Valores extremos y GPD con aplicación a portafolios de inversión

Isauro Medina Murúa A01193935
Gerardo Montoya A01282506
Alejandro Mejía A01173117
Julio Valiñas A01540158
Mario Modiano A01023625

28 de septiembre, 2018

Resumen

El propósito de este trabajo es estimar el Valor en Riesgo (VaR), media y varianza de tres portafolios de inversión en la frontera eficiente, cada uno con tres distintas varianzas. Cabe mencionar, este análisis tiene una sensibilidad numérica considerable, por lo cual se sugiere tomar las conclusiones como complementarias a otros resultados. Como futuro trabajo, podrían explorarse algunas ideas para mejorar la estabilidad del procedimiento.

Se reporta el VaR de tres portafolios, cada uno con percentiles 1 %, 0.1 % y 0.01 %. Se utiliza *points over threshold* (POT) y *Generalized Pareto Distribution* (GPD) para estimar el valor en riesgo.

La información fue recuperada de Yahoo Finance, en donde se encuentran datos históricos del valor de las acciones de: Grupo Alfa, Gentera, Grupo Carso y OHL México. Se utilizan los métodos media-varianza de Markowitz y elementos de la teoría de valores extremos (POT y GDP), para poder estimar los riesgos y la rentabilidad esperada, en base a los portafolios seleccionados, que se encuentran sobre la frontera eficiente (frontera de Pareto entre media-varianza).

Abstract

The purpose of this work is to estimate the Value at Risk (VaR), the variance and the mean of three investment portfolios lying on the efficient frontier, each with three different variance values. It is worth noting, this analysis is rather unstable, so it suggested to take conclusions as complementary to other results. An exploration of possibilities to improve stability could be considered as future work.

The VaR of three portfolios is shown each with the percentiles 1 %, 0.1 % and 0.01 % are used. To estimate the tail distribution, *points over threshold* (POT) and *Generalized Pareto Distribution* (GPD) models are fit to finally estimate the value at risk.

The data used is recovered from Yahoo Finance, where we found and selected the following stocks: Grupo Alfa, Gentera, Grupo Carso and OHL México. Markowitz's mean-variance method is used as well as and elements of extreme value theory (POT and GDP), in order to determine the risks and expected return, based on efficient portfolio combinations found on the efficient frontier (Pareto frontier between mean and variance).

1. Introducción

El análisis estadístico de los precios de acciones es un área con una cantidad de trabajo extensa. Generalmente, lo que más atención recibe es un análisis basado en series de tiempo. En este enfoque, dada una secuencia de n variables aleatorias x_n se utilizan procesos para estimar la distribución de un dato t unidades en el futuro x_{n+t}. Sobre este método destaca la metodología Box-Jenkins [4].

Sin embargo, estos métodos suelen dar buenas aproximaciones de la distribución de probabilidad (PDF) de x_{n+t} cerca de la media de x_{n+t}, y obtienen peores resultados en las colas de esta distribución.

Por esta razón, para predecir valores extremos es más adecuado un modelo específico para las colas. Incluso dentro de las distribuciones de colas anchas, existen diversos tipos [5]. En este trabajo, se discutirá únicamente la distribución de Pareto generalizada (GPD) ya que su uso es más común en análisis de portafolios [5].

El análisis se hará en dos pasos:

- Primero, seleccionar un portafolio eficiente (método media-varianza) [2].
- Después, estimar el Valor en Riesgo (VaR) utilizando GPD aplicada a las colas (puntos sobre umbral) [3].

Para este trabajo, se utilizan los datos históricos de las empresas seleccionadas; éstos disponibles al público proporcionados por Yahoo finance.

2. Contexto

2.1. Análisis media-varianza [2]

En esta sección, se da una descripción básica del método media-varianza de Markowitz, también conocido como teoría moderna del portafolio.

Se tienen N empresas diferentes para invertir una cantidad fija de capital (en este caso, $N = 4$). Como este capital debe dividirse entre las 4, definimos una variable aleatoria G que será una combinación lineal de las $N = 4$ empresas a invertir:

$$G = \sum_{i=1}^{N} X_i = w_1 X_1 + w_2 X_2 + w_3 X_3 + w_4 X_4$$

Donde X_i es la variable aleatoria que toma los valores de los rendimientos aritméticos de la empresa i y los pesos w_i deberán sumar 1.
Dadas las medias y varianzas de los datos históricos de estas empresas, la idea del método media-varianza es precisamente calcular la media y la varianza de la distribución de G. Para esto, simplemente se utilizan las propiedades del operador de esperanza y varianza.

