor valor esperado en todos los posibles estados de la naturaleza futuros deben tener un menor riesgo.
* Tercero, si . Esto se conoce como subaditividad y lo podemos asociar al principio de diversificación. El riesgo de un portafolio de un activo tiene que ser menor o igual a la suma de sus riesgos individuales.
* Cuarto, si tenemos la constante entonces . Se conoce como homogeneidad positiva e implica que un aumento proporcional de de un activo o portafolio, debiese aumentar el riesgo del activo o portafolio en .
* Finalmente, si es que existe un retorno adicional en de , tenemos que . Su nombre es invariante a la translación esto implica que, si sumo un retorno constante e independiente mi riesgo disminuirá justo en ese valor.

# VALUE AT RISK

Una forma de entender el Value-at-Risk() según Engle et al. (2004) es definirlo como la máxima perdida que tendrá un activo o un portafolio en un cierto periodo de tiempo, dado un cierto nivel de confianza. En la práctica implica estimar la cola izquierda de la distribución retornos(R) dado un , el sera matemáticamente definido como:

Lo que buscamos es dado una función de distribución acumulada de retornos, el valor ínfimo que pertenece a los reales tal que su probabilidad acumulada hasta el punto sea mayor o igual a este. La figura 1 nos muestra gráficamente lo expresado en la ecuación y comprueba lo fácil que es interpretar el VaR.

En Artzner et al. (1999) se muestra que el no cumple con la propiedad de subaditividad si es que la distribución conjunta de los precios no es normal, por lo que es una de las primeras críticas que se le efectúa como medida de riesgo, pero dada su simpleza y relevancia practica sigue siendo importante su estimación.

Gráfico

Descripción generada automáticamente

**Figura 2:** Función de distribución acumulada con datos simulados a partir de una distribución normal estandar, para el calculo del VaR con un 95% de confianza.

## Predicción.

La predicción implica modelar como una función de la información () en , entonces en forma general podemos plantear el problema de predecir como:

Donde acumula toda la información relevante hasta .

## Literatura.

Al ser el VaR un cuantil de una distribución es necesario entender el comportamiento de esta distribución, un resumen de hechos estilizados presentes puede ser encontrado en Cont (2000) y sus referencias. Algunas de particular interés son las relacionadas al comportamiento de la cola de la distribución, como:

* Ausencia de autocorrelación.
* Colas anchas, condicionales e incondicionales.
* Asimetría.
* Agrupamientos de volatilidad.
* Autocorrelación de retornos absolutos.

Una definición formal del VaR como medida de riesgo de mercado se encuentra en el trabajo de Artner (1999).

Este trabajo es mucho mas general, y define el riesgo tanto como presenta una estructura para el análisis, medición y construcción de métricas de riesgo. En particular de nuestro interés es el riesgo de mercado.

En principio fueron dos métodos los utilizados para estimar el VaR, el primero uno completamente no paramétrico donde se ocupaba simplemente el cuantil histórico o de una simulación. Una refinación permitía suponer una distribución de probabilidad teórica para la serie, lo que permitía a través de la estimación de los parámetros a partir de la muestra derivar el cuantil de interés, generalmente se usa una normal o t de Student. Estos métodos son los mas simples, y son relativamente rápidos en términos computacionales. Ambos responden al supuesto de estacionariedad en la distribución de retornos.

Pero Engle (1982) muestra que la varianza condicional podría no ser constante, haciendo que modelos de distribución constante fallaran en capturar la heterocedasticidad. Señala que caracterizar a la varianza como un proceso autoregresivo del error permite hacernos cargos de los agrupamientos de volatilidad, estos son los modelos ARCH.

Una generalización es hecha en Bollerslev (1986), llamada GARCH, para permitir incorporar memoria de largo plazo mediante una estructura más flexible a través de rezagos de la varianza.

Ambos trabajos asumen normalidad en los residuos, una extensión es hecha en Bollerslev (1987) para incorporar una distribución-t, esto permite tomar en cuenta los excesos de curtosis, que existen incluso después de controlar por agrupamientos de volatilidad, esto es lo que implica las colas anchas condicionales Cont (2000).

