Entrega: curso de datos extremales

Laura Montaldo, CI: 3.512.962-7

2024-02-06

Índice

Resumen

Your abstract goes here.

Motivación y objetivo del estudio

(Gonzalo y Segura 2024) Los índices de $S\&P$ son una familia de índices de renta variable[1] diseñados para medir el rendimiento del mercado de acciones en Estados Unidos que cotizan en bolsas estadounidenses. Ésta familia de índices está compuesta por una amplia variedad de índices basados en tamaño, sector y estilo. Los índices están ponderados por el criterio *float-adjusted market capitalization* (FMC). Además, se disponen de índices ponderados de manera equitativa y con límite de capitalización de mercado, como es el caso del $S\&P$ 500. Este este sentido, el $S\&P$500 entraría en el conjunto de índices ponderados por capitalización bursátil ajustada a la flotación (ver $S\&P$ Dow Jones Indices). El mismo mide el rendimiento del segmento de gran capitalización del mercado estadounidense. Es considerado como un indicador representativo del mercado de renta variable de los Estados Unidos, y está compuesto por 500 empresas constituyentes.

Se busca crear un indicador de una posible crisis bursátil. Como variable de referencia de toma la relación de precios al cierre de ayer sobre la de hoy

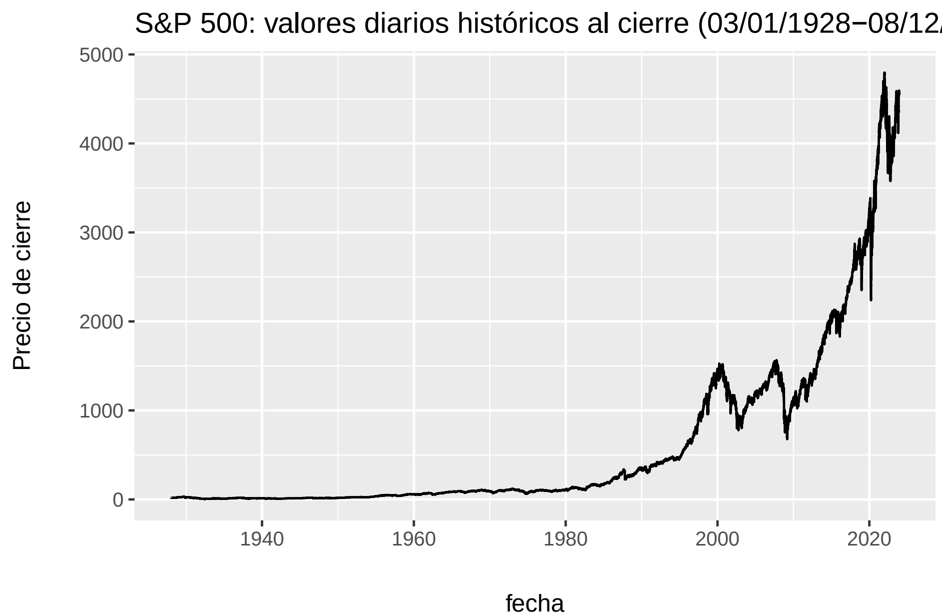
$$Indicador_t = \frac{Precio_{t-1}}{Precio_t}, \quad \text{para } t = 1, ..., T \tag{1}$$