El operador de esperanza es un operador lineal: [AGREGAR REF]

$$E[G] = w_1 E[X_1] + w_2 E[X_2] + w_3 E[X_3] + w_4 E[X_4]$$

Los estimados de las esperanzas de cada X_i son simplemente el promedio de los datos históricos de la empresa i-ésima. La varianza no es un operador lineal, sin embargo, la varianza de una suma puede escribirse de la siguiente forma [2 AGREGAR REF]:

$$Var[G] = \sum_{i=1}^{4} w_i^2 Var[X_i] + \sum_{i \neq j}^{4} w_i w_j Covar[X_i, X_j]$$

El cálculo de la varianza de G puede realizarse más simplemente con una matriz de covarianzas. Por ejemplo, para nuestros caso, X_1 es Alfa, X_2 es Gentera, X_3 es Grupo Carso, y X_4 es OHL México. Como la matriz de covarianzas es simétrica, todos los valores aparecen dos veces.

Entonces, para calcularse la suma de todas las covarianzas posibles, basta con sumar todos los términos diferentes de 1 en la matriz de covarianzas. Como cada uno aparece exactamente 2 veces, cada valor puede simplemente multiplicarse por 2.

Aplicando estas propiedades a la distribución G en nuestros datos, obtenemos las siguientes fórmulas para calcular la media y varianza de un portafolio:

$$E[G] = w_1 \overline{X_1} + w_2 \overline{X_2} + w_3 \overline{X_3} + w_4 \overline{X_4}$$

donde $\overline{X_i}$ es el promedio de la empresa i. Para la varianza:

$$Var(G) = \sum_{i=1}^{4} w_i^2 Var[X_i] + \dots$$

$$2w_1 w_2 Cov(X_1, X_2) + 2w_1 w_3 Cov(vX_1, X_3) + \dots$$

$$2w_1 w_4 Cov(X_1, X_4) + 2w_2 w_3 Cov(X_2, X_3) + \dots$$

$$2w_3 w_4 Cov(X_3, X_4)) + 2w_2 w_4 Cov(X_2, X_4)$$

2.2. Valores extremos (GEV y GPD) y valor en riesgo

En esta sección, se discuten brevemente algunas ideas útiles de la teoría de valores extremos.

Para los análisis sobre el cuerpo de una distribución, un ajuste de una curva normal es suficiente. Pero en muchas áreas de aplicación (las finanzas incluidas), al alejarse de la media los datos empíricos no decaen exponencialmente como lo hace una distribución normal. A grandes razgos, se dice que una distribución es de *colas pesadas* al caso en que una cola decae más lento que una curva exponencial [5].

Una vez que se ajusta un modelo de PDF a los valores extremos, se pueden realizar predicciones como el VaR, que será de interés en esta aplicación.

Una cuestión que podría considerarse primordial de un valor extremo es: ¿cómo saber qué valores son extremos?. En este trabajo se mencionarán dos, desarrollando la segunda para la aplicación.

La primer forma de definir un valor extremo es localmente. Se realiza una partición del conjunto de datos históricos, por ejemplo en semanas. En este caso, se definen los valores extremos como los máximos de cada partición. Esto en teoría de valores extremos se conoce como *máximos de bloques*, o *block maxima*. Es común encontrar análisis de este tipo usando los máximos de cada año. Se sabe que al extraer esta clase de máximos, los nuevos valores siguen una distribución generalizada de valores extremos [5], o GEV por sus siglas en inglés. Esta es la distribución clásica en teoría de valores extremos. Sin embargo, se menciona únicamente como contexto y no será utilizada.

La segunda forma de definir un valor extremo es globalmente. Se toman los n valores más grandes en los datos históricos, y se ajusta una distribución a estos datos. O bien, se escoge un umbral donde los puntos arriba de éste se consideran como extremos. Este enfoque se conoce como *Points Over Threshold (POT)*. Igual que en el caso de máximos de bloques, se sabe que los puntos en POT siguen una distribución generalizada de Pareto [5], o GPD por sus siglas en inglés. Su PDF es la siguiente:

$$f(x) = \frac{1}{\sigma}(1 + k\frac{x-\mu}{\sigma})^{-1-\frac{1}{k}}, k \neq 0$$
$$= (\frac{1}{\sigma})e^{-\frac{x-\mu}{\sigma}}, k = 0$$

La segunda se obtiene tomando el límite como k tiende a 0.

3. Método

3.1. Datos y preprocesamiento

Fechas Las fechas que se utilizaron para este análisis fueron escogidas de forma que sean exactamente las mismas fechas para cada registro. Esto implica que para el caso que una serie de datos sea más pequeña que las demás, las series más grandes solamente se pueden utilizar en fechas coincidentes con las pequeñas.