En general, la autocorrelación de los retornos absolutos es asociados con una propiedad de memoria de largo plazo, para tener esto en consideración es que Dish (1993) desarrolla el modelo APARCH.

Modelo integrados eran conocido para modelar la media de series macroeconómicas y financieras, pero la incorporación de un componente I(1), era visto demasiado restrictivo para varianza dado que asume la persistencia infinita del shock. Para esto Baillie (1996) propone modelo integrados pero de forma fraccional I(d), conocidos como FIGARCH, lo que permite ahora al shock caer de manera hiperbólica con el tiempo, el cual tiene mas memoria que la caída exponencial de un I(0).

Un problema de los modelos anteriores es que implican asumir una distribución específica para los retornos, otra forma es asumir una función explícitamente para un determinado cuantil, o incluso para la densidad completa, esta fue la idea de los modelos, semi paramétricos, lineales de cuantiles propuesto por Koenker and Bassett. Una extensión que permite un componente autoregresivo y otro no lineal, es conocido como modelos CAVIAR los cuales fueron desarrollado por Engle (2004), ya con el desarrollo madurado de la familia de modelos GARCH Engle desarrolla un modelo de cuantiles autoregresivos donde cada uno de sus modelos incorpora algún hecho estilizado documentado de la distribución de retornos, mayor detalle en la sección 4. Los resultados empíricos muestran resultados superiores al GARCH, lo que sugiere una nuevo hecho estilizado que implica que la cola tiene un comportamiento diferente al centro de la distribución.

Distinto trabajos han comparado como es el rendimiento de distinto modelos. En Kuester (2005) se compara GARCH-t combinado con métodos desde *Extreme Value Theory (EVT)* con simulaciones históricas y el CAVIAR, el resultado muestra una clara ventaja del GARCH-t y un pobre rendimiento del CAVIAR. En un trabajo que busca específicamente tratar con el exceso de curtosis y la asimetría de la distribución de retornos, es que Ovaert (2010) prueba que un ARCH con una t-asimétrica es mejor una normal, no puede decir lo mismo sobre modelos de *EVT*. Halbleib prueba con una composición de modelos GARCH y CAVIAR obteniendo resultados mejores que los modelos por separado. Degiannakis (2013) muestra que independiente del horizonte o serie de tiempo usada modelos FIGARCH no ofrecen un rendimiento superior al GARCH, descartando evidencia de memoria de largo plazo. Braione (2016) prueba diferentes distribuciones para distintas configuraciones de GARCH, demostrando que el t-asimétrico es por lejos el mejor a través de todas las métricas y niveles de confianza. En Frezza et al. (2022) se propone un proceso estocástico regularizado para los precios, lo que le permite obtener un VaR que esta disponible para obtener rendimientos similares al GARCH y superiores al CAVIAR.

El uso de aprendizaje supervisado en finanzas, y en particular en riesgo, esta bien documentado, ver por ejemplo Mashrur et al. (2020). En Junge et al. (1998) se modela el VaR como una *Mixture Density Networks,* que es un tipo de red neuronal ocupada para modelar densidades condicionales. En particular ocupan una Gaussian Mixture Networks (GMN), en la cual primero existe una capa densa completamente conectada, donde se modela distribuciones sencillas (Normal) después una capa final las combina para hacer una función de densidad más compleja. Schittenkopf et al. (2000) extiende el GMN para permitir a la varianza ser heterocedasticas, ósea es una red tipo GARCH, llamada Recurrent Gaussian Mixture Networks (RGMN) la cual modifica la capa intermedia que generaba los parámetros de las densidades condicionales para permitir una varianza no constante. Du et al. (2019) modifica ligeramente la idea anterior, y modela directamente la media y la varianza a través de una red neuronal recurrente conocida como long shot-term memory (LSTM) curiosamente es el primer trabajo donde la simulación histórica tiene el mejor resultado, LSTM rinde mejor que el GARCH. Ormaniec et al. (2022) ocupa la misma idea anterior desarrolla una LSTM enfocado en probarlo en contexto altamente heterocedasticos lo mide contra un GARCH, obteniendo similar performance en datos simulados, pero LSTM exhibe un claro mejor desempeño con data de mercado.

