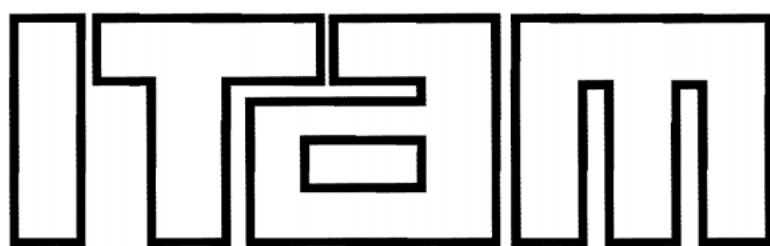


INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



Verosimilitud Perfil para Tasa de Cambio Efectiva Real en  
México

**T E S I S**  
QUE PARA OBTENER EL TÍTULO DE  
LICENCIADA EN ACTUARÍA  
P R E S E N T A  
**MARÍA FERNANDA ALCALÁ DURAND**

**ASESOR: DR. JOSÉ LUIS FARAH IBÁÑEZ**



Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada “**Verosimilitud Perfil para Tasa de Cambio Efectiva Real en México**”, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación.

María Fernanda Alcalá Durand

---

FECHA

---

FIRMA



---

---

# Agradecimientos

Linda, Refugio, Federico, Gustavo: para los egipcios, el cielo se sostenía en cuatro pilares.

Tapen. Farah. Leticia.

Y todos los demás, porque por chiquita que haya sido su ayuda, recuerden que

$$\sum_{x=1}^{\infty} \frac{1}{x}$$

se va a infinito por los valores pequeños. ¡Gracias!



---

# Índice general

<b>Agradecimientos</b>	<b>iv</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Máxima Verosimilitud</b>	<b>5</b>
2.1. Introducción . . . . .	5
2.2. Conceptos básicos y definición . . . . .	6
2.2.1. Métodos comúnmente utilizados para inferencia estadística . . . . .	7
2.3. El método de Máxima Verosimilitud . . . . .	8
2.3.1. Intervalos de verosimilitud . . . . .	10
2.3.2. Propiedades de los EMV . . . . .	12
<b>3. Verosimilitud Perfil</b>	<b>15</b>
3.1. Definición . . . . .	15
3.1.1. Información: . . . . .	17
3.2. Ventajas con respecto a Verosimilitud . . . . .	18
3.3. Ejemplos . . . . .	19
3.4. Limitaciones . . . . .	21
<b>4. El Problema de Maximización y las Envolventes</b>	<b>23</b>
4.1. El Teorema de la Función Implícita . . . . .	23
4.2. Envolventes de una familia de funciones . . . . .	24
4.3. El Teorema de la Envolvente en Economía . . . . .	25



---

4.3.1. Visualización gráfica del Teorema de la Envolvente . . . . .	28
<b>5. Análisis de riesgo</b>	<b>31</b>
5.1. Introducción . . . . .	31
5.2. Valor en Riesgo - VaR . . . . .	32
5.2.1. Definición . . . . .	32
5.2.2. Descripción y Características . . . . .	33
5.2.3. Limitaciones . . . . .	34
5.3. Teoría de Valores Extremos . . . . .	35
5.3.1. Distribución de Valores Extremos Generalizada y el método de Block Maxima	35
5.3.2. Distribución Pareto Generalizada y el método de Excesos sobre un Umbral .	41
<b>6. Problema: el Tipo de Cambio Efectivo Real</b>	<b>45</b>
6.1. Descripción . . . . .	45
6.1.1. Relevancia . . . . .	47
6.2. Datos . . . . .	48
6.2.1. Enfoque DGVE - Niveles . . . . .	51
6.2.2. Enfoque DGVE - Rendimientos . . . . .	59
6.2.3. Enfoque DPG . . . . .	64
6.3. Interpretación . . . . .	65
<b>7. Conclusiones</b>	<b>67</b>
7.1. La situación actual . . . . .	68
<b>A. Adecuación del modelo</b>	<b>71</b>
A.1. Niveles . . . . .	71
A.2. Rendimientos . . . . .	73
<b>B. Comportamiento de los parámetros para la DGVE</b>	<b>75</b>
<b>Bibliografía</b>	<b>78</b>

---

# Capítulo 1

## Introducción

*If you keep on saying things are going to be bad, you have a good chance of being a prophet.*

Isaac Bashevis Singer

Cuando las cosas salen mal, ¿qué tan mal pueden salir? ¿Podemos establecer medidas precautorias estar listos para afrontar situaciones adversas, sin llegar a un exceso?

Existen diversas herramientas para ofrecer soluciones a este problema. La solución más sencilla sería tomar los datos históricos (claro, suponiendo que se cuenta con ellos), ajustarles un modelo adecuado, calcular algún percentil alto y tomar las acciones necesarias para cubrir ese riesgo. Sin embargo, dado que los eventos extremos para los cuales se busca tener precauciones son eventos con poca *probabilidad*, no se suelen tener muchos datos para conocer cómo se comportan estos fenómenos en situaciones muy fuera de lo esperado.

Sin embargo, eso no significa que todo esté perdido. Cuando nos interesan específicamente las *colas* de una distribución, es decir, sus eventos extremos, podemos recurrir a herramientas como la Distribución Generalizada de Valores Extremos o distribuciones Pareto Generalizadas, relacionadas con el comportamiento de los datos después de un cierto umbral. Así, si tenemos datos históricos de una empresa, no solamente sabremos qué tan probable es que una pérdida exceda un valor alto e

---

imposible de cubrir con reservas, sino que podemos también estimar la media en exceso - la pérdida promedio una vez que se rebasa el umbral crítico. Estos datos serían pertinentes, por ejemplo, para asegurarse contra este tipo de pérdidas, ya que tener reservas enormes, evidentemente, no sería la estrategia óptima.

En este trabajo se analizará el comportamiento del Tipo de Cambio Efectivo Real, un índice que se creó como una alternativa a los tipos de cambio sencillos, visto desde un enfoque de Verosimilitud Perfil tomando como parámetros de interés  $Q_\alpha$  ó  $Q_{1-\alpha}$ , un cuantil bastante bajo o uno bastante alto respectivamente, y  $\xi$ . Los primeros parámetros son de interés, pues justamente se busca conocer los niveles extremos a los que podría llegar el Tipo de Cambio Real Efectivo; el segundo lo es también pues es el parámetro que describe el tipo de Distribución de Valores Extremos que ajusta mejor a los datos.

El REER, (*Real Effective Exchange Rate* por sus siglas en inglés), es un promedio ponderado de los Tipos de cambio de todas las economías extranjeras con las que una economía local comercia, ponderado por el tamaño de tal intercambio comercial. El REER es un número sin una interpretación directa: no indica ningún tipo de cambio; sin embargo, al eliminar efectos causados por la inflación y la falta de un patrón definido para comparar todas las monedas, se vuelve una herramienta muy útil para identificar el comportamiento de una moneda. No es necesario deflactarlo para eliminar los efectos del tiempo - un alza significa una apreciación (lo cual puede significar que una moneda esté sobrevalorada).

En los primeros capítulos, se sentarán las bases en las que descansa la teoría estadística a ser utilizada en este trabajo: se hará una breve explicación de los métodos de Máxima Verosimilitud y Verosimilitud Perfil, su origen y los ámbitos en los que se suele utilizar. También se presentarán propiedades, ventajas y limitaciones de los métodos, así como ejemplos teóricos.

En el capítulo 4, se hará una revisión del Teorema de la Función Implícita, que será utilizado más adelante para implementar el algoritmo de maximización para Verosimilitud Perfil, y el Teorema

---

de la Envolvente. Se presentará la interpretación que se suele dar a este Teorema desde el punto de vista económico, con lo que será más clara la relación de la Envolvente con la Verosimilitud Perfil, y se mostrarán ejemplos gráficos que ayudarán a entender la intuición de por qué la Verosimilitud Perfil funciona. Así, se subrayará la pertinencia del método de Verosimilitud Perfil para encontrar puntos óptimos de los parámetros de interés.

En el capítulo 5, se describirá el VaR como medida de riesgo estudiable desde el lente de Valores Extremos y se describirá para evaluar su pertinencia en el ámbito. Asimismo, se hablará extensamente de la Teoría de Valores Extremos, desde la idea en la que está basada hasta un análisis de las diferencias entre los tres casos de la Distribución Generalizada de Valores Extremos. Se explicará también el modelo de Pareto Generalizado, sus características y el método de Picos Sobre un Umbral que se utiliza para ajustarlo.

En el capítulo 6, el central de este trabajo, se presentarán los datos elegidos: el Tipo de Cambio Efectivo Real (REER, por sus siglas en inglés) mensual de México, la historia de tal índice junto con sus ventajas para conocer y comparar el comportamiento del Tipo de Cambio de una moneda *doméstica* no solamente en términos de cada una de las monedas del resto del mundo, sino como un indicador consolidado incluyente de todas las monedas *extranjeras* con las que comercia. Se obtendrán primero estimadores máximo verosímiles para los parámetros originales, para después presentar resultados del método de Verosimilitud Perfil, los cuales nos permitirán tener mejores estimadores de los parámetros de interés, pues estamos estudiando solamente su comportamiento. Se analizarán tanto los niveles del REER, como sus rendimientos.

Por último, se concluirá acerca de los resultados obtenidos, y se hablará acerca de posibles nuevos enfoques para este método.



---

## Capítulo 2

# Máxima Verosimilitud

En esta sección se presentan los conceptos básicos de inferencia estadística y diversos métodos para estimación. Se profundizará en el método de máxima verosimilitud y sus características; asimismo, se analizará a los estimadores máximo verosímiles, sus ventajas y cualidades.

### 2.1. Introducción

¿Qué tanto podemos describir una población, si conocemos una pequeña parte de ella? La inferencia estadística es una herramienta útil para acercarse a este problema. Si tenemos un conjunto de observaciones y un modelo asociado a los parámetros, podemos generalizar desde una parte al todo; por supuesto, con un grado de incertidumbre. Tal incertidumbre proviene de la aleatoriedad de la muestra.

Los posibles valores para los parámetros, independientemente del método que usemos para asignarles valores, se llaman *estimadores*. Los matemáticos y estadísticos del siglo XIX y principios del XX desarrollaron numerosos métodos para acercarse a este concepto, que fue vastamente estudiado. Prevalcieron varios de estos métodos, cada uno con ventajas y desventajas, para obtener dichos estimadores. Algunos tienen un uso más extenso actualmente; uno de los principales es el método de máxima verosimilitud con diversas variantes; entre ellas: condicional, marginal y perfil.

---

## 2.2. Conceptos básicos y definición

Supongamos que una característica de los elementos en una población puede ser representada por medio de una variable aleatoria  $X$  con densidad  $f_X(\cdot; \theta)$ , conocida salvo por el parámetro  $\theta$ . Además, supongamos que podemos obtener una muestra aleatoria  $X_1, X_2, \dots, X_n$  de observaciones independientes idénticamente distribuidas con densidad  $f_X(\cdot; \theta)$ . En situaciones reales, suele ser imposible conocer a la población, así que solamente disponemos de una pequeña parte de la información. Entonces, queremos conocer lo más posible el modelo subyacente del cual proviene. El interés de estimar el parámetro  $\theta$  (o alguna función de él) no solamente radica en describir el todo a partir de una parte; también puede ser útil para muchos otros aspectos, que van desde formular predicciones hasta crear nuevos modelos.

La inferencia estadística se encarga precisamente de abordar este problema. Por medio de ella, podemos extraer de las observaciones obtenidas información que describa de una forma más general su comportamiento. Así, es una técnica cuyos resultados pueden ayudarnos a tomar decisiones; desarrollar e innovar; e incluso comprender mejor nuestro entorno.

Por ejemplo, si obtenemos datos y suponemos que provienen de una distribución normal, la inferencia estadística nos proporciona herramientas para estimar los parámetros  $\mu$  y  $\sigma$ , los cuales nos describirían por completo la campana de la distribución de la que provienen dichos datos. De esta manera, a partir de una muestra se obtiene un modelo que describe a la población, dentro de un cierto margen de error que proviene de la selección del modelo y la aleatoriedad de los datos.

Es importante asignar un modelo correcto a los datos. Si tenemos observaciones negativas, podemos excluir automáticamente algunas distribuciones, con base en su soporte, por ejemplo, la Poisson. La suposición de la distribución es un problema previo a la inferencia estadística, pero existen herramientas para evaluar qué modelos ajustan mejor a los datos, como análisis de residuales, Glivenko Cantelli, etcétera. Estas herramientas tienen ventajas y desventajas, así que la posibilidad de inferencia acerca de los datos queda a nuestro criterio. En este trabajo, se trabajará bajo la suposición de que el modelo ajustado a los datos es correcto.

---

La cantidad de datos afecta en gran medida las conclusiones a las que podríamos llegar. Evidentemente, si estimáramos la altura promedio de una población tomando solamente un par de datos, tendríamos gran incertidumbre. A medida de que tengamos más datos, la incertidumbre acerca del verdadero valor de  $\theta$  disminuirá. En el contexto de verosimilitud, la incertidumbre es una buena medida de qué tan “buenas” o “estables” son nuestras estimaciones. Por esta razón, es necesario medir la incertidumbre acerca de las estimaciones; son poco útiles si no sabemos qué tan *verosímiles* son.

### 2.2.1. Métodos comúnmente utilizados para inferencia estadística

No existe un único estimador para los parámetros. Algunos de los métodos más utilizados para estimación puntual son:

- Método de igualación de momentos - fue propuesto por Pearson en 1894 [21], citado por Cáceres Hernández [4]. Es el más antiguo y sencillo para obtener parámetros poblacionales. Es por esto que se suele utilizar para un primer acercamiento a la inferencia estadística. Consiste en igualar los momentos muestrales con los poblacionales, tantos como parámetros haya en el modelo. Como los momentos poblacionales son función de los parámetros desconocidos, esto resulta en un sistema de ecuaciones. Sin embargo, no suele utilizarse para modelos con muchos parámetros.
- Método de igualación de percentiles - fue propuesto por Owen [19]. Produce un modelo con  $p$  percentiles, escogidos arbitrariamente, que coinciden con los datos (representado con la distribución empírica). Un gran problema con este método surge con las variables aleatorias discretas, pues los percentiles pueden no ser únicos. Esta limitación se puede abordar suavizando los percentiles, ajustando rectas entre los puntos. Mientras va aumentando el tamaño de muestra, la función va ajustando mucho mejor a los puntos.
- Método de mínimos cuadrados - fue propuesto por Gauss en 1875 [24]. Se utiliza para estimar los parámetros de un modelo lineal. Parte de un conjunto de puntos observados y se pretende



---

que la función teórica pase lo más cerca posible de esos puntos. Entonces minimiza la distancia (los errores al cuadrado) respecto a una recta.