Este fue precisamente el caso, en el que los datos de Alfa y Grupo Carso tienen más registros que OHL México y que Gentera, es decir, llevan más tiempo cotizando en la bolsa. Entonces, el primer paso fue comparar las fechas disponibles en Yahoo Finance, y encontramos que las fechas en las que se coincide son del 30 de diciembre del 2010 al 21 de noviembre del 2018. Estos fueron los datos de la muestra para cada empresa.

Transformación Se realizaron únicamente dos transformaciones a los datos: reemplazar los *missing values* con el valor del día anterior, y después calcular el rendimiento aritmético.

3.2. Media-varianza

Esta ecuación puede resolverse con muchos métodos. Por razones de simplicidad, recorremos al método más sencillo: muestreo aleatorio. Consiste en generar 4 números pseudoaleatorios que sumen 1, y utilizarlos como pesos w_1, w_2, w_3, w_4. Calculamos entonces la media y varianza de éste portafolio, y lo ponemos en una gráfica.

Este conjunto de puntos va a generar una región de portafolios posibles en el plano varianza-media. Esta gráfica se muestra en la figura (1).

El objetivo ahora es encontrar 3 puntos sobre la frontera eficiente de portafolios estimada, uno con bajo riesgo, uno con medio, y otro con alto riesgo. Aquí la varianza es una medida del riesgo financiero.

Escogimos 3 puntos que nuestra simulación catalogó como posibles mostrados en la figura (2).

3.3. Valores extremos

El uso de máximos de bloques es poco común en el análisis financiero [5]. Por esto, se utilizará otra distribución para colas discutida antes: GPD.

Algo también valioso de este análisis es que la distribución es una teórica para finanzas [5]. Esto significa que incluso cuando existan franjas sin datos en la base de datos, la distribución puede realizar interpolación y extrapolación.

En el método de *points over a threshold*,se ajusta una distribución GPD. Esto se realiza de la siguiente manera:

1. Utilizar solamente los datos de la cola de las pérdidas

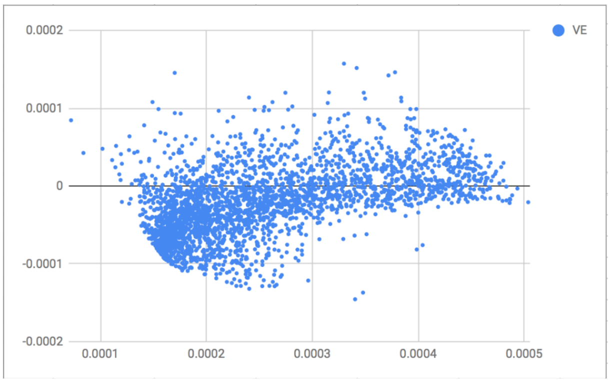
2. Realizar un análisis exploratorio de los datos: gráficas QQ-plot, espectrogramas, histogramas, pruebas Anderson-Darling, etc.

3. Encontrar el número de puntos extremos, utilizando el estimador de Hill [1].

4. Ajustar la distribución GPD a nuestros datos (obtener parámetros).

5. Encontrar los percentiles de la distribución (valor en riesgo).

Toda esta sección se realizó con el uso del software Extremes [3] y también se utilizó MATLAB.



(a) Figura (1)

Menor riesgo encontrado			
W ALFAA	W GENTERA	W GCARSOA1	W OHLMEX
0.0001622183853	0.1812119802	0.03367559943	0.784950202
Mediana del riesgo			
W ALFAA	W GENTERA	W GCARSOA1	W OHLMEX
0.008042970582	0.2518725645	0.07564950896	0.664434956
Mayor riesgo encontrado			
W ALFAA	W GENTERA	W GCARSOA1	W OHLMEX
0.07969173717	0.1225099145	0.009471757924	0.7883265904