El enfoque anterior se caracteriza por intentar aprender los parámetros que gobierna una densidad, o variadas densidades combinadas y a partir de esto obtener una predicción del cuantil. Otro enfoque es modelar directamente el cuantil, utilizando la regresión de cuantil y en particular la *pinball loss* combinada con algoritmos mas complejos.

Petneházi (2021) usa una red neuronal convolucionada (QCNN) de cuantil para predecir el VaR la cual toma ventaja de poder explotar la autocorrelación, la Q solo implica que se modificó la función de perdida a una pinball loss. Sus resultados muestran probes resultados en el caso de entrenamiento univariado, dado que propone otro experimento donde calcula un joint QCNN que implica entrenar con todas las series de retornos disponible la red neuronal, esto le permite obtener resultados competitivos contra el GARCH y regresión de cuantiles.

Finalmente, un enfoque alternativo corresponde al uso de predicción conforme, este es un método que genera conjuntos de predicción, dado un nivel de confianza, los cuales vienen con ciertas garantías dado ciertos supuestos. Para problemas de regresión y clasificación se encuentran en el trabajo pionero de Vovk (2005). El problema es que asume iid, conocido como exchangeability, lo que sabemos que no es válido para la distribución de retornos, para más detalle sobre posibles transgresiones a este supuesto ver Barber et al. (2022). Xu (2021,2022) modifica el supuesto para permitir por errores estacionarios y con strongly-mixing process, el método es conocido como Embpi. De momento no es de conocimiento que se haya ocupado directamente para calcular el VaR, pero en Kath (2021) se ocupan junto a un algoritmo de bosques aleatorios para la predicción de precios de electricidad los cuales son no estacionarios y con presencia de no linealidades y Embpi rinde parecido y mejor que métodos existentes, como lo es regresión de cuantiles más bosques aleatorios. La conexión entre conjunto de predicción y el VaR es clara, dado que el VaR será simplemente la banda inferior.

De lo anterior, es evidente que el resultado del mejor modelo no es universal y es dependiente de los datos, esto implica que es necesario probarlo en activos con diferentes comportamientos para poder robustecer la conclusión. Por otro lado, modelo de aprendizaje profundo son visto para obtener resultados similares cuando se los comparan con los modelos que han mostrado un rendimiento superior con anterioridad, pero cuando se reduce el universo de modelos comparados es más probable que el método logre mejores resultados que los modelos bases propuesto, esto implica la necesidad de establecer una cantidad suficientes de modelos base.

# MODELOS

Se presentan 4 familias de modelos, donde cada uno contendrán un conjunto de potenciales modelos. El primero es el normal, donde parametrizamos la función de densidad como una normal. Después están la familia de modelos GARCH, que permite hacernos cargo en caso de un proceso autorregresivo y, o, de varianza no constantes. Para no depender de una distribución de probabilidad especifica, la familia de modelo CAVIAR modela directamente el cuantil de interés. Y finalmente, los modelos de aprendizaje supervisado, en particular una red neuronal conocida como red neuronal recurrente (RNN) la cual ha tenido un notable éxito cuando se trata de predecir datos secuenciales.

El retorno () para el periodo lo definimos como y ,donde corresponde al precio.

El hace referencia al Value-at-Risk con un 1- de confianza. Lo que nos interesara sera modelar el del día siguiente al que nos encontramos condicional en la información que tenemos hoy. Lo que será denotado , donde se refiere a la información en t.

## Normal

Aquí suponemos que , lo que hace sencillo estimar el dado que sera constante. Por lo tanto, vendrá dado por:

Donde es el cuantil de la normal estándar.