- Método de Máxima Verosimilitud - Fue introducido por Fisher en 1922 [22], aunque su verdadero origen está en disputa. En términos generales, busca los valores que hacen más probable la obtención de los datos empíricos a partir del modelo supuesto.

En el presente trabajo, se abordará el problema por medio del método de Máxima Verosimilitud y se analizará también una de sus variantes, la Verosimilitud Perfil. Además, dado que estos métodos proporcionan estimadores puntuales, se presentará una alternativa para medir la incertidumbre.

## 2.3. El método de Máxima Verosimilitud

El método de Máxima Verosimilitud fue nombrado como tal por primera vez por Fisher en 1922, aunque ya había presentado el procedimiento en 1912; además, ya había sido propuesto y utilizado anteriormente por Gauss y Laplace, entre otros. Surge entre diversas propuestas para estimación de parámetros, y de la mano de él nacen muchos conceptos como suficiencia, consistencia y verosimilitud.

Consideremos  $X$  una variable aleatoria discreta de la cual se obtienen las observaciones  $X_1, \dots, X_n$ . La *función de verosimilitud*, para variables aleatorias discretas, se define como la función de probabilidad conjunta de las observaciones, vista como función del parámetro  $\theta$ ; esto es:

$$\begin{aligned} L(\theta) &= L(\theta; x_1, \dots, x_n) = C(x_1, \dots, x_n)P(x_1, \dots, x_n; \theta) \\ &\propto P(X_1 = x_1, \dots, X_n = x_n; \theta), \end{aligned}$$

donde  $C(x_1, \dots, x_n)$  es una constante respecto al parámetro  $\theta$ .

Es importante notar que la función de probabilidad conjunta, si la muestra es independiente idénticamente distribuida, se puede descomponer como el producto de las funciones de probabilidad. Se puede pensar, de una manera muy simplificada, a la función de verosimilitud como la probabilidad de haber ob-

---

servado esa muestra.

En el caso en el que  $X$  es una variable aleatoria continua, ya que no tenemos puntos con probabilidad positiva (estrictamente, todos tienen probabilidad cero), podemos tomar en cuenta la densidad de cada una de nuestras observaciones en un cierto intervalo:

$$f(x)h \quad f(x \in [x, x + h])$$

Obtenemos una noción parecida a una probabilidad como la del caso discreto si hacemos tender cada uno de esos intervalos a cero, con lo cual, en el límite, estaríamos trabajando con funciones de densidad. Entonces, la aproximación continua a la función de verosimilitud, es:

$$L(\theta) = L(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta)$$

Se pueden consultar las bases de esta aproximación en DeGroot [8].

Para efectos de esta tesis, supondremos que no existen valores de  $x$  para los que la función de densidad tenga una singularidad, es decir, puntos en los cuales no exista la función de densidad localmente). Para más información sobre singularidades y un método para identificarlas, se puede consultar el trabajo de Zinde-Walsh y Galbraith [29]. Este problema se puede atacar con la función de verosimilitud correcta o exacta. Para más detalles, se puede consultar el artículo de Bolívar et al. [3].

El estimador de máxima verosimilitud (EMV) de  $\theta$  se define como el valor  $\hat{\theta} \in \Theta$  que maximiza la función de verosimilitud, en donde  $\Theta$  es el espacio parametral y comprende todos los posibles valores de  $\theta$ . En este sentido, para muestras que provienen de una variable aleatoria discreta, es el que hace más probable haber obtenido la muestra. Por otro lado, para muestras que provienen de densidades continuas, tenemos una intuición parecida; sin embargo, debemos recordar que la función de densidad no indica probabilidad. Así,  $\hat{\theta}$  es el candidato más plausible para ser el verdadero parámetro que origina los datos, es decir, el más verosímil bajo la realización observada.

En muchos casos es mucho más conveniente trabajar con la *función de logverosimilitud*, definida como el logaritmo natural de la función de verosimilitud:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta).$$

---

Por ejemplo, en el caso de la función de verosimilitud de una muestra aleatoria, se transforman productos de densidades en sumas. Además, debido el logaritmo es una función estrictamente creciente, el argumento que maximiza a  $l(\theta)$  es el mismo que maximiza a  $L(\theta)$ . Entonces, para encontrar  $\hat{\theta}$ , buscamos el valor de  $\theta$  que maximiza la función de logverosimilitud.

En caso de que la logverosimilitud sea diferenciable, podemos obtener su derivada, a la que llamaremos función *Score*,  $S(\theta)$ , e igualarla a cero para obtener el máximo relativo. Para asegurar que  $\hat{\theta}$  es un máximo local; verificamos que

$$l''(\theta) = S'(\theta) < 0$$

Cuando tenemos funciones que involucran múltiples parámetros,  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ , entonces la logverosimilitud es una función con  $k$  variables, y el EMV  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$  es el vector de parámetros que la maximiza. Generalmente, cuando la logverosimilitud es diferenciable,  $\hat{\theta}$  se puede obtener resolviendo las  $k$  ecuaciones simultáneas  $S_i(\theta) = 0$  para  $i = 1, \dots, k$  [13].

### 2.3.1. Intervalos de verosimilitud

En la sección anterior, se busca estimar el verdadero valor del parámetro con un solo candidato. Este tipo de inferencia se denomina *puntual*, ya que selecciona un único valor dentro del espacio parametral. Tal inferencia se complementa con la descripción de la información sobre la que se basa la selección puntual. En este contexto, por medio de la función de verosimilitud podemos obtener intervalos o regiones del espacio parametral con un nivel de verosimilitud relativa de, al menos, un determinado nivel  $c$ . De esta manera, en vez de tener un único valor puntual, se obtiene un intervalo con cierta plausibilidad  $c$ .

Primero, se considerará la estandarización de la verosimilitud, que se denominará *verosimilitud relativa*, esto es:

$$R(\theta) = \frac{L(\theta)}{\max L(\theta)} = \frac{L(\theta)}{L(\hat{\theta})}.$$

Podemos notar que  $0 \leq R(\theta) \leq 1 \quad \forall \theta \in \Theta$ . Para efectos de esta tesis, supondremos que este estimador es único, pero esta no es una condición que se cumpla en cualquier caso. Entonces, los estimadores con verosimilitud relativa cercanos a uno serán más plausibles; aquellos con valores cercanos a cero serán poco plausibles.

---

Definimos los *intervalos de verosimilitud* a nivel  $c$ , como el conjunto de valores de  $\theta$  tales que  $R(\theta) \geq c$ . Bajo ciertos supuesto de regularidad, que se cumplen en la mayor parte de los modelos simples, obtenemos una familia de intervalos anidados, que siempre contendrán al EMV, pero no necesariamente estos son simétricos alrededor de él. Los intervalos de verosimilitud nos proporcionan una idea de la dispersión y asimetría alrededor del estimador máximo verosímil.

Es importante destacar que, a un mismo nivel  $c$ , es mejor tener intervalos más pequeños, pues esto indica que hay menos incertidumbre acerca del verdadero valor de  $\theta$ . Muestras con mayor cantidad de datos nos darán, naturalmente, intervalos de menor longitud. Más adelante se ahondará sobre esta afirmación y las propiedades de la verosimilitud.

A continuación, se presenta un ejemplo del método de Máxima Verosimilitud, en el cual se obtendrá el estimador del parámetro  $\Theta$  de una distribución exponencial.

### Ejemplo 1

Si  $X$  tiene distribución exponencial tal que  $E[X] = \theta$ , entonces su función de densidad es:

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \quad \text{para } x > 0 \quad \theta \in (0, \infty)$$

Si tenemos una muestra aleatoria de  $X$ ,  $x_1, x_2, \dots, x_n$ , entonces la función de verosimilitud es:

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-x_i/\theta} = \theta^{-n} \exp\left(-\frac{1}{\theta} \sum x_i\right).$$

La función de logverosimilitud y su derivada son, respectivamente:

$$l(\theta) = -n \log \theta - \frac{1}{\theta} \sum x_i; \quad S(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum x_i$$

Resolvemos la ecuación  $S(\theta) = 0$  y obtenemos  $\hat{\theta} = \frac{1}{n} \sum x_i = \bar{x}$  como estimador máximo verosímil de  $\theta$ . Es importante verificar que  $S'(\hat{\theta}) < 0$ , asegurando así que obtenemos un máximo relativo.

Consideremos ahora los siguientes datos como una muestra aleatoria proveniente de una distribución exponencial con un parámetro  $\theta$  desconocido:

---

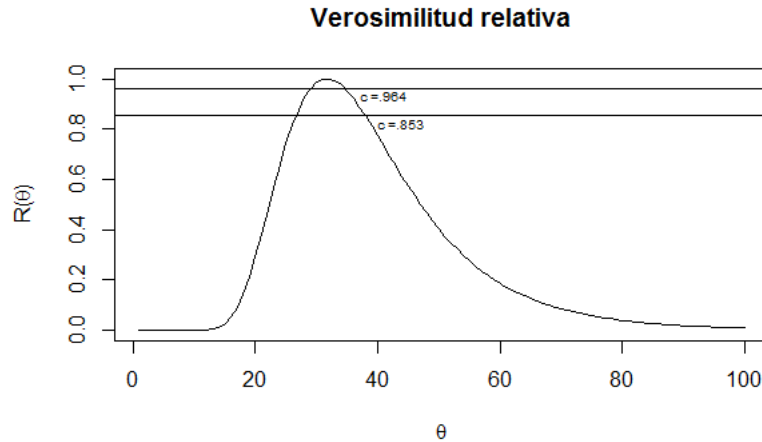

$$\{4, 65, 23, 12, 54, 23, 31, 28, 49, 27\}$$

Como ya hemos obtenido la expresión cerrada para el EMV del parámetro  $\theta$ , podemos obtenerlo directamente. Además, verificamos por medio de  $S'(\hat{\theta})$  que hayamos obtenido un máximo.

$$\hat{\theta} = \frac{1}{n} \sum x_i \Rightarrow \hat{\theta} = 31,6$$

$$S'(\hat{\theta}) = -0,010014420$$

En la siguiente gráfica, observamos el comportamiento de estimadores de  $\theta$ , así como los intervalos de verosimilitud con niveles  $c_1=0.147$  y  $c_2=0.036$ , utilizados comúnmente porque corresponden aproximadamente a intervalos de confianza de 95 y 99 %, respectivamente.



Al ser  $\theta$  un parámetro unidimensional, podemos observar que definir un intervalo de verosimilitud es separar, por medio de una línea paralela al eje de  $\theta$ , a los valores que tienen una plausibilidad mayor o menor al valor  $c$  de interés. Podemos notar también que, como se mencionó antes, son intervalos anidados que contienen al estimador máximo verosímil, pero no forzosamente están centrados en él.

### 2.3.2. Propiedades de los EMV

Hogg y Craig [11] habla de propiedades importantes de los estimadores máximo verosímiles (EMV):

- Consistencia - converge en probabilidad al parámetro que estima.

- 
- Normalidad asintótica - a medida que el tamaño de muestra tiende a infinito, la distribución del EMV tiende a ser Normal con media igual al parámetro a estimar, y matriz de covarianzas inversa a la Matriz de Información de Fisher, definida, bajo condiciones de regularidad débiles (existencia de la información de Fisher en todos los puntos e intercambiabilidad de las operaciones de derivación e integración), con elemento típico:

$$(I(\theta))_{i,j} = E \left[ \left( \frac{\partial}{\partial \theta_i} \log f(X; \theta) \right) \left( \frac{\partial}{\partial \theta_j} \log f(X; \theta) \right) \middle| \theta \right].$$

- Eficiencia - su varianza alcanza la cota inferior de Cramér-Rao, ligada a la Información de Fisher, cuando el tamaño de muestra tiende a infinito.
- Invarianza funcional- es decir, el estimador máximo verosímil de una función biyectiva del parámetro es sencillamente la función evaluada en el EMV.
- Sesgo - en general, no se puede asegurar que un estimador máximo verosímil sea insesgado. Un estimador insesgado es aquel que cumple  $E[\hat{\theta}] = \theta$ , es decir, su valor esperado es igual al valor real del parámetro. En muchos casos, se encuentran estimadores asintóticamente insesgados; el límite del sesgo asociado a ellos es cero cuando el tamaño de muestra tiende a infinito.

## Limitaciones

Aunque el estimador máximo verosímil es asintóticamente insesgado, un gran sesgo puede persistir con muestras pequeñas. Además, debido a que las muestras son aleatorias, siempre existe la posibilidad de que tengamos una muestra muy poco representativa.

Cuando no se puede obtener una expresión cerrada para el estimador máximo verosímil, hay que recurrir a métodos numéricos susceptibles a problemas de resolución. Por otro lado, si estamos trabajando con un modelo con muchos parámetros, aumentamos la complejidad de los algoritmos y, por lo tanto, el tiempo máquina.

Además, se pueden obtener óptimos que no son globales: la segunda derivada negativa nos asegura que obtendremos un máximo, pero no que será un máximo global. Cualquier máximo local satisfaría la restricción de maximización que proponemos, así que este problema de convergencia podría causar un grave error de estimación. Sin embargo, para atacar este problema, existen técnicas como el algoritmo *MultiStart* para aumentar la posibilidad de obtener óptimos globales. Este algoritmo, en específico, toma

---

varios puntos de inicio distribuidos en toda la superficie de optimización. Uno de los principales problemas es que es posible que muchos de estos puntos lleguen al mismo óptimo, y por lo tanto es menos eficiente que métodos más sencillos cuando se aplica a funciones con un solo punto óptimo. Se puede obtener más información de este tipo de algoritmos en Csendes [7].

---

## Capítulo 3

# Verosimilitud Perfil

En los libros de texto de inferencia estadística, generalmente encontramos ejemplos de modelos con pocos parámetros. Nunca se trata en papel con normales multivariadas o modelos logísticos con más de unos cuantos parámetros. Sin embargo, en la práctica, los modelos suelen tener muchos parámetros, lo cual complica la interpretación de los modelos resultantes. Una gráfica en 2D de una distribución normal es fácilmente comprendida, pero la representación visual de un modelo con 20 parámetros es imposible; es mucho más fácil analizar el comportamiento de los parámetros uno por uno.