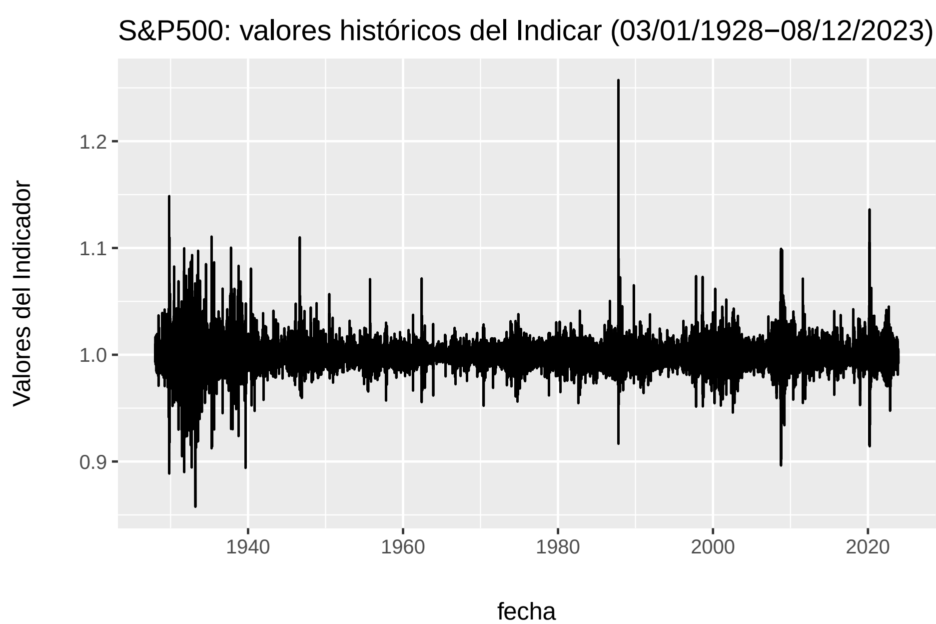
Interpretación del Indicador:

- Si el $Indicador_t \leq 1$, el precio de cierre de hoy es mayor o igual que el de ayer, lo cual podría ser considerado una señal positiva.
- Si el $Indicador_t > 1$, el precio de cierre de hoy es menor que el de ayer, lo cual podría considerarse una señal de alerta.

En las siguiente figura @ref(fig:plot1) se muestra la evolución histórica desde la fecha 03/01/1928 hasta 08/12/2023 del precio al cierre del día del indicar S&P 500.

[1] En inglés se llaman equity indices


S&P 500: valores diarios históricos al cierre (03/01/1928−08/12/
Precio de cierre
5000
4000
3000
2000
1000
0
1940
1960
1980
2000
2020
fecha


S&P500: valores históricos del Indicar (03/01/1928−08/12/2023)
Valores del Indicador
1.2
1.1
1.0
0.9
1940
1960
1980
2000
2020
fecha

Marco Teórico

Teoría asintótica clásica y las distribuciones extremales y sus dominios de atracción

Siguiendo [?] se dice que tenemos datos extremos cuando cada dato corresponde al máximo o mínimo de varios registros. Son un caso particular de evento raro o gran desviación respecto a la media.

Asumiremos que nuestros datos son *iid* (independientes e idénticamente distribuidos, son dos suposiciones juntas). Esta doble suposición suele no ser realista en aplicaciones concretas (ninguna de sus dos componentes, incluso) pero para comenzar a entender la teoría clásica, la utilizaremos por un tiempo.

Si tenemos datos $X_1, ..., X_n$ *iid* con distribución F, entonces $X_n^* = max(X_1, ..., X_n)$ tiene distribución F_n^* dada por $F_n^*(t) = F(t)_n$. Si conocemos la distribución F conoceríamos la distribución F_n^* , pero en algunos casos la lectura que queda registrada es la del dato máximo y no la de cada observación que dio lugar al mismo, por lo que a veces ni siquiera es viable estimar F. Pero aún en los casos en que F es conocida o estimable, si n es grande, la fórmula de F_n^* puede resultar prácticamente inmanejable. En una línea de trabajo similar a la que aporta el Teorema Central del Límite en la estadística de valores medios, un teorema nos va a permitir aproximar F_n^* por distribuciones más sencillas. Este es el Teorema de Fischer-Tippet-Gnedenko (FTG, para abreviar) que presentaremos en breve.

Como $X_1, ..., X_n$ iid, definimos $Y_i = -X_i$ para todo valor de i, entonces $Y_1, ..., Y_n$ iid y además $min(X_1, ..., X_n) = -max(Y_1, ..., Y_n)$ la teoría asintótica de los mínimos de datos iid se reduce a la de los máximos, razón por la que nos concentramos aquí en estudiar el comportamiento asintótico de los máximos exclusivamente.

Las distribuciones extremales

Las distribuciones extremales son tres: la distribución de Gumbel; la distribución de Weibull; la distribución de Fréchet.