Una forma de calcular y es reemplazarlas por sus contra-parte muéstrales, que es la forma que se ocupara en este trabajo. Una forma general de definir los parámetros es

y ,

donde si entonces, se usa toda la información histórica lo que produce una línea eventualmente plana que responderá solo ante presencia de outliers en el caso del arribo de nueva información. Si la idea es permitir a la media y la varianza depender del tiempo, pero la distribución subyacente sigue siendo la normal. Si la serie es estacionaria en su forma fuerte, el modelo rolling debiese converger hacia el otro modelo.

El problema es que en la literatura está bien documentada la asimetría y exceso de curtosis, por ejemplo, en Engle (1982) y Bollerslev (1986-87) lo que indicaría que el resultado no debiese ser el óptimo dado que no logramos capturar esos hechos estilizados.

## GARCH

Los modelos GARCH son una familia de modelos que permite modelar varianzas heterocedasticas, por eso el nombre de Generalized AutoRegresive Conditional Hetorocedasticty. Engle (1982) desarrolla un primer modelo llamado ARCH, que es un caso particular del GARCH solo incorporando el componente autoregresivo del error. Por lo tanto, el modelo será de particular relevancia cuando la varianza de la predicción se correlacione con el error pasado Engle (1982). Si asumimos que la media es constante para todo , podemos representar un ARCH(q), donde la varianza condicional sigue un proceso lineal en el error, con las siguientes ecuaciones

donde , corresponde a la información en a la distribución normal y la media.

Posterior a esto, Bollerslev (1986) los generaliza permitiendo incorporar rezagos de la varianza condicional, mejorando problemas de autocorrelación de la varianza. Entonces si continuamos con el ejemplo, para un GARCH(p,q) la varianza condicional vendrá dada por

donde .

Ding et al. (1993) muestran una persistencia en la autocorrelación de , lo que explicaría la asimetría de la volatilidad condicional a los retornos positivos y negativos. Para solucionarlo propone el siguiente modelo para la varianza condicional conocido como el APARCH(p,q),

donde es un parámetro más a estimar. y .

En Baillie et al. (1996) se muestra la relevancia de considerar la persistencia de los shocks mediante la integración de la serie, pero se señala que un modelo de integración completa, IGARCH, o un no integrado, GARCH, son a lo mejor muy extremos y hace que perdamos información relevante. Una solución es usar integración fraccional, esto permite incorporar la presencia de persistencia de shock que decaen de forma hiperbólica a diferencia de la forma exponencial del GARCH, y infinita persistencia del IGARCH, como se muestra en la figura 2.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

**Figura 2:** Comparación de la caída hiperbólica del FIGARCH (roja) versus la caída exponencial que presenta el GARCH (azul) tras un shock. Se realizo con datos simulados.

Estos modelos se conocen como FIGARCH(p,d,q), y siguiendo la notación en Degiannakis et al. (2013) tenemos que la varianza condicional será

donde son operadores de rezagos de orden p y q respectivamente y un operador de orden z es definido como . El operador de diferenciación fraccional viene definido por la siguiente función

donde , notar que no incluye 0 ni 1 porque en ese caso serian un GARCH y IGARCH respectivamente.

Hasta este punto se trató con la heterocedasticidad, autocorrelación de la varianza, asimetría en las respuesta y memoria de largo plazo, pero aún existe el problema de que retornos presentan colas más anchas, exceso de curtosis, que una normal. En un trabajo más aplicado Bollerslev (1987) extiende el GARCH para permitir una distribución-t estandarizada

donde corresponde a la distribución-t con grados de libertad.

Hasta ahora solo hemos asumidos distribuciones simétricas, y sabemos a partir de los hechos estilizados señalados que valores más grandes son probables a ubicarse en el lado negativo de la distribución. Esto provoca que la distribución se concentre más en ese lado. Una forma de solucionarlo es asumir una distribución-t asimétrica, el cual se encuentra en varios trabajos como un modelo base difícil de superar de manera consistente, ver Braione et al. (2016) o Halbleib et al. (2012).