Los *parámetros de ruido* aparecen irremediablemente en nuestros esfuerzos por crear modelos *mejores* para nuestros datos. Por supuesto, la definición de un modelo *mejor* es subjetiva, pero en la generalidad esto suele significar más parámetros. Sin embargo, muchas veces existe un subconjunto de los parámetros del modelo que no son realmente de nuestro interés, sino que sirven para completar el modelo.

En esta sección presentaremos, como una opción para atacar este problema, el método de Verosimilitud Perfil. Asimismo, analizaremos sus ventajas y desventajas con respecto al método de verosimilitud y a otros como verosimilitud marginal y condicional.

### 3.1. Definición

A pesar de que el concepto de verosimilitud perfil fue usado en modelos econométricos desde la década de los 50, y que Fisher utilizó una versión muy simple del método en una tabla que publicó en 1956, Sprott y



---

Kalbfleisch [23] presentaron por primera vez el método llamado *verosimilitud maximizada* como un método general para eliminar parámetros de ruido en *Application of Likelihood Methods to Models Involving Large Numbers of Parameters* (1970).

La función de verosimilitud perfil es una herramienta estadística que nos permite estimar parámetros de interés en presencia de parámetros de ruido. Es un método más general que otras verosimilitudes, pues no depende de una estructura especial, tal como lo requieren otras formas de inferencia, como las verosimilitudes marginal y condicional.

No se debe confundir a la función verosimilitud perfil con las de verosimilitud condicional y marginal. Las segundas requieren que la función de verosimilitud se pueda factorizar de una manera específica; la primera, no. Además, los métodos son sumamente diferentes, aunque todos resultan en funciones de densidad con menos parámetros.

Para definir estas dos verosomilitudes, definamos una transformación de  $x$  a  $(v, w)$  tal que la distribución marginal de  $v$  o la condicional de  $(v|w)$  se pueden expresar como función de solamente los parámetros de interés. Sea  $\delta = (\theta, \lambda)$  el vector de parámetros, del cual nos interesa  $\theta$ .

La verosimilitud marginal busca eliminar los parámetros de ruido por medio, justamente, de la integración (o suma) marginal respecto a ellos - es un método muy parecido a la manera Bayesiana de deshacerse de parámetros que no son de interés - sin embargo, dado que la verosimilitud no es una función de probabilidad, integrar sobre un parámetro para sacarlo puede no tener el efecto correcto. Si la verosimilitud se puede expresar de la siguiente manera:

$$\begin{aligned} L(\theta, \lambda) &= p_{\theta, \lambda}(v, w) \\ &= p_{\theta}(v) p_{\theta, \lambda}(w|v) \\ &= L_1(\theta) L_2(\theta, \lambda) \end{aligned}$$

entonces, la verosimilitud marginal de  $\theta$  es:

$$L_1(\theta) = p_{\theta}(v)$$

---

La verosimilitud condicional, busca no depender de ellos estableciendo su valor como una estadística suficiente. La perfil, por otro lado, busca escribir a los parámetros de interés en términos de los parámetros de ruido.

$$\begin{aligned} L(\theta, \lambda) &= p_\theta(v|w)p_{\theta, \lambda}(w) \\ &= L_1(\theta)L_2(\theta, \lambda) \end{aligned}$$

entonces, la verosimilitud marginal de  $\theta$  es:

$$L_1(\theta) = p_\theta(v|w)$$

Para más detalles acerca de estas tres variantes de verosimilitud, se puede consultar Pawitan [20].

Si tenemos un vector de parámetros  $\delta = (\theta, \lambda)$ , donde los parámetros incluidos en  $\theta$  son de interés, entonces podemos definir a la función de verosimilitud perfil, o maximizada, como:

$$L_P(\theta; y) = \max_{\lambda} L(\theta, \lambda; y).$$

Es decir, la verosimilitud perfil de  $\theta$ ,  $L_P(\theta; y)$ , se obtiene maximizando la función de verosimilitud  $L(\theta, \lambda; y)$  sobre  $\lambda$  pero fijando  $\theta$ . Así, reemplazamos al parámetro de ruido,  $\lambda$ , por su estimador máximo verosímil restringido a un cierto valor de  $\theta$ , en la función de verosimilitud global  $L(\theta, \lambda; y)$ .

La función de verosimilitud perfil se puede tratar entonces como una verosimilitud. Se puede desde normalizar hasta obtener intervalos de verosimilitud de ella.

### 3.1.1. Información:

Reescribamos, por comodidad de notación, a nuestro vector de parámetros  $\theta$  como  $(\theta_1, \theta_2)$  en donde  $\theta_1$  es el subconjunto de parámetros que son de interés, y  $\theta_2$  contiene a los parámetros de ruido.

Entonces podemos escribir a la Matriz de Información de Fisher como:

---


$$I(\hat{\theta}) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

con inversa:

$$I(\hat{\theta})^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix}$$

Las verosimilitudes perfiles para cada parámetro pueden ser aproximadas por medio de:

$$\log L(\theta_i) \approx -\frac{1}{2}(I^{ii})^{-1}(\theta_i - \hat{\theta}_i)^2$$

y el error estándar de  $\hat{\theta}_i$  es la raíz de los elementos diagonales de  $I^{-1}(\hat{\theta})$ :

$$se(\hat{\theta}_i) = \sqrt{I^{ii}}$$

Se puede consultar Pawitan [20] para los desarrollos anteriores, así como para un análisis más profundo acerca de los métodos comunes utilizados lidiar con parámetros de ruido.

## 3.2. Ventajas con respecto a Verosimilitud

El hecho de no estimar a los parámetros de ruido proporciona dos ventajas principales: menor costo computacional y un mejor conocimiento del comportamiento de los parámetros de interés.

Trabajar con menos parámetros es menos costoso computacionalmente. Si existe forma cerrada para el estimador máximo verosímil de cada parámetro, una sustitución directa resulta en trabajar con menos ecuaciones simultáneas y, por ende, algoritmos menos complejos. Si, por otro lado, los estimadores tienen que ser computados numéricamente, aún estaremos buscando maximizar una nueva función con respecto a menos variables, lo cual es, evidentemente, menos costoso.

Además, incluso cuando es de interés estudiar un gran número de parámetros, generalmente es más fácil entender el modelo si observamos un parámetro a la vez. Es importante destacar que es posible obtener el EMV perfil de cada parámetro por separado, y el conjunto de tales estimadores para todos los parámetros asegura el máximo de la función de verosimilitud original. Recordemos que los estimadores máximo verosímiles maximizan, en esa combinación, la función de densidad conjunta. Por lo tanto, al obtener la

---

maximización de un subconjunto cualquiera con respecto a su complemento, la función obtenida contendrá a los estimadores máximo verosímiles originales.

Otra ventaja importante es que se puede utilizar el método de Verosimilitud Perfil cuando se tiene una gran cantidad de datos faltantes o censurados dentro de las observaciones. Al trabajar con menos parámetros, se hacen menos suposiciones acerca de restricciones, y se obtienen estimaciones más estables de parámetros acerca de los que se tiene suficiente información. Se puede consultar más acerca de esta aplicación en Zhang [28].

También podría ser de gran utilidad en el caso de un modelo con una gran cantidad de parámetros y pocos datos. Imaginemos un caso extremo en el cual tuviéramos  $n$  parámetros y  $k \ll n$  observaciones, de manera que repercutiera en una varianza paramétrica considerablemente grande <sup>1</sup>. Entonces, al hacer inferencia solamente para los parámetros de interés, tendremos un modelo más equilibrado y con más información (de cierta manera) para la certeza de tales parámetros. Este es un tema en el cual se puede ahondar en futuros trabajos.

Utilizar el Teorema de la Envolvente para estudiar el comportamiento *perfil* de los parámetros de interés nos permite tener un conocimiento más profundo de cómo se comportan estos, y sus puntos óptimos, respecto a los de ruido. Esto cobra gran importancia en ámbitos como el económico, en los cuales es de interés estudiar las variaciones de ciertos parámetros en el largo plazo, describiendo su comportamiento respecto a otros. En realidad, esta es la aplicación que parecería más directa y significativa del método de Verosimilitud Perfil.

### 3.3. Ejemplos

Consideremos el modelo normal con  $\mu \in \mathbb{R}$  y  $\sigma > 0$ . Sus funciones de densidad, verosimilitud y logverosimilitud están dadas por:

$$f(x_1, \dots, x_n; \mu, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} \prod_{i=1}^n \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right)$$

---

<sup>1</sup>Ya que, en general, los parámetros suelen tener varianza paramétrica asintótica a su cota inferior de varianza mientras el tamaño de la muestra, es decir, la *certeza* del valor del parámetro disminuye si el tamaño de muestra es menor.

---


$$L(\mu, \sigma) = \frac{c}{\sigma^n} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right)$$

$$l(\mu, \sigma) = \log c - n \log \sigma - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}$$

Tomemos primero el caso en el que  $\mu$  es nuestro parámetro de interés, y  $\sigma$  el de ruido.

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum (x_i - \mu)^2}{\sigma^3}$$

Obtengamos ahora el máximo, al resolver:

$$\max_{\sigma} l(\mu, \sigma; y) = 0$$

$$-\frac{n}{\hat{\sigma}} + \frac{\sum (x_i - \mu)^2}{\hat{\sigma}^3} = 0 \quad \Leftrightarrow \quad \hat{\sigma} = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

Ya que hemos obtenido  $\hat{\sigma}(\mu)$ , es decir, el argumento que maximiza  $\sigma$  sujeto a  $\mu$ , podemos insertar esta nueva expresión en nuestra función de logverosimilitud original:

$$\begin{aligned} l(\mu, \hat{\sigma}(\mu)) &= \log c - n \log \hat{\sigma}(\mu) - \frac{\sum (x_i - \mu)^2}{2\hat{\sigma}(\mu)^2} \\ &= \log c - n \log \sqrt{\frac{\sum (x_i - \mu)^2}{n}} - \frac{\sum (x_i - \mu)^2}{2\sqrt{\frac{\sum (x_i - \mu)^2}{n}}} = \log c - \frac{n}{2} \left(1 + \log \frac{\sum (x_i - \mu)^2}{n}\right) \end{aligned}$$

Busquemos ahora el máximo:

$$\begin{aligned} \frac{\partial l(\mu, \hat{\sigma}(\mu))}{\partial \mu} &= \frac{n^2 \sum (x_i - \mu)}{\sum (x_i - \mu)^2} \\ \frac{\partial l(\mu, \hat{\sigma}(\mu))}{\partial \mu} = 0 &\Leftrightarrow \hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} \end{aligned}$$

El mismo procedimiento para el caso inverso, en el que  $\sigma$  es nuestro parámetro de interés, y  $\mu$  el de ruido:

$$\frac{\partial l}{\partial \mu} = \frac{\sum (x_i - \mu)}{\sigma^2}$$

Obtiene su máximo en

$$\max_{\mu} l(\sigma, \mu; y) = 0$$

---


$$\hat{\mu}(\sigma) = \frac{\sum x_i}{n}$$

Entonces podemos insertar  $\hat{\mu}(\sigma)$  en nuestra función de logverosimilitud original:

$$l(\sigma, \hat{\mu}(\sigma)) = \log c - n \log \sigma - \frac{-\sum (x_i - \frac{\sum x_i}{n})^2}{2\sigma^2}$$

$$\frac{\partial l}{\partial \sigma} = \frac{-n}{\sigma} + \frac{\sum (x_i - \frac{\sum x_i}{n})^2}{\sigma^3} \quad \frac{\partial l}{\partial \sigma} = 0 \Leftrightarrow \hat{\sigma}^2 = \frac{\sum (x_i - \frac{\sum x_i}{n})^2}{n}$$

Es de esperarse que los estimadores máximo verosímiles perfil (por separado) coincidan con los máximo verosímiles. Además, son parámetros independientes en el sentido de su matriz de varianzas covarianzas; con ceros fuera de la diagonal en la Matriz de Información de Fisher; es decir, ninguno guarda información acerca del otro.

### 3.4. Limitaciones

La verosimilitud perfil puede arrojar resultados poco intuitivos. Laos [14] presenta ejemplos para mostrar que estos problemas generalmente se deben a singularidades en la función de densidad. La función de verosimilitud perfil hereda tales singularidades. Sin embargo, si acotamos a la función de verosimilitud, tal que sea proporcional a una probabilidad, tales singularidades desaparecen.

También ilustra ciertas desventajas que podrían provenir de que la función de verosimilitud perfil aparentara un exceso de precisión que no es realista para la muestra observada, o si se tuvieran problemas de estimación con muchos parámetros de ruido: podría arrojar estimadores que no fueran insesgados o consistentes. Generalmente, en problemas que no presentan irregularidades, el sesgo es combatible con muestras grandes. Sin embargo, si se trabaja con una cantidad demasiado grande de parámetros de ruido (con respecto a la cantidad de datos en la muestra). Pawitan [20] presenta un ejemplo bastante claro de este caso, en el cual asume  $N$  observaciones y  $N + 1$  parámetros. En este tipo de situaciones, si la función de verosimilitud cumple la estructura requerida, podría ser más conveniente usar otros métodos de inferencia.

---

---

## Capítulo 4

# El Problema de Maximización y las Envolventes

En esta tesis, el enfoque elegido para optimizar las estimaciones de los parámetros de interés, contrastados a los parámetros de ruido, se basa en estudiar el comportamiento de los segundos en términos de los primeros. De esta manera, se reduce la dimensionalidad del problema. El fundamento matemático de este enfoque se encuentra en el Teorema de la Envolvente.

### 4.1. El Teorema de la Función Implícita

Sea  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  una función con derivadas parciales continuas en una región  $U$  y sea  $(x_0, y_0)$  un punto interior de  $U$  tal que  $F(x_0, y_0) = 0$  y  $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$ . Entonces existen regiones abiertas  $A$  que contiene a  $x_0$  y  $B$  que contiene a  $y_0$ , tal que existe una función  $f : A \rightarrow B$  para la cual:

$$F(x, f(x)) = 0$$

para todo  $x \in A$ , donde  $y_0 = f(x_0)$  y  $f(x)$  es continuamente diferenciable:

$$\frac{\partial f}{\partial x} = -\frac{F_x(x, f(x))}{F_y(x, f(x))}$$

para todo  $x \in A$  y  $y \in B$ .