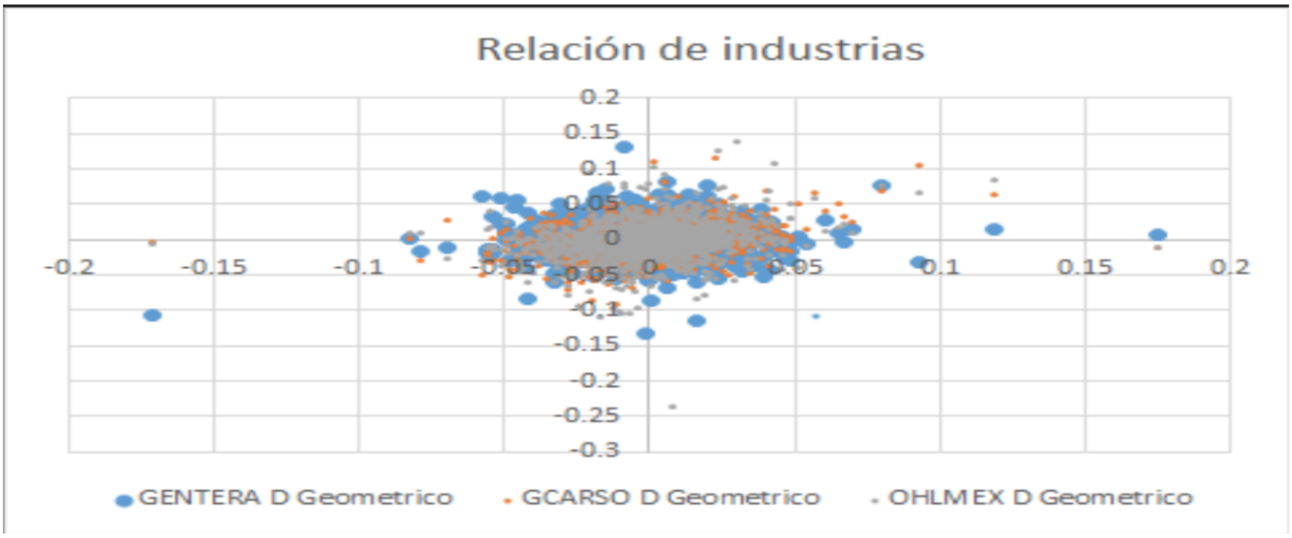
(b) Figura (2)

4. Resultados

En la figura (1), es importante notar que una gran densidad de portafolios queda bajo el eje de varianza. Esto significa que tienen un valor esperado negativo; pérdidas. Esto es de esperarse, ya que 3 de las 4 empresas seleccionadas tienen un promedio negativo en los datos históricos. Esto se refleja en una gráfica con los rendimientos geméricos de las 4 empresas superpuestos en una gráfica de correlación con las acciones de ALFA.

Existen diferencias significativas entre el Valor en Riesgo según los datos históricos y el que fue calculado con valores extremos mediante una extrapolación prospectiva, las cuales son de entre 7 mil pesos hasta 11 mil pesos. Con esto, comprobamos que en efecto hay diferencias económicas cuando analizamos el VaR en base a los datos historicos (los rendimientos geométricos y aritméticos) y con valores extremos. Los estimados para tres niveles de VaR de cada uno de los tres portafolios se muestran en la tabla.

Relación VaR portafolio						
Nivel de riesgo/Tipo de VaR	VaR Según datos historicos			VaR con valores extremos		
	1%	5%	10%	1%	0.1%	0.01%
Portafolio de bajo riesgo	$ 50,300.25	$ 27,700.92	$ 19,860.66	$ 57,713.13	$ 76,401.14	$ 113,645.23
Portafolio de medio riesgo	$ 41,267.81	$ 23,650.09	$ 16,835.03	$ 53,180.33	$ 66,437.37	$ 93,778.30
Portafolio de alto riesgo	$ 49,406.54	$ 27,838.71	$ 19,955.41	$ 56,823.59	$ 75,663.74	$ 114,661.53



5. Reflexiones finales

En conclusión, debido a la volatilidad y el bajo rendimiento de las acciones de nuestro portafolio, el VaR es alto en proporción al capital invertido. Tanto el VaR como el GPD le pueden ser de gran utilidad a un inversionista que busque un portafolio de inversión que vaya de acuerdo con su actitud al riesgo, ya que dichas herramientas le proporcionan la opción de un portafolio de bajo, mediano y alto riesgo. También, el inversionista podrá saber con cierta exactitud, el máximo capital que puede llegar a perder con sus inversiones y sus respectivas probabilidades.

Otro punto importante es que un inversionista debe tomar siempre en cuenta los valores extremos y los cisnes negros, ya que son dos elementos que le pueden provocar grandes pérdidas a un banco o inversionista. Por lo tanto se deben analizar con gran meticulosidad. Si nosotros fuéramos los inversionistas, no invertiríamos en un portafolio conformado por las acciones que tenemos, ya que la mayoría de los portafolios tienen esperanza negativa.

Referencias

[1] Hill B.M. (1975) *A simple general approach to inference about the tail of a distribution*. Ann. Stat., vol. 3, pg. 1163–1174.

[2] Markowitz, H.M. (March 1952). "Portfolio Selection". The Journal of Finance. 7 (1): 77–91.

[3] Extremes (Version 2,0) [Computer software]. (2009). Recuperado 25 de septiembre de 2018, de http://extremes.gforge.inria.fr/

[4] G. E. Box, , G. M. Jenkins, , G. C. Reinsel,G. M. Ljung, , and G. M Ljung. (2015). *Time Series Analysis : Forecasting and Control*.

[5] Longin, F. (2016). Extreme events in finance : a handbook of extreme value theory and its applications. Recuperado de https://0-ebookcentral.proquest.com.millenium.itesm.mx