De lo anterior, notamos que tenemos diferentes combinaciones para el modelo a ocupar para la varianza como la distribución que se asume, a esto falta agregarle el modelo para la media, en este trabajo consideramos solo modelos con media igual a 0, constante y autoregresivas.

Ósea respectivamente.

Por lo tanto, la familia de modelos GARCH será una combinación de modelos para la media, más un modelo para la varianza y finalmente un supuesto para la distribución de los residuos.

Dado esto para obtener una predicción a partir de estos datos ocupamos la siguiente formula.

Donde tanto la media y la varianza son obtenidas a través de predicciones condicional en la información presente, y representa el cuantil de la distribución teórica de retornos escogida.

Finalmente, la estimación de los parámetros de los modelos se hace mediante la maximización de la función likelihood, donde básicamente es ajustar unos datos a una distribución mediante la optimización de ciertos parámetros. Si la distribución es normal y simétrica es trivial armar la función de optimización a partir de las densidades de cada observación. En caso de que la distribución es no simétrica, o no gaussiana, necesitamos maximizar la función de quasi-likelihood que implica aproximar una densidad de probabilidad a la data. Las propiedades del likelihood para un GARCH(p,q) se pueden encontrar en Ling y Lin (1998) y las propiedades del quasi-likelihood se pueden encontrar en Ling y Lin (1997) aplicado a un FIGARCH(p,d,q).

Esta familia de modelo se encuentra en la libraría arch\_model [[2]](#footnote-2) de Python.

## CAVIAR

Fue desarrollado por Engle y Mangananelli (2004), es un modelo de cuantiles autoregresivo que optimiza los parámetros mediante una regresión del tipo cuantiles (Koenker and Bassett,1978), lo que permite modelar el VaR sin tener que asumir una distribución.

En el modelo lineal propuesto por Koenker and Bassett (1978) una función del tipo es la que busca ser ajustada y se minimiza la siguiente función, llamada pinball loss, para obtener los parámetros de la regresión para el cuantil tenemos que

Segun Engle et al. (2004) la autocorrelación en la varianza debiese de alguna forma ser transmitida hacia el dado que estos se encuentran íntimamente ligados. Para esto propone en principio un modelo general

donde simplificamos la notación, solo por un tema practico Notamos que el proceso involucra un termino autoregresivo y otro que permite mapear desde un numero arbitrario de rezagos de los retornos o incluso, en su forma más general, que no es tratada en este trabajo ni en el trabajo de Engle et al. (2004), podría incluir cualquier otro observable.

En particular Engle et al. (2004) presenta cuatro especificaciones cada una respondiendo a un hecho estilizado distinto.

El primero es que el VaR siga un proceso adaptativo, donde cada vez que hagamos un hit, ósea que el retorno real sea menor al VaR estimado, se aumente el VaR y cada vez que no lo pase se empiece a disminuir ligeramente.

Adaptive (AD):

El problema es que solo aprende de retornos negativos grandes, que es cuando corrige, por lo tanto, retornos positivos o cercanos al VaR no aportan. Notar además que tiene coeficiente igual a la unidad en su rezago.

Después tenemos tres procesos con reversión a la media, pero que difieren en la forma en impacta los retornos pasados.

Symmetric absolute value (SAV):

En este modelo el impacto de los retornos pasados es simétrico, está presente la idea que se intenta incorporar en modelos de volatilidad de que los retornos absolutos pasados están relacionados a la dispersión de la distribución de retornos actual, Bollerslev (1986).

Naturalmente, el otro modelo implica deshacerse de ese supuesto e incorporar la asimetría de la respuesta. Nuevamente nos encontramos con lo hecho por Ding et al. (1993) que menciona que retornos pasados se encuentran asociados a la dispersión actual de la distribución, pero el impacto que tiene los negativos es distinto a la de los positivos.

Asymmetric Slope (AS):

Donde y .

El tercero es un modelo del tipo GARCH(1,1).

Indirect GARCH(1,1) (IG):

Para obtener los parámetros en cada uno de esos modelos se siguió el algoritmo propuesto el trabajo en Engel et al. (2004).