El Teorema de la Función Implícita se generaliza fácilmente para funciones  $F : \mathbb{R}^n \rightarrow \mathbb{R}$ .



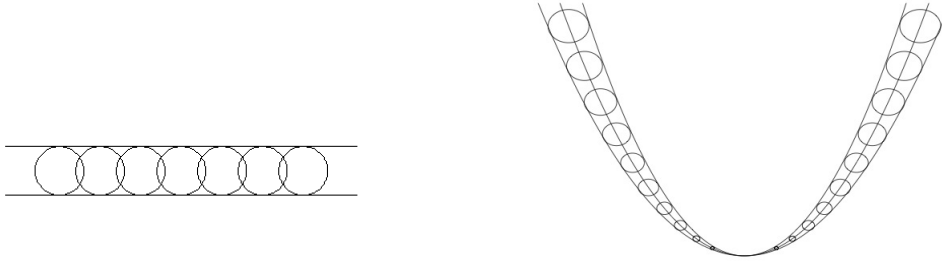
---

## 4.2. Envoltentes de una familia de funciones

Definimos a una *familia de curvas* como  $f(x, y, c)$ , una función continuamente diferenciable en todos los parámetros, en donde cada miembro de la familia se corresponde con un valor, perteneciente a un intervalo dado, de  $c$ . Por ejemplo, podemos definir a la familia de círculos concéntricos alrededor del origen, o de radio dado y cuyos centros pertenecen a la recta  $x = a$ , rectas perpendiculares a un eje o, incluso, todas las normales a una curva dada. Asimismo, se pueden considerar tanto familias de curvas multiparamétricas, como familias de superficies.

Con el enfoque de Courant y John [6], si consideramos a la familia de rectas tangentes a una curva plana  $C$ , podemos entonces definir a  $C$  como la *envolvente* de la familia de rectas. En general, la envolvente de una familia de curvas será tangente en cada uno de sus puntos a alguna curva perteneciente a dicha familia. Es decir, la envolvente es el lugar geométrico de las intersecciones de curvas vecinas.

Algunos ejemplos sencillos son las envolventes de una familia de círculos con centro sobre una recta y radio constante, o de una familia de círculos cuyos centros se encuentran sobre una parábola y con radio creciente de manera lineal, presentados en las siguientes figuras.



Tomemos entonces una curva  $f(x, y, c) = 0$  y una curva vecina  $f(x, y, c + h) = 0$ . En su punto de intersección, se cumplirán las siguientes ecuaciones:

$$\frac{f(x, y, c + h) - f(x, y, c)}{h} = 0$$
$$f(x, y, c + h) = 0, \quad f(x, y, c) = 0$$

Se busca entonces expresar a  $x$  y  $y$  como funciones del parámetro  $c$ , para obtener la ecuación de la curva

---

envolvente en términos de  $c$ , por medio de dos funciones continuamente diferenciables  $x = x(c)$  y  $y = y(c)$ . Entonces, la envolvente evaluada en  $c$  tocará a la curva de la familia  $f(x, y, c) = 0$ . Al derivar con respecto a  $c$ , se llega a:

$$f_x \frac{dx}{dc} + f_y \frac{dy}{dc} + f_c = 0$$

Además, se puede sustituir  $x = x(c)$  y  $y = y(c)$  en lugar de  $x$  y  $y$  en la ecuación  $f(x, y, c) = 0$ , sujeto a que  $c$  pertenezca al intervalo original, lo que resulta en una identidad en  $c$ . Esto quiere decir que, para todos los puntos de la envolvente:

$$f_x \frac{dx}{dc} + f_y \frac{dy}{dc} = 0.$$

Por lo tanto, la envolvente debe satisfacer la condición  $f_c(x, y, c) = 0$  y se tienen las dos condiciones suficientes para obtener la envolvente.

### 4.3. El Teorema de la Envolvente en Economía

Supongamos que se quiere maximizar una función  $f(x; p)$ , con  $p = (a, b) \in \mathbb{R}^n \times \mathbb{R}^k$  bajo ciertas restricciones, pero es de interés conocer el comportamiento del subconjunto  $a$  de los  $p$  parámetros dentro de la maximización de la función. Una primera intuición del Teorema de la Envolvente es que, para un modelo de la forma anterior, nos permite conocer los efectos marginales de un subconjunto de cambios de  $b$  en  $a$ . Recordemos que, para que este teorema sea aplicable, es necesario que la función de  $b$  esté bien definida en una vecindad alrededor de  $a$ , es decir, que exista una vecindad en la cual sea aplicable el Teorema de la Función Implícita. Localmente, la solución del problema sería una función diferenciable  $b(a)$ .

Al escribir a la función con una reparametrización  $f(x; p) = f(x; a, b(a))$  podemos intuir que los parámetros  $a$  tienen dos efectos en la función. El primero, más bien explícito, es cómo afecta a la función directamente. El segundo tiene que ver con su dependencia del resto de los parámetros y cómo son afectados estos por los parámetros de interés  $a$ .

En economía, el Teorema de la Envolvente se suele utilizar para resolver problemas de optimalidad en el largo plazo. Se utiliza para observar los efectos de ciertos parámetros en los equilibrios, pues gracias al TE es comprobable que, en el largo plazo, solamente los efectos directos prevalecen.

---

Se quiere resolver el problema de maximización:

$$\max_{\theta} f(\theta, \delta) \quad \text{s.a.} \quad g(\theta, \delta) = 0$$

En el cual  $f : \mathbb{R}^m \rightarrow \mathbb{R}, \theta = (\theta_1, \dots, \theta_m)$  contiene variables llamadas *de control*, las cuales son análogas a las llamadas variables *de ruido* en el contexto de verosimilitud perfil, y  $\delta$  es un parámetro que afecta  $f$  o  $g$ , pero cuya maximización no se realiza.

Si la función es suficientemente diferenciable (es decir, en la generalidad, que el determinante de su matriz Jacobiana sea  $\neq 0$ ), podemos obtener la función Langrangiana:

$$\mathcal{L}(\theta, \delta, \lambda) = f(\theta, \delta) - \lambda[g(\theta, \delta)]$$

con condiciones de primer orden:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{\partial f(\theta^*, \delta)}{\partial \theta_i} - \lambda \frac{\partial g(\theta^*, \delta)}{\partial \theta_i} = 0 \quad \forall i \in 1, \dots, m$$

Así, bajo ciertas condiciones, se puede obtener la solución a este problema por medio del conjunto de soluciones  $(\hat{\theta})_i^*$  que ahora se encuentran expresadas en términos del parámetro de interés.

Se expresa, entonces, la función original como una nueva función del parámetro  $\delta$ . Esta función es una maximización de  $f$ :

$$F(\delta) = f(\theta^*(\delta), \delta)$$

Cuya derivada total es definida por:

$$\frac{dF(\delta)}{d\delta} = \sum_{i=1}^m \frac{\partial f}{\partial \theta_i} \frac{d\theta_i^*}{d\delta} + \frac{\partial f}{\partial \delta}$$

Y que, sustituyendo las condiciones de primer orden, se convierte en:

$$\frac{dF(\delta)}{d\delta} = \sum_{i=1}^m \lambda^* \frac{\partial g}{\partial \theta_i} \frac{d\theta_i^*}{d\delta} + \frac{\partial f}{\partial \delta}$$

Por otro lado, si se obtiene la derivada total con respecto a  $\delta$  de la restricción original, se obtiene:

---


$$\frac{\partial g}{\partial \delta} = \sum_{i=1}^m \frac{\partial g}{\partial \theta_i} \frac{d\theta_i^*}{d\delta}$$

Que se puede sustituir en la ecuación anterior, con lo cual obtenemos el *Teorema de la Envolvente*:

$$\frac{\partial F(\delta)}{\partial \delta} = \frac{\partial f}{\partial \delta} - \lambda^* \frac{\partial g}{\partial \delta} = \frac{\partial \mathcal{L}}{\partial \delta}$$

Se puede estudiar el Teorema de la Envolvente, así como su aplicación en diversos problemas económicos en Varian [27] y en las notas de Economics 101 de Albouy [1].

### Interpretación del TE

El Teorema de la Envolvente, aplicado a maximización de funciones de utilidad (o, en su caso, minimización de funciones de costo) puede interpretarse de la siguiente manera. Sea un problema de maximización:

$$M(a) = \max_x f(x, a)$$

En el cual  $a$  podría ser el precio o el ingreso,  $x$  un factor de demanda y  $M(a)$  el valor maximizado de los beneficios como función del precio. Sea  $x(a)$  el valor que maximiza la función. Esto quiere decir que el valor optimizado de la función es igual a la función evaluada en la  $x$  que la optimiza,  $M(a) = f(x(a), a)$ . Podría ser de interés evaluar el comportamiento de  $M(a)$  ante cambios de  $a$ .

Se aplica el Teorema de la Envolvente sin restricciones, entonces:

$$\frac{\partial F(\delta)}{\partial \delta} = \frac{\partial f}{\partial \delta} = \frac{\partial \mathcal{L}}{\partial \delta}$$

Obtenemos:

$$\frac{dM(a)}{da} = \left. \frac{\partial f(x, a)}{\partial a} \right|_{x=x(a)}$$

Lo cual puede interpretarse como que la derivada de  $M$  con respecto a  $a$  es igual a la derivada parcial de  $f$  con respecto a  $a$ , evaluada en el punto óptimo  $x(a)$ . Esto quiere decir que podemos conocer los óptimos de la función original con respecto a variaciones de solamente parámetros de interés.

### Interpretación del parámetro $\lambda$

Tomemos el caso en el que la maximización de interés fuera:

---


$$\max_{\theta} f(\theta, \delta) \quad s.a. \quad g(\theta, \delta) = c$$

Es decir, ahora existe una variable más dentro de la restricción, por lo tanto, el máximo depende ahora de una variable extra. Reescribamos la restricción como

$$g(\theta, \delta) = c \Leftrightarrow g(\theta, \delta) - c = 0 \Leftrightarrow g'(\theta, \delta, c) = 0$$

Evaluando el Teorema:

$$\frac{\partial F}{\partial c} = \frac{\partial f}{\partial c} - \lambda^* \frac{\partial g'}{\partial c} = \frac{\partial \mathcal{L}}{\partial c}$$

Pero

$$\frac{\partial f}{\partial c} = 0 \quad \frac{\partial g'}{\partial c} = -1$$

Entonces,

$$\frac{\partial F(\delta)}{\partial \delta} = 0 - \lambda^*(-1) = \frac{\partial \mathcal{L}}{\partial \delta}$$

$$\frac{\partial F(\delta)}{\partial \delta} = \frac{\partial \mathcal{L}}{\partial \delta} = \lambda^*$$

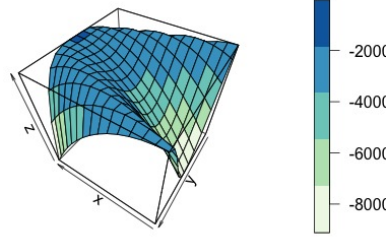
Se puede interpretar al parámetro  $\lambda$  como el efecto que tienen cambios en la restricción sobre la función de interés.

#### 4.3.1. Visualización gráfica del Teorema de la Envolvente

Tomemos la siguiente función, de la cual nos interesa el comportamiento de  $y$ :

$$z(x, y) = -(x - y^2)^2 - x$$

**Funcion de interés**



Aplicando el Teorema de la Envolvente, obtenemos:

$$\frac{\partial z}{\partial(x)} = -2(x - y^2) + 1$$

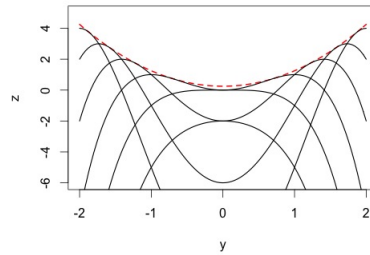
Que al igualar a cero, nos devuelve el máximo:

$$x^* = 0,5 + y^2$$

Ya que conocemos el comportamiento de  $a$  con respecto al valor óptimo de  $x$ , lo reemplazamos en la función original a optimizar:

$$z(x^2(y), y) = -((0,5 + y^2) - y^2)^2 - (0,5 + y^2) = 0,25 + y^2$$

**Envolvente con respecto a x**



Con lo cual podemos ver que la envolvente describe el comportamiento de los máximos de la función cuando  $x$  está en su óptimo. De esta manera, se puede ver a la Verosimilitud Perfil como el problema de encontrar la envolvente de la familia de funciones de distribución, tomando como parámetros variables los parámetros de interés, y haciendo la maximización sobre los de ruido.

---

---

## Capítulo 5

# Análisis de riesgo

En esta sección se hablará de la teoría del análisis de riesgo; su finalidad, y su utilidad en varios aspectos. Después se abordarán dos herramientas de análisis de riesgos: el Valor en Riesgo (VaR) y la teoría de Valores Extremos. Al final, se comentará la relevancia de estas técnicas y se discutirá la aplicabilidad de ellas a algunos problemas.

### 5.1. Introducción

Vivimos rodeados de aleatoriedad. Desde los fenómenos naturales, hasta los experimentos científicos y los indicadores financieros; podemos encontrar el factor del azar en casi cualquier ambiente.

Podemos pensar, por ejemplo, en una variable aleatoria relacionada con las pérdidas que podría sufrir una empresa. El primer paso lógico sería hacer inferencia acerca del valor esperado de la variable. Para ir un poco más lejos, podríamos hablar de su varianza. Sin embargo, aún nos haría falta una consideración más: los extremos. Las pérdidas negativas no causan mucho revuelo, pues son ganancias. Sin embargo, un evento extremo en la cola derecha (pérdida alta con baja probabilidad), podría ser catastrófico.

A pesar de ser valores “extremos”, no podemos sencillamente ignorarlos. Por ejemplo, si tomamos el cuantil .95 de una distribución (es un cuantil muy común para el VaR), estableceremos un umbral que esperamos que se supere una de cada veinte veces. En el caso de que nuestros datos fueran anuales, nos hablaría de un evento extremo una vez cada veinte años. Para una empresa, es un lapso bastante largo en



---

el cual pueden cambiar drásticamente las condiciones; para un edificio que debe mantenerse en pie en una zona sísmica, no. Entonces, es importante tener herramientas que nos ayuden a estar preparados para tales valores extremos.

## 5.2. Valor en Riesgo - VaR

El Valor en Riesgo (VaR) surgió justo después de la caída de la bolsa de 1987, mejor conocida como *Lunes Negro*. Los analistas cuantitativos (también llamados *quants*) aplicaban modelos estadísticos estándares, los cuales hacían que una crisis pareciera muy improbable. Sin embargo, sus métodos fueron puestos en duda en el momento en el que la caída comenzó a expandirse por todo el mundo.

Fue necesario, entonces, dejar de ver solamente valor esperado o varianza; voltear hacia una nueva medida de riesgo que nos hablara de eventos catastróficos. Surgió el VaR, que en palabras de Jorion [12], es un valor umbral tal que la probabilidad de que la pérdida lo exceda, sea mayor o igual a un nivel fijo; *i.e.* el  $VaR_q$  es un percentil alto de la distribución de la variable de pérdida.

En el artículo de Ugray y Marti [25] se menciona que se estableció en varios grupos comerciales en diversas instituciones financieras, notablemente *Bankers Trust* y *J. P. Morgan*, la cual publicó la metodología y dio libre acceso a valores estimador de los parámetros necesarios en 1994. Esta fue la primera vez que el VaR fue expuesto a un público no conformado por quants. Sin embargo, a pesar de que ya estaba siendo usado, en 1990 aún no se habían estandarizado el nombre ni la definición.

McNeil et al. [16] considera al VaR como la medida preferida de riesgo de mercado; se utiliza en administración de riesgos y control financiero. Se puede consultar más información acerca del VaR y su uso como estándar de medición de riesgo financiero en Jorion [12].

### 5.2.1. Definición

El Valor en Riesgo (*Value at Risk*, VaR) es un percentil alto de la distribución de la variable de pérdida  $X$ . En otras palabras, para  $q \in (0.5, 1)$ :

$$VaR_q(x) = F^{-1}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$$

---

### 5.2.2. Descripción y Características

Es una especie de cota superior para una pérdida que se excede solamente en pocas ocasiones. Al ser un percentil alto, se suelen generar problemas graves de estimación y pronóstico. Por otro lado, el VaR no nos indica el tamaño potencial de la realización de la variable en caso de excederse el VaR.

#### Análisis de coherencia

Podemos examinar las propiedades del VaR como medida de riesgo.

**Monotonía.** Tomemos dos variables de pérdida,  $X$  y  $Y$ . La propiedad de monotonía nos dice que si  $X > Y$  ( $X$  domina estocásticamente a  $Y$ ), entonces su VaR será más grande.

Proposición:  $x \geq y \Rightarrow VaR_q(x) \geq VaR_q(y)$

*Demostración.*  $F_x(z) = P(x \leq z) \leq P(y \leq z) = F_y(z)$

$\therefore VaR_q(x) = \inf\{x \in \mathbb{R} : F_x(x) \geq q\} \geq \inf\{y \in \mathbb{R} : F_y(y) \geq q\} = VaR_q(y)$   $\square$

**Homogeneidad positiva.** Tomemos la variable de pérdida  $X$ , por ejemplo asociada a un activo. La homogeneidad positiva nos dice que si compramos el doble de ese activo, el VaR será exactamente el doble. Es decir, el riesgo crece linealmente con cada unidad de la variable de pérdida  $X$  extra a la que nos enfrentemos.

Proposición  $\forall c \geq 0, VaR_q(cx) = cVaR_q(x)$

*Demostración.* Sean  $y = cx$ . Entonces,  $P(Y \leq y) = F_x(\frac{y}{c})$

$\therefore VaR_q(y) = \inf\{y \in \mathbb{R} : F_x(\frac{y}{c}) \geq q\} = cVaR_q(x)$   $\square$

**Invarianza ante traslaciones.** Tomemos de nuevo una variable de pérdida  $X$  y asociémosla a un activo. La invarianza ante traslaciones podría ser interpretada como que el VaR no cambiará si nuestra riqueza inicial es menor o mayor, pues depende íntegramente de la variable de pérdida.

Proposición  $\forall c \geq 0, VaR_q(x + c) = VaR_q(x) + c$

*Demostración.* Sea  $y = x + c$ . Entonces,  $P(Y \leq y) = F_x(y - c)$

Si  $y = VaR_q(x) + c$ , se tiene que  $F_x(y - c) = F_x(VaR_q(x) + c - c) = F_x(VaR_q(x))$

$\therefore VaR_q(x + c) = VaR_q(y) = \inf\{y \in \mathbb{R} : F_x(y - c) \geq q\} = VaR_q(x) + c$   $\square$

---

**Subaditividad.** Tomemos dos variables de pérdida  $X$  y  $Y$ . La subaditividad nos aseguraría que, si la empresa  $A$  posee el riesgo de pérdida  $X$ , y la empresa  $B$ , el riesgo  $Y$ , entonces, una fusión de las empresas no generaría riesgo adicional.

Para probar subaditividad, tendríamos que demostrar que  $VaR_q(x+y) \leq VaR_q(x) + VaR_q(y)$ . Pero para demostrar que el VaR no la cumple, tomaremos un contraejemplo.

Sean  $x \sim F(x) = 1 - x^{-\frac{1}{2}} x \geq 1$  ;  $y \sim F(y) = 1 - y^{-\frac{1}{2}} y \geq 1$  independientes.

Definamos dos riesgos:  $Z = X + Y$ ;  $W = 2X$ . Entonces,

$$\begin{aligned} F_Z(z) &= \int_1^\infty f_Y(y) F_X(z-y) dy \\ &= \int_1^z \frac{y^{-\frac{3}{2}}}{2} \left[ 1 - (z-y)^{-\frac{1}{2}} \right] dy = 1 - \frac{2\sqrt{z-1}}{z} \quad z \geq 2 \end{aligned}$$

Además,

$$F_W(w) = P(W \leq w) = P(X \leq \frac{w}{2}) = F(x) = 1 - \frac{w^{-\frac{1}{2}}}{2}, w \geq 2$$

Entonces,

$$1 - \frac{2\sqrt{z-1}}{z} < 1 - \frac{2\sqrt{2z}}{z} = P(2x \leq z), z \geq 2$$

$\therefore VaR_q(x+y) > VaR_q(2x) = VaR_q(x) + VaR_q(y)$

Luego, el VaR es monótono, homogéneo e invariante ante translaciones, pero no es subaditivo.

### 5.2.3. Limitaciones

En general, el problema de subaditividad puede ser evitado. Para todas las distribuciones elípticas, es decir, aquellas generalizaciones de la normal multivariada, el VaR sí es subaditivo. En la situación ideal en la que todos los portafolios de riesgo pueden ser representados como combinaciones lineales del mismo conjunto de distribuciones elípticas, el VaR sí será una medida coherente de riesgo. Se puede encontrar esta demostración en [15]

El VaR no nos indica el tamaño potencial de la pérdida en caso de excesos. Además, generalmente se usan valores de  $q \geq 0.9$ , lo cual puede generar problemas graves de estimación y pronóstico.

Encontramos uno de los principales problemas del VaR cuando nuestros datos provienen de una distri-

---

bución de colas pesadas. Dado que se espera que en una muestra haya pocas observaciones pertenecientes a la *cola derecha* (es decir, percentiles muy altos), contamos con pocos datos para ajustar el comportamiento de la cola. Entonces, debido a la escasez de datos para la inferencia respecto a la cola, el VaR podría arrojar un valor demasiado bajo.

### 5.3. Teoría de Valores Extremos

A grandes rasgos, en la literatura existen dos grandes enfoques respecto a la Teoría de Valores Extremos. El primero, llamado *Block Maxima*, se enfoca en las observaciones extremas (generalmente los máximos), por medio de una partición de los datos en bloques que no forzosamente son del mismo tamaño. El segundo, llamado *Excesos Sobre un Umbral*, define una cota a partir de la cual es de interés conocer los excesos de la variable.

*Block Maxima* es un método tradicional que se ha utilizado con datos que tienen cierta estacionalidad, como variables climáticas. Al ser un método que requiere separar los datos en bloques, es natural pensar en, por ejemplo, los máximos anuales de precipitación. Por otro lado, el método de *excesos sobre un umbral* utiliza los datos más eficientemente, así que ha ganado terreno y popularidad en aplicaciones recientes. Se puede encontrar más información acerca de ambos métodos en McNeil et al. [15] y Gilli y Këllezli [10]

#### 5.3.1. Distribución de Valores Extremos Generalizada y el método de Block Maxima

Supongamos que tenemos  $n$  realizaciones independientes de la variable aleatoria  $X$ , con funciones de densidad y distribución  $f(x)$  y  $F(x)$ , respectivamente. Denotemos a estas  $n$  realizaciones como  $x_1, \dots, x_n$ . Para conocer el comportamiento extremo de esta variable, es de interés el comportamiento de el último estadístico de orden:

$$M_n = \max\{x_1, \dots, x_n\}$$

cuya distribución es:

$$F_{M_n} = F_x^n(x).$$

Analicemos la convergencia. Sabemos que:

$$P(M_n \leq x) = F^n(x), \quad x \in \mathbb{R}, n \in \mathbb{N}$$

---

y podemos definir al extremo derecho de  $F$  como:

$$x_F = \sup\{x \in \mathbb{R} \mid F(x) < 1\}$$

Entonces, en general, tendremos que:

$$P(M_n \leq x) = F^n(x) \rightarrow 0 \quad \text{si } n \rightarrow \infty \quad \text{para } x < x_F$$

Por otro lado, si  $x_F < \infty$ , para  $x \geq x_F$  tenemos  $F^n(x) = 1$ . Entonces, observemos que

$$M_n \xrightarrow{pr} x_F, \quad \text{si } n \rightarrow \infty \quad \text{y } x_F < \infty.$$

Supongamos ahora que  $M_n$  converge en distribución, bajo una normalización adecuada: por medio de una sucesión de constantes  $d_n$  y  $c_n > 0$ , tal que,  $\forall n$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} P((M_n - d_n)/c_n \leq x) &= \lim_{n \rightarrow \infty} P(M_n \leq c_n x + d_n) \\ &= \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x) \end{aligned}$$

para alguna  $H(x)$  no degenerada. Si esta existe, entonces se dice que  $F$  pertenece al dominio de atracción de  $H$ .

Así, si una distribución es no colapsada, entonces el Teorema de Fisher Tippet nos asegura que convergerá a una de las tres distribuciones de valores extremos.

**Teorema de Fisher Tippet 1.** Sea  $(X_n)$  una sucesión de variables aleatorias iid. Si existen constantes  $c_n > 0, d_n \in \mathbb{R}$ , y una distribución  $H$  no degenerada tal que

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H,$$

entonces  $H$  pertenece a una de las tres distribuciones de valores extremos:

$$\begin{aligned} \text{Fréchet: } \phi_\alpha(x) &= \begin{cases} 0, & x \leq 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases} & \alpha > 0 \\ \text{Weibull: } \psi_\alpha(x) &= \begin{cases} e^{-(-x)^\alpha}, & x \leq 0 \\ 1, & x > 0 \end{cases} & \alpha > 0 \end{aligned}$$

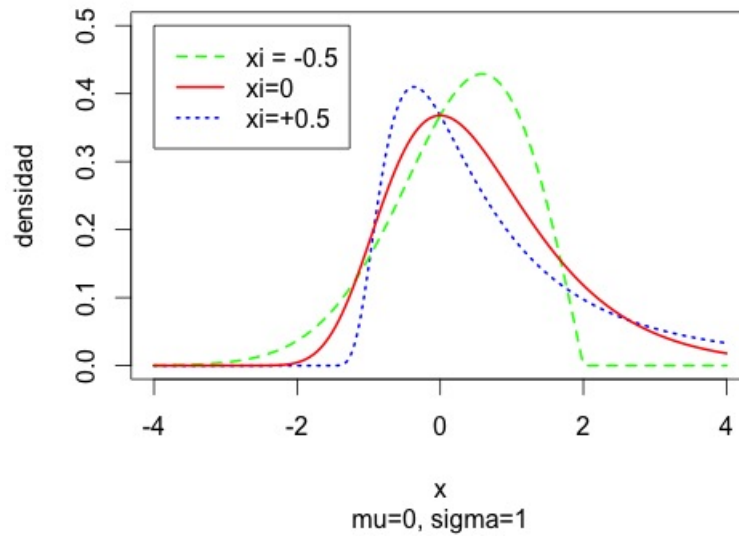
---


$$\text{Gumbel: } \Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}$$

Jenkinson (1955) y von Mises (1954) sugirieron la *Distribución de Valores Extremos Generalizada* (DVEG), con la siguiente representación:

$$H_{\xi}(x) = \begin{cases} e^{-(1+\xi x)^{-1/\xi}} & \xi \neq 0 \\ e^{-e^{-x}} & \xi = 0 \end{cases} \quad 1 + \xi x > 0$$

### Distribución Generalizada de Valores Extremos



Podemos analizar gráficamente las diferentes distribuciones. Para el caso en el que  $\xi < 0$ , Weibull, podemos observar que la cola derecha (ligera) cae drásticamente. Este efecto es mucho más notorio comparando con el caso  $\xi > 0$ , Fréchet, en la cual se puede distinguir que la caída de la cola es mucho más lenta (pesada). Así, el caso Weibull será más adecuado para modelos acotados, y el Fréchet para fenómenos que puedan crecer con menos mesura, digamos, más *catastróficos*.

Así podemos obtener la Fréchet con  $\xi = \alpha^{-1}$ , la Weibull con  $\xi = -\alpha^{-1}$ , y la Gumbel como el caso límite cuando  $\xi = 0$ . Podemos incluir también parámetros de escala y localización,  $\mu$  y  $\sigma$  respectivamente, de la misma manera:

$$H_{\xi, \mu, \sigma}(x) = H_{\xi}\left(\frac{x - \mu}{\sigma}\right)$$

---

Al separar en  $m$  bloques nuestros datos, por medio del método Block Maxima, podemos obtener el valor máximo de cada uno de nuestros bloques. De esta manera, proponemos a la colección de estos valores,  $\{X_{n1}, X_{n2}, \dots, X_{nm}\}$  como realizaciones de la variable aleatoria  $X_n$ , es decir, el último estadístico de orden. Así, podemos hacer inferencia acerca de los parámetros de la distribución a la que pertenece.

En general, no conocemos la distribución de la cual provienen los datos, así que la estimación del parámetro  $\xi$  se vuelve verdaderamente útil para identificarla. Una manera muy simple de tener un primer acercamiento es por medio de la *gráfica empírica de media en exceso*. La media en exceso, definida como:

$$M(u) := E[X - u | X > u]$$

La cual puede interpretarse como la pérdida extra esperada, una vez que se rebasó un umbral  $u$ . La media en exceso empírica queda definida entonces como:

$$\hat{M}(u) = \frac{\sum (X_i - u) I_{[X_i > u]}}{\sum I_{[X_i > u]}}$$

Si la pendiente de la gráfica de media en exceso empírica es positiva,  $F$  pertenece al dominio de atracción Fréchet (que tiene colas polinomiales); si es negativa, pertenece al Weibull (colas ligeras); y si es cero, pertenece al dominio de atracción Gumbel. Se puede consultar más información acerca de la media en exceso en Embrechts y Mikosch [9]

Es importante notar que los parámetros de localización y escala dependerán de las constantes  $d_n$  y  $c_n$  que utilicemos; sin embargo, el parámetro  $\xi$  de forma quedará definido independientemente de ellas. Esto quiere decir que siempre es posible encontrar constantes  $d_n$  y  $c_n$  que, en el límite, produzcan una forma “estándar” de  $H_\xi$ .

Por último, es importante recalcar que, a pesar de que el análisis se centra en la cola derecha de la distribución (el máximo de las observaciones), podemos fácilmente analizar la distribución del mínimo de ellas, es decir, la cola izquierda, con una sencilla reparametrización:

$$\min\{X_1, \dots, X_n\} = -\max\{-X_1, \dots, -X_n\}$$

---

## Función de verosimilitud para DVEG

Bajo el supuesto de independencia, la función de logverosimilitud que buscamos maximizar es:

$$L(\xi, \mu, \sigma; x) = \sum \log h(x_i)$$

En donde

$$h(\xi, \mu, \sigma; x) = \frac{1}{\sigma} \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi - 1} \exp\left(-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}\right)$$

es la función de densidad si  $\xi \neq 0$  y  $1 + \xi \frac{x - \mu}{\sigma} > 0$ . Si, por el contrario,  $\xi = 0$ , entonces la función de densidad es:

$$h(\xi, \mu, \sigma; x) = \frac{1}{\sigma} \left(\frac{x - \mu}{\sigma}\right) \exp\left(-\exp\left(\frac{x - \mu}{\sigma}\right)\right)$$

Con la reparametrización para unificar a la Weibull, la Fréchet y la Gumbel en la Distribución de Valores Extremos Generalizada, la función de logverosimilitud está dada por:

$$\begin{aligned} h(a, b, c; x) &\propto \sum_{i=1}^n \log f(x_i; a, b, c) \\ &= -n \log b - \left(1 + \frac{1}{a}\right) \sum_{i=1}^n \log \left(1 + c \frac{x_i - a}{b}\right) \sum_{i=1}^n \left(1 + c \frac{x_i - a}{b}\right)^{\frac{1}{a}} \end{aligned}$$

Con los parámetros = 
$$\begin{cases} a = \mu, b = \sigma & \text{con el modelo Gumbel} \\ a = \mu - \sigma, b = \frac{\sigma}{\beta}, c = \frac{1}{\beta} & \text{con el modelo Frechet} \\ a = \mu + \sigma, b = \frac{\sigma}{\beta}, c = -\frac{1}{\beta} & \text{con el modelo Weibull} \end{cases}$$

No es posible obtener analíticamente los estimadores máximo verosímiles para la DGVE; sin embargo, existen muchos algoritmos para la obtención de tales valores, muchos de los cuales contienen arreglos para asegurar obtención de máximos globales, mientras se cumplan ciertas condiciones de regularidad (las mismas requeridas para la Información de Fisher). La implementación del método de máxima verosimilitud, entonces, se hará de forma numérica.

## Función de verosimilitud perfil para DVEG

Nuestro interés se centra en el  $VaR$ , el cual está definido como un percentil de la distribución, y cuyo valor depende de todos los parámetros de la distribución. La verosimilitud perfil no nos otorga estimaciones de los parámetros de ruido, así que es necesario otro acercamiento: existe una forma cerrada para los



---

percentiles de las distribuciones de Valores Extremos. Si es de interés el percentil  $\alpha$ , entonces:

$$Q_{1-\alpha} = \begin{cases} \mu - \sigma(-\log(1-\alpha))^{1/\xi} & (Weibull) \\ \mu - \sigma \log(\log(1-\alpha)) & (Gumbel) \\ \mu + \sigma(-\log(1-\alpha))^{-1/\xi} & (Frechet) \end{cases}$$

Por lo tanto, podemos expresar a  $\mu$  en términos de los parámetros  $\sigma$  y  $\xi$  y el percentil de interés, que reparametrizaremos como  $Q_{1-\alpha}$ , eliminando  $\mu$ :

$$\mu(Q_{1-\alpha}, \sigma) = \begin{cases} Q_{1-\alpha} - \sigma(-\log(1-\alpha))^{1/\xi} & (Weibull) \\ Q_{1-\alpha} - \sigma \log(\log(1-\alpha)) & (Gumbel) \\ Q_{1-\alpha} + \sigma(-\log(1-\alpha))^{-1/\xi} & (Frechet) \end{cases}$$

Por lo tanto, se puede hacer inferencia sobre nuestro parámetro de interés  $Q_{1-\alpha}$ , por medio de la nueva función de logverosimilitud:

$$L(\xi, Q_{1-\alpha}, \sigma; x) = \sum \log(h(x_i)),$$

en donde

$$h(\xi, Q_{1-\alpha}, \sigma; x) = \frac{1}{\sigma} \left( 1 + \xi \frac{x - \mu(Q_{1-\alpha}, \sigma)}{\sigma} \right)^{-1/\xi-1} \exp \left( - \left( 1 + \xi \frac{x - \mu(Q_{1-\alpha}, \sigma)}{\sigma} \right)^{-1/\xi} \right)$$

para los modelos Weibull y Fréchet, y

$$h(\xi, Q_{1-\alpha}, \sigma; x) = \frac{1}{\sigma} \left( \frac{x - \mu(Q_{1-\alpha}, \sigma)}{\sigma} \right) \exp \left( - \exp \left( \frac{x - \mu(Q_{1-\alpha}, \sigma)}{\sigma} \right) \right)$$

para el modelo Gumbel.

De igual manera que con el modelo simple, no es posible obtener analíticamente las expresiones para los estimadores, así que la implementación del método se hará de forma numérica.

---

### 5.3.2. Distribución Pareto Generalizada y el método de Excesos sobre un Umbral

Una de las grandes debilidades del método de *Block Maxima* es la poca eficiencia al usar los datos; al utilizar solamente la máxima observación de cada bloque, perdemos la mayor parte de nuestra información. Los métodos basados en excesos sobre un nivel determinado conservan todos los datos que cumplan con la condición de sobrepasar un umbral, resultando en un menor desperdicio. Bajo este enfoque, buscamos conocer el comportamiento de la distribución  $F_u$ , llamada *distribución condicional de excesos* y definida como:

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(y + u) - F(u)}{1 - F(u)}, \quad 0 \leq y \leq x_F - u$$

En donde  $u$  es un *umbral* dado,  $y = x - u$  son los excesos y  $x_F$  es el extremo derecho de  $F$ , definido en la sección anterior.

El principal modelo para excesos sobre un umbral es la *Distribución Pareto Generalizada* (DPG), cuya función de densidad está dada por:

$$G_{\xi, \beta}(y) = \begin{cases} 1 - (1 + \xi \frac{y}{\beta})^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-y/\beta) & \xi = 0 \end{cases}$$

con  $\beta > 0$ ; y  $0 \leq y \leq (x_F - u)$  cuando  $\xi \geq 0$  o  $0 \leq y \leq -\beta/\xi$  cuando  $\xi < 0$ . Los parámetros  $\xi$  y  $\beta$  son, respectivamente, de forma y escala.

La DPG lleva el calificativo “generalizada” porque, al igual que la DVEG, abarca distintos casos según el valor de su parámetro. Cuando  $\xi > 0$ , se refiere a una distribución Pareto con parámetros  $\alpha = 1/\xi$  y  $\kappa = \beta/\xi$ . Si  $\xi = 0$ , se tiene una distribución exponencial; además, la DPG es continua en este punto, pues es el caso límite cuando  $\xi \rightarrow 0$ . Por último, si  $\xi < 0$ , tenemos una distribución Pareto tipo II. El parámetro  $\xi$  describe la cola de la distribución: mientras más grande sea  $\xi$ , más pesada será la cola.

El método de excesos sobre un umbral se basa en el supuesto de que, para una distribución  $F$  con extremo derecho  $x_F$ , existe un umbral  $u$  para el que  $F_u(x) = G_{\xi, \beta}(x)$  para  $0 \leq x < x_F - u$  y alguna  $\xi \in \mathbb{R}$  y  $\beta > 0$ . Esto quiere decir que el umbral  $u$  debe ser alto para satisfacer las condiciones; sin embargo, mientras más alto sea  $u$ , quedarán menos observaciones por encima de él.

---

Cabe destacar que no existen algoritmos satisfactorios para la selección de  $u$  [10]. Una DPG con un umbral más alto mantendrá el mismo parámetro  $\xi$ , mientras que el parámetro  $\beta$  crecerá de manera lineal respecto al umbral. Esto quiere decir que la selección del umbral se puede basar en la observación del crecimiento de  $\nu$  frente a alteraciones en  $u$ ; un umbral adecuado será aquel en el que la gráfica de la media en exceso deje de tener un comportamiento lineal. Los detalles de esta propiedad de los parámetros, así como de la selección del umbral, se pueden consultar en McNeil et al. [15].

La función de logverosimilitud de la DPG se define como:

$$L(\xi, \beta|y) = \begin{cases} -n \log \beta - (\frac{1}{\xi} + 1) \sum \log(1 + \frac{\xi}{\beta} y_i) & \text{si } \xi \neq 0 \\ -n \log \beta - \frac{1}{\beta} \sum y_i & \text{si } \xi = 0. \end{cases}$$

Al igual que con la DGVE, la DPG no tiene expresiones cerradas para los estimadores máximo verosímiles; sin embargo, también existen algoritmos que nos permiten obtener tanto los valores de los estimadores, como su varianza.

## Adecuación del modelo DPG

Existen diversas pruebas para verificar que los datos cumplen con la hipótesis de distribución Pareto, entre las que se encuentra la *gráfica de media en exceso* mencionada en la sección anterior.

La distribución Pareto (con todas sus generalizaciones) es la única caracterizada por cumplir la Ley de van der Wijk [26]. Esta ley fue concebida originalmente en el contexto de medición de riqueza, y enuncia que los ingresos promedio de la población sobre un cierto nivel  $u$  es proporcional a  $u$  mismo, es decir

$$\frac{\int_u^\infty t f(t) dt}{\int_u^\infty f(t) dt} = cu \quad c > 0$$

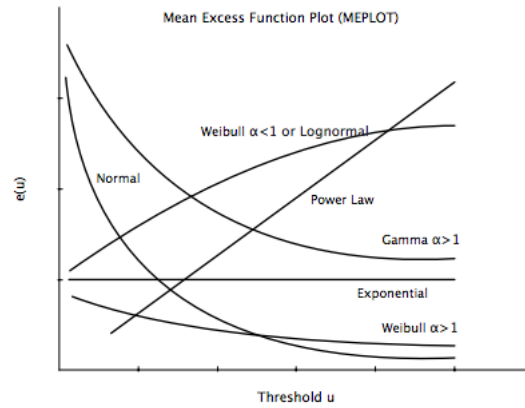
Es decir, la distribución Pareto está caracterizada por un comportamiento lineal respecto a  $u$  de la función de medias en exceso. En específico, el comportamiento de esta función para la distribución Pareto Generalizada es

$$e(u) = \frac{\beta + \sigma u}{1 - \sigma} \quad \beta + \sigma u > 0$$

En la siguiente gráfica, tomada de Cirillo [5] se presentan las formas típicas de las gráficas de medias

---

en exceso para distribuciones comunes. Recordemos que la distribución Pareto es parte de las *power law*.



Tal representación gráfica es coherente con la idea intuitiva de cada uno de los modelos: aquellos pertenecientes a la ley de potencias presentan un crecimiento lineal de la media en exceso al aumentar el umbral; los pertenecientes al dominio normal y Weibull presentan un comportamiento decreciente relacionado a sus colas ligeras; por último, el caso exponencial es sumamente fácil de identificar, pues su media en exceso es constante, independientemente del umbral.



---

## Capítulo 6

# Problema: el Tipo de Cambio Efectivo Real

En este capítulo se hará un análisis, tomando como instrumento principal la inferencia estadística con modelos paramétricos, del Tipo de Cambio Efectivo Real (REER por sus siglas en inglés, Real Effective Exchange Rate). En particular, se tratará como un problema de Valores Extremos.

### 6.1. Descripción

En 1944 se establecieron los acuerdos de Bretton Woods. Fue un esfuerzo después de la Segunda Guerra Mundial para establecer reglas y procedimientos estandarizados para regular el sistema monetario internacional (gracias a ellos, también, se fundó el Fondo Monetario Internacional). La base consistía en que las políticas monetarias de todos los países debían apegarse a un tipo de cambio fijo respecto a los dólares estadounidenses. Esto se logró por medio del Patrón Dólar, con el cual se estableció un tipo de cambio fijo entre el dólar (y la libra esterlina) y el oro. Esto permitió la cooperación y redujo la devaluación y las oportunidades de arbitraje en tipos de cambio. Esta época tuvo gran estabilidad económica, además de pocas fluctuaciones en los tipos de cambio mundiales. Sin embargo, en 1971 se eliminó el Patrón; comenzó a haber más inestabilidad en los tipos de cambio relativos, así que considerarlos por separado ya no reflejaba los cambios que en realidad estaban teniendo. Se puede leer más acerca de los acuerdos y sus consecuencias en Mikesell [17].

---

El Tipo de Cambio Efectivo Real (REER) es el promedio ponderado de una moneda *doméstica* relativo a un índice (o canasta) de otras monedas, ajustado también por inflación. Las ponderaciones de cada moneda *extranjera* son proporcionales a las balanzas de intercambio relativas; es decir, aquellos países con los que el país doméstico tiene mayor volumen de comercio tendrán mayor ponderación. Se diferencia del Tipo de Cambio Efectivo Nominal en que este último se mide con las partes nominales, es decir, sin tomar en cuenta la diferencia entre el poder adquisitivo de los consumidores de todos los países. El tipo de cambio real es solamente el tipo de cambio nominal multiplicado por los índices de precios de ambos países involucrados. Es importante destacar que ambos índices no están definidos para cada moneda extranjera por separado, sino que son un solo número que expresa cómo se comporta la moneda doméstica frente al conjunto de las demás.

Así, el REER es útil para determinar el valor de una moneda *doméstica* individual, respecto a las demás monedas en el índice. Da una referencia acerca del desempeño y movimiento de la moneda doméstica con relación al resto del mundo, como si fuera una sola entidad. Las monedas extranjeras que se consideran son generalmente las más comunes y con mayor oferta y demanda: como dólares (EUA) y yenes (Japón). De esta manera, podemos saber cuánto pagaría un consumidor individual por un bien importado, pero a nivel justamente de consumo. Este “precio” incluye tarifas y aranceles asociados a la importación del bien. En una situación “ideal”, una moneda nunca estaría mal valorada respecto al resto del mundo, y todos los consumidores tendrían estarían indiferentes entre comprar los productos en su economía doméstica o importarlos, pues costarían lo mismo en términos reales. A esto se le llama *Hipótesis de Paridad de Poder Adquisitivo*.

Es importante notar que, para poder hacer una verdadera comparación, es necesario transformar al índice de nominal a real. Supongamos que la tasa de cambio nominal de pesos mexicanos a dólares estadounidenses es  $15MXN = 1USD$ , por facilidad. Si decimos que el nivel de precios es mayor en EUA que en México, eso quiere decir que un dólar comprará en México más que lo que quince pesos comprarán en EUA. Esto, por un lado es una disparidad en poder de consumo; por otro, presenta una oportunidad de arbitraje: un individuo podría obtener ganancias cambiando sus dólares por pesos, comprando bienes en México y vendiéndolos de nuevo en EUA por dólares. Por último, se puede interpretar como una sobrevaloración del dólar. Si se estudia un bien producido en ambas economías de manera idéntica, y el dólar está sobrevalorado, entonces la tasa de cambio será favorable para México, pues es relativamente más barato importar tal bien desde México que desde EUA.

---

Este simple proceso se puede generalizar a todos los bienes producidos en una economía. De esta manera, el REER nos permite saber si una moneda está sobre o infravalorada. Si la moneda doméstica se vuelve más cara en términos de las demás (ya sea en término nominal o real), entonces las exportaciones se volverán menos valiosas en términos de la doméstica, y las importaciones serán más baratas en términos relativos. Esta relación también funciona, en sentido inverso, si la moneda doméstica se abarata. De esta manera, entonces, se puede medir la competitividad de exportaciones.

Desde 1971 ha habido muchas maneras de medir el REER, dado que las canastas de consumo no son iguales en cada país, además de que los datos acerca del volumen de intercambios comerciales no siempre han estado completos para todos los países. Otro obstáculo subjetivo al determinar el REER es la ponderación que recibe cada país según su importancia respecto al país doméstico. Es importante destacar, también, que los valores nominales del REER cambian a través del tiempo debido al cambio en el año base elegido. Sin embargo, dado que es solamente un cambio de escala, mantienen la misma relación entre ellos, sin importar el año base.

### 6.1.1. Relevancia

Un aumento en el nivel del índice significa una apreciación de la moneda, y viceversa. En teoría, si la moneda doméstica se vuelve más cara en términos de las extranjeras, entonces las importaciones se volverán relativamente más baratas, y las exportaciones tendrán menos valor.

De esta manera, este índice es, sin duda, sumamente útil para investigación, análisis y creación de políticas. Sin embargo, se puede notar que tiene un enfoque global, pues toma en cuenta todas las monedas extranjeras con las que un país comercia. Esto quiere decir que, para un inversionista con interés en solamente una o dos monedas, habría herramientas mejores; sin embargo, para los creadores de políticas monetarias, es información valiosa para decidir acerca de las intervenciones.

La relevancia de este índice estudiado desde el enfoque de Valores Extremos puede ser ejemplificada con un caso sucedido en 1992, comúnmente llamado *Black Wednesday*. La Libra esterlina pertenecía al Mecanismo Europeo de Cambio (el precursor del actual Euro), lo cual conllevaba políticas económicas y monetarias que aseguraran que el tipo de cambio entre las monedas pertenecientes no fluctuara más de 6 %. Sin embargo, la creencia generalizada de que la Libra esterlina estaba sobrevaluada y la excesiva especula-



---

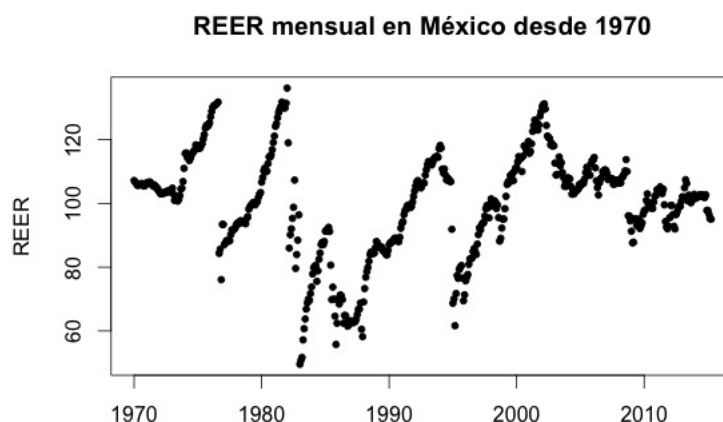
ción acerca de ella, especialmente por parte de George Soros, llevó a que la Libra se derrumbara y cayera por debajo del nivel mínimo para pertenecer al MCE. Esto resultó en una pérdida de millones de Libras para el Reino Unido, y una ganancia de alrededor de un millón para Soros.

También, en los últimos años, Rusia sufrió una gran inestabilidad con el Rublo. Desde 2011 tuvo devaluaciones anuales fuertes, causadas por el cambio en su balanza de pagos: generalmente esta cuenta tenía un balance positivo, fuertemente soportado por las reservas de petróleo. Sin embargo, ya en el segundo trimestre del 2013, su balanza de pagos estaba rondando el cero, con tendencia negativa. Además, los conflictos políticos recientes (entre ellos las controversias de Crimea y el este de Ucrania) resultó en una disminución de la entrada de capital extranjero. En respuesta a esta situación, el Banco Central de Rusia tomó una serie de medidas el septiembre pasado, en el cual aplicó medidas de intervención cambiaria, y anunció que entraría en el mercado en caso de amenazas a la estabilidad financiera. Ofreció préstamos a los bancos a una tasa de 8% anual, causando que estos se concentraran en comprar moneda extranjera. Podemos recordar en este momento la crisis que sufrió en 2008-2009, en la cual la presión sobre el rublo terminó pocos días después de que el Banco dejó de emitir préstamos. Se puede obtener más contexto acerca de esto en el artículo de Sergey Aleksashenko [2].

Dado que no existe un método estandarizado para determinar si una moneda está bien valuada, la búsqueda de comportamiento extremo en el REER se presenta como una buena alternativa. Las oportunidades de especulación se pueden presentar tanto si una moneda está sobrevalorada como en el caso contrario, por lo tanto, son de interés eventos extremos que se encuentren en ambas colas, *i.e.* tanto percentiles muy altos, como muy bajos.

## 6.2. Datos

Los datos usados en esta tesis fueron obtenidos de la base de datos del Fondo Monetario Internacional y proporcionados por el Instituto Tecnológico Autónomo de México. Comprenden el Tipo de Cambio Efectivo Real mensual de México (REER), desde enero de 1970 hasta abril de 2015, un total de 544 observaciones.



Se pueden notar claramente dos picos con explicación conocida: el primero en 1982 y el segundo, alrededor de 1995.

Ambos son marcas importantes en la historia económica de México. Recordemos que las políticas de intercambio comercial implementadas a mediados del siglo pasado, diseñadas para subsidiar la importación para la industria doméstica, causó que los bienes y servicios aumentaran sus precios al triple entre 1954 y 1976. El problema fue que estos mismos precios aumentaron solamente al doble en Estados Unidos, nuestro principal socio comercial. En este caso, el valor del peso debía haber sido depreciado, pero el gobierno mantuvo una tasa de cambio fija contra el dólar. La principal consecuencia de esto fue una caída monumental en las exportaciones, llegando a menos de 4% del PIB en 1971. La recuperación, aunque pequeña, llegó a mediados de los 70 con un aumento en el precio del petróleo y el descubrimiento de nuevos yacimientos.

Esta política de protectorado siguió hasta el 18 de Febrero de 1982, cuando el gobierno decidió devaluar de 27 a 47 Pesos por Dólar, además de la nacionalización de la banca. Esto causó que el PIB de México se redujera 42%, debido a su valuación en USD. Por otro lado, la deuda, también valuada en USD, se incrementó efectivamente. Este es el primer pico que se puede ver en los datos.

Se esperaba que esta política conllevara una disminución en las importaciones, sin embargo esto no sucedió. Para contrarrestar esto, el gobierno confiscó todos los valores en dólares, a un valor mucho menor del real, e impuso un fuerte control sobre el cambio de divisas. Sin embargo, estas medidas no fueron sos-

---

tenibles por el gobierno, conllevando un desplome del peso: en 1981, la tasa de cambio era 26:1; en 1983 se encontraba en 144:1, y en 1985 alcanzó 372:1.

La segunda gran devaluación del Peso, no causada por factores aleatorios, sucedió en 1994.

Durante el gobierno de Salinas, el Banco de México mantuvo una política <sup>1</sup> de tasa de cambio que sobrevaluaba el peso. Los productos extranjeros eran baratos en términos de moneda local, dando la impresión de que gracias al PRI se había aumentado el poder de consumo.

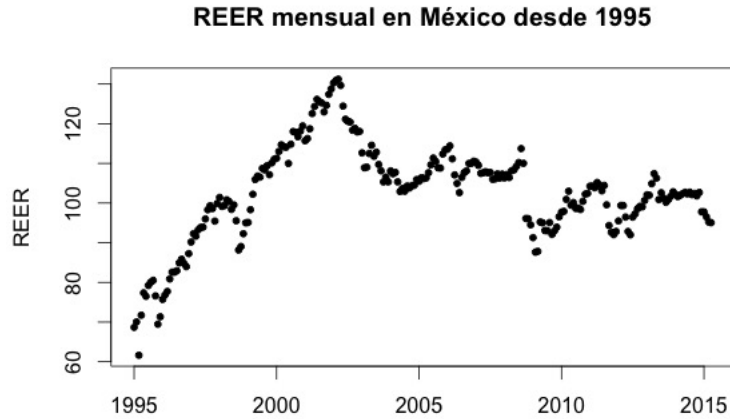
Las tasas de interés en México eran considerablemente más altas que en EUA y el gobierno mantenía sus políticas enfocadas en mantener una tasa de cambio estable, aunque gravemente sobrevaluada. La combinación de estos dos factores incentivó la inversión en bancos y fondos mexicanos, así como la seguridad de empresas y bancos de tener sus deudas en Dólares.

A finales de 1994, la presión sobre el tipo de cambio debido a la sobrevaluación era ya evidente. Las empresas y bancos buscaban retirar sus fondos en Pesos y convertirlos a Dólares lo más rápido posible, antes de que fuera inminente una transición a un tipo de cambio flotante. Esto sucedió en diciembre de 1994.

La entrada de capital extranjero en forma de inversión en mercado de valores y bonos de gobierno llegó a un balance negativo, que continuó durante 1995. La estrategia de Salinas de sobrevaluar artificialmente el peso tuvo consecuencias caras: los productos mexicanos, relativamente caros en términos de Dólares, se volvieron poco competitivos en mercados internacionales. Los inversionistas de todo el mundo dejaron de creer que el gobierno pudiera mantener tal tasa artificial, así que poco después de la sucesión presidencial, en la cual Zedillo tomó el poder, comenzaron a vender los valores en Pesos que poseían. Cuando la tasa de cambio Dólar-Peso fue convertida en flotante, las deudas valuadas en pesos duplicaron su valor en unos cuantos días.

---

<sup>1</sup>Esta política, conocida como *crawling peg*, consiste en mantener el tipo de cambio de la moneda local dentro de un rango de valores (o “bandas”) permitiendo una devaluación lenta a través del tiempo. El Banco trató de controlar esto con una devaluación de 15 %, pero este ajuste controlado sirvió de poco.



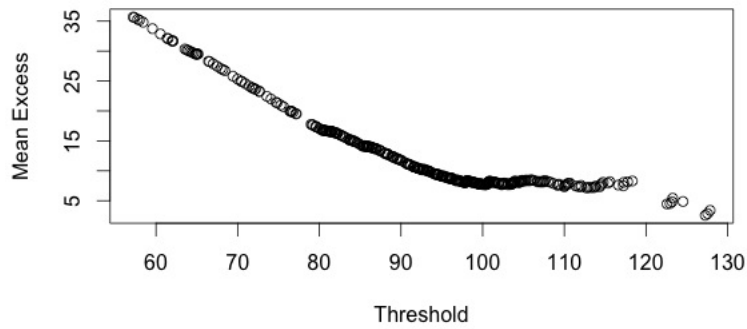
Un primer análisis exploratorio de los datos no revela una evidente tendencia ni estacionalidad, así que no es inverosímil trabajar con ellos con un acercamiento diferente al de series de tiempo. Al final, se analizará levemente también el comportamiento de los rendimientos.

Se pueden encontrar detalles acerca de la adecuación del modelo en el Apéndice 1.

### 6.2.1. Enfoque DGVE - Niveles

Para la implementación del modelo de Valores Extremos, se aplicó el método *Block Maxima* a los datos. Recordemos que para cumplir las condiciones para el Teorema de Fisher-Tippet, es necesario que tanto el número de muestras (bloques) como el número de observaciones en cada bloque sea grande. Se propone una construcción de bloques con  $\sqrt[n]{n}$ , en donde  $n$  es el total de datos con que se cuenta. Para el caso de datos completos, de un total de 544 observaciones se construyeron 23 bloques con tamaño variable entre 23 y 24 observaciones. Para el caso después de 1995, con 15 bloques de entre 15 y 16 observaciones.

Como se mencionó en el capítulo **Análisis de Riesgo**, el primer acercamiento a la identificación del dominio de atracción al cual pertenece la función  $F$  relacionada a la variable aleatoria de interés, es por medio de la gráfica de media en exceso, que se presenta a continuación.



Podemos observar que la pendiente de esta gráfica es negativa, lo que sugiere ajustar un modelo Weibull. Además, dada la naturaleza de los datos, es de interés estimar tanto valores extremos en la cola derecha como en la cola izquierda. Se hará un análisis para  $VaR_{0,95}$  y para  $VaR_{0,05}$

### Máxima verosimilitud

En esta sección haremos dos análisis paralelos: el primero, tomando los datos completos; el segundo, tomando datos solamente a partir de 1995.

Al estimar los parámetros por medio de la función *gev* del paquete *evir* de R, se obtienen los siguientes resultados. Valores negativo para  $\xi$  de nuevo sugiere un modelo Weibull. Por otro lado, tomando en cuenta los valores de error estándar para cada parámetro que arroja el paquete, obtenemos los siguientes intervalos de confianza:

### Datos completos

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.50876	-0.252	0.00476
$\sigma$	5.66488	7.216	10.76712
$\mu$	112.69028	115.368	118.04572

---

### Datos desde 1995

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.8394	-0.0458	-0.03231
$\sigma$	10.93799	12.0309	16.84711
$\mu$	92.9612	99.3014	105.6416

Ambos conjuntos de datos sugieren un modelo Weibull, dado que son mayoritariamente negativos. A pesar de que el primero contiene tanto al cero (modelo Gumbel) como valores positivos, se hará el análisis basado en Weibull, creíble dado el gran peso negativo que tienen los intervalos de confianza. Podemos estimar el VaR para dos casos extremos: sub y sobrevaluación, es decir, cuantiles bastante bajos y bastante altos. Con los parámetros estimados por máxima verosimilitud, se obtienen las estimaciones:

Datos	$VaR_{0,05}$	$VaR_{0,95}$
Desde 1970	79.40079	126.7413
Desde 1995	85.13219	115.7837

La diferencia entre los estimadores para los datos completos y los datos a partir de 1995 es clara: recordemos que gracias al tipo de cambio flotante y al libre comercio de divisas, con los niveles más recientes se tiene mucha menor variabilidad y, por lo tanto, los valores extremos delimitan un intervalo mucho menor.

El primer paso para llegar a la verosimilitud perfil con el VaR como parámetro de interés, es realizar el cambio de variable. Recordemos que, en el caso de la DGVE, se puede obtener cualquier cuantil por medio de una función de los parámetros  $\mu$ ,  $\sigma$  y  $\xi$ . De esta manera, “intercambiaremos” el parámetro  $\mu$ , que no es de especial interés, por el nuevo parámetro  $Q_{1-\alpha}$ , el cual es equivalente al VaR para un  $\alpha$  pequeña. Además, al tener a  $Q_{1-\alpha}$  como variable, podremos cuantificar, por medio del Hessiano (es decir, la matriz de información de Fisher), un intervalo de confianza para su estimador.

Para esto, utilizaremos la función *optim* del paquete *optimx* de R, la cual realiza minimizaciones de funciones con distintos algoritmos. Obtenemos tanto nuevos estimadores máximo verosímiles para el modelo, como intervalos de confianza por medio de la matriz de Información de Fisher.

Tomaremos como comportamiento de interés el de la cola derecha, pues es el que acarrea consecuencias negativas más pronto; entonces trabajaremos con dos conjuntos de parámetros:  $(\xi, \sigma, Q_{0,95})$  para medir situaciones en las que la moneda esté demasiado sobrevalorada, y  $(\xi, \sigma, Q_{0,05})$  en el caso contrario.

---

## Datos completos

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.508	-0.252	0.004
$\sigma$	5.66488	7.216	10.76712
$Q_{0,05}$	73.743	79.400	85.056
$Q_{0,95}$	120.274	126.741	143.208

## Datos desde 1995

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.8394	-0.4458	-0.05231
$\sigma$	5.414	10.012	14.610
$Q_{0,05}$	80.789	85.132	92.1931
$Q_{0,95}$	100.3861	105.784	119.528

Es importante destacar algunas características de la función *optim*. Por un lado, dado que minimiza la función objetivo, se trabaja con  $-h(\xi, \sigma, Q_{1-\alpha})$ . Por otro lado, está disponible una gama de algoritmos para realizar la minimización de la función objetivo. A continuación, se presentan las definiciones, obtenidas de Nocedal y Wright [18] y algunos comentarios acerca de cada algoritmo:

**Nelder y Mead (1965)** - este método, que utiliza un método simplex, devolvió un valor de convergencia igual a cero, lo cual significa que el método terminó exitosamente, pero un gradiente *NA*. Esto está definido en la documentación de la función como que el método no fue adecuado para la función.

**CG Fletcher and Reeves (1964)** - método basado en gradientes conjugados, devolvió los mismos mensajes que el método de Nelder y Mead. Este es un método cuyo uso principal es resolución de sistemas lineales.

**BFGS (1970)** - este algoritmo cuasi-Newtoniano usa tanto los valores de la función, como los gradientes, para crear la "superficie" que debe ser optimizado. Este método funcionó bien para los datos, devolviendo en cada iteración tanto el mismo punto máximo como una matriz Hessiana positiva, lo cual permitió obtener los errores estándar de los parámetros.

---

**SANN Belisle (1992)** - es una variante del *simulated annealing*, el cual es un método que utiliza métodos estocásticos para optimizar. Como agregado, este método funciona para funciones no diferenciables. Esta implementación simula puntos candidatos a través de un kernel Gaussiano. Este método, al ser estocástico y estar definido con un número de iteraciones limitadas, proporciona diferentes óptimos en cada iteración.

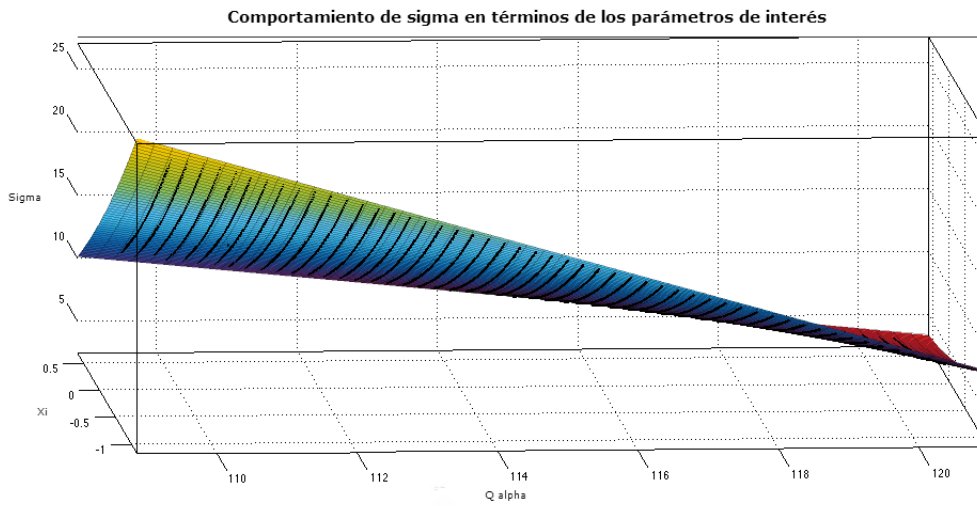
Se escogió el método **BFGS** por simplicidad y consistencia. Para el enfoque de verosimilitud perfil, dado que no existe una expresión cerrada para los estimadores que nos interesan, un método que regrese el mismo resultado, y por lo tanto permita replicar el experimento para comparar ventajas, es pertinente. El método **SANN** se sale de los límites de esta tesis, pues debería hacerse un análisis más profundo de convergencia del algoritmo. Tomando en cuenta que el límite de iteraciones dentro del algoritmo se debe a la alta complejidad computacional que tiene, un enfoque posible sería repetir el experimento y obtener un *estimador del estimador*. Sin embargo, al añadir un parámetro de varianza (pues los estimadores de los parámetros de interés no serían únicos), los intervalos de confianza de los parámetros serían más extensos, lo cual afectaría una de las principales ventajas de la verosimilitud perfil: la ganancia en precisión gracias a intervalos de confianza más estrechos en los parámetros de interés (por supuesto, a cambio de no tener estimadores de los de ruido).

## Verosimilitud perfil

Nuestro principal parámetro de interés es  $Q_{1-\alpha}$ , pues es justamente la medida de riesgo que se tomará. También es de interés  $\xi$ , para asegurarnos de que nuestro modelo Weibull es realmente un buen ajuste a los datos. Por lo tanto, se aplicará el método de Verosimilitud Perfil tomando a  $\sigma$  como parámetro de ruido.

No es posible obtener una forma analítica de los estimadores máximo verosímiles, sin embargo, podemos conocer el comportamiento de  $\sigma$  respecto a los parámetros de interés  $Q_{1-\alpha}, \xi$  por medio del Teorema de la Función Implícita. El comportamiento se puede apreciar en la siguiente figura:





Recordemos que, al aplicar el método de verosimilitud perfil, no se obtienen estimadores de los parámetros de ruido, a cambio de obtener *mejor* información acerca del comportamiento de los parámetros de interés.

Los resultados que se obtienen son:

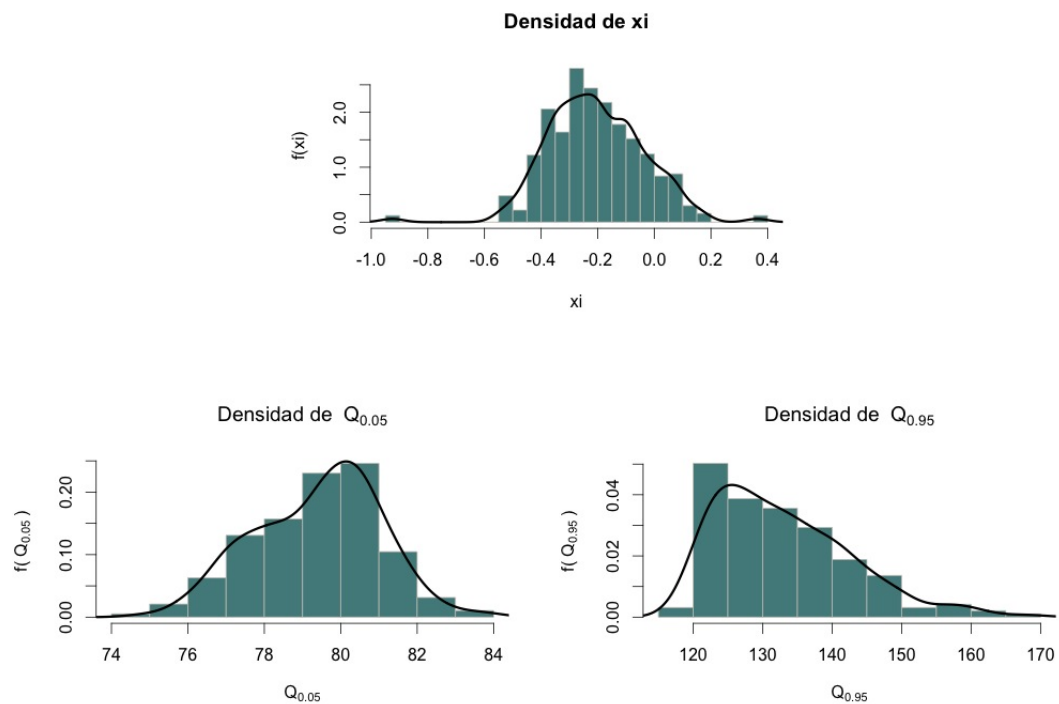
#### Datos completos

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.5120	-0.2780	0.1098
$Q_{0,05}$	77.0933	79.3961	81.5314
$Q_{0,95}$	118.4428	123.0527	149.4644

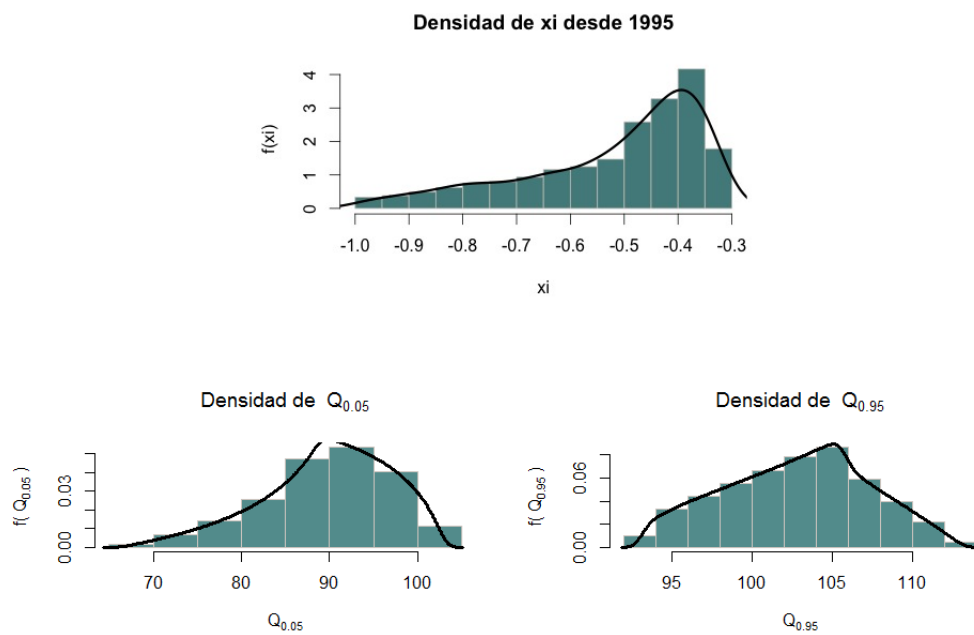
#### Datos desde 1995

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.9215	-0.4639	-0.3307
$Q_{0,05}$	75.9598	89.70512	100.2058
$Q_{0,95}$	95.00274	102.7723	110.0867

Los cuales representan el comportamiento de los pares de parámetros  $\{Q_{0,05}, \xi\}$  y  $\{Q_{0,95}, \xi\}$  cuando  $\sigma$  se encuentra en su óptimo. Las distribuciones perfil quedan representadas de la siguiente manera, para los datos completos:



Y para los datos desde 1995:



---

Existe una diferencia tangible entre las distribuciones de los estimadores perfil entre los datos completos y los reducidos (desde 1995). A pesar de que los reducidos proporcionan valores que serían interpretados como menor riesgo (el  $Var_{0,05}$  es más alto y el  $Var_{0,95}$ , más bajo), tales estimadores tienen distribuciones con forma que sugiere una mayor varianza.

Podemos observar que ninguno de los parámetros tiene una distribución simétrica. El parámetro  $\xi$  conserva la forma general de la distribución conjunta anterior a