Distribución de Gumbel

Se dice que una variable tiene distribución de Gumbel si su distribución es:

$$\Lambda(x) = exp\{-e^{-x}\} \quad \text{para todo } x \text{ real}$$

Distribución de Weibull

Se dice que una variable tiene distribución de Weibull de orden $\alpha > 0$ si su distribución es:

$$\Psi_\alpha(x) = \begin{cases} exp{-}(-x)^\alpha & si\ x < 0 \\ 1 & \text{en otro caso} \end{cases}$$

Distribución de Fréchet

Se dice que una variable tiene distribución de Fréchet de orden $\alpha > 0$ si su distribución es:

$$\Phi_\alpha(x) = \begin{cases} exp\{-x^{-\alpha}\} & si\ x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Teorema 1: Relaciones entre las versiones standard de las distribuciones extremales

X tiene distribución $\Phi_\alpha(x)$ si y sólo si $(-1/X)$ tiene distribución $\Psi_\alpha(x)$ si y sólo si $log(X^\alpha)$ tiene distribución Λ.

Teorema 2: Algunos datos de las distribuciones extremales

Parte 1 Si X tiene distribución $\Lambda^{(\mu,\beta)}$ entonces tiene:

a) Valor esperado: $E(X) = \mu + \beta\gamma$, donde γ es la constante de Euler-Mascheroni, cuyo valor aproximado es 0.5772156649.

b) Moda: μ

c) Mediana: $\mu - \beta \log(\log 2) \approx \mu - 0.36651\beta$.

d) Desviación estándar: $\beta\pi\sqrt{6} \approx 1.2825\beta$.

e) Si $X^+ = \max(X, 0)$, entonces $E(X + k)$ es finito para todo valor de k natural.

f) Para simular computacionalmente X, se puede tomar U uniforme en $(0, 1)$ y hacer $X = \mu - \beta \log(-\log U)$.

Parte 2 Si X tiene distribución $\Psi_\alpha^{(\mu,\beta)}$ entonces tiene:

a) Valor esperado: $E(X) = \mu + \beta\Gamma(1 + 1/\alpha)$.

b) Moda: μ si $\alpha \leq 1$ y $\mu - \beta\{(\alpha - 1)/\alpha\}^{(1/\alpha)}$ si $\alpha > 1$.

c) Mediana: $\mu - \beta \log(2)^{(1/\alpha)}$.

d) Desviación estándar: $\beta\{\Gamma(1 + 2/\alpha) - \Gamma(1 + 1/\alpha)^2\}^{1/2}$.

Parte 2 Si X tiene una distribución $\Phi_\alpha^{(\mu,\beta)}$ entonces se tiene:

a) Valor esperado: $E(X) = \mu + \beta\Gamma(1-1/\alpha)$ si $\alpha > 1$, ∞ en caso contrario.

b) Moda: $\mu + \beta\Gamma(1-1/\alpha)$ si $\alpha > 1$.

c) Mediana: $\mu + \beta \log(2)^{(-1/\alpha)}$.

d) Desviación estándar: $\beta|\Gamma(1-2/\alpha) - \Gamma(1-1/\alpha)^2|$ si $\alpha > 2$, ∞ si $1 < \alpha \leq 2$.

Teorema 3: Fischer-Tippet-Gnedenko (FTG)

Si $X_1, ..., X_n$ *iid* con distribución F “continua”, llamamos F_n^* a la distribución de $max(X_1, ..., X_n)$ y n es grande, entonces existen μ real y $\beta > 0$ tales que alguna de las siguientes tres afirmaciones es correcta:

1) F_n^* se puede apromixar por la distribución de $\mu + \beta Y$ con Y variable con distribución Λ.

2) Existe $\alpha > 0$ tal que F_n^* se puede aproximar por la distribución de $\mu + \beta Y$ con Y variable con distribución Φ_α.

3) Existe $\alpha > 0$ tal que F_n^* se puede aproximar por la distribución de $\mu + \beta Y$ con Y variable con distribución Φ_α.

Lo anterior equivale a decir que la distribución del máximo de datos “continuos” e *iid*, si n es grande, puede aproximarse por una Gumbel, una Fréchet o una Weibull. Una aproximación será válida dependiendo de la distribución de F. En este sentido, cuando F sea normal entonces F_n^* se puede aproximar como una Gumbel. Cuando F sea uniforme, se puede aproximar F_n^* como una Weibull y cuando F sea Cauchy entonces F_n^* se puede aproximar por una Fréchet.

Gonzalo, Perera, y Angel Segura. 2024. *Notas del curso*.