1. Inicializamos una matriz de parámetros, con números aleatorios desde una distribución uniforme.
2. Obtenemos una cantidad(k) de posibles candidatos a parámetros a partir del ajuste en el modelo.
3. Optimizamos k veces usando cada unos de los k vectores de parámetros como vector inicial, ayuda con la no convexidad y para no quedar atrapado en un mínimo local. El método de optimización es el de Nelder-Mead.
4. De las k optimizaciones, de las que tengan una convergencia satisfactoria lo que se ve chequeando la matriz hessiana, se escoge la con el menor residuo.

Actualmente no existe una librería en Python que lo haga directo, por lo que el código fue realizado por el autor.

## RNN

Redes neuronales representan un mapeo no lineal F(X) sobre datos de potencialmente alta dimensionalidad, por medios de capas ordenas de manera jerárquica Dixon et al. (2020) y Murphy (2022). La más conocida es la feedforward networks, que es una secuencia de capas que se ordenan vía composición, ocupando la notación de Dixon et al. (2020) tenemos que la forma funcional de una Deep feedforward networks será igual a

donde es una función semi-afine, univariada, no lineal y continua. son los pesos y los sesgos.

Los parámetros, en general, son obtenidos vía propagación hacía atrás que implica ir tomando las derivadas parciales desde la capa mas externa a la interna usando regla de la cadena, la optimización es con el algoritmo de descenso de gradiente estocástico.

### Red Neuronal Recurrente

Son redes neuronales hechas para predecir secuencia, basta mostrar la autocorrelación para entender que las series de tiempo son un caso particular. Puede ser vista como una generalización no lineal de los modelos autoregresivos, Dixon (2020).

donde . Entonces no es mas que un mapeo no lineal de rezagos de una variable. Esto implica que en caso de idd, recuperamos la arquitectura de *feedforward.*

La función de perdida en general es el error cuadrático medio,

lo que permite modelar el valor esperado, para modelar un cuantil una modificación natural, siguiendo lo hecho por Petneházi (2021), es modificar la función de perdida para que esta sea la *pinball loss* presentada junto a los modelos CAVIAR. Entonces podemos predecir al VaR como

A la RNN con función de perdida tipo *pinball loss,* llamaremos modelos QRNN.

Diagrama

Descripción generada automáticamente

La siguiente pregunta es como escogemos los datos de entrada y la arquitectura de la red. Para la primera Dixxon et al. (2020) recomienda usar hasta el último rezago significativo en el gráfico de autocorrelación parcial. Para la segunda debemos seguir el sesgo-varianza *trade-off ,* ver Dixxon et al.,2020 y Murphy (2022),en general debiésemos necesitar unas capas sin muchas neuronas y solo un par de capas densas dado lo que menciona Ormaniec et al. (2022), esto es problema de muestra pequeña y altamente no lineal.

XXXXX Foto de la arquitectura

### Predicción conforme

# APLICACIÓN

## Datos

## Métricas

## Resultados

# CONCLUSIONES

# REFERENCIAS BIBLIOGRÁFICAS

Utilizar norma APA (un solo listado – sin separar por tipo de fuente – en orden alfabético, y sangría francesa.

Para mayores detalles revisar: <http://biblioteca.usach.cl/sites/biblioteca/files/documentos/dhi-apa_0.pdf>

# **ANEXOS**

En nueva página, opcional. Si no tiene anexos, eliminar la página.

Los anexos corresponden a aquellos materiales encontrados en la búsqueda de información y que son relevantes para la tesis.

# APÉNDICES

En nueva página, opcional. Si no tiene apéndices, eliminar la página.

Los apéndices son materiales elaborados por el autor para la investigación (encuestas, entrevistas).

1. Todo el código que replica los resultados y modelos ocupados, además de las tablas y gráficos se encuentra en XXXXXX insertar github [↑](#footnote-ref-1)
2. <https://arch.readthedocs.io/> para más detalles sobre la librería. [↑](#footnote-ref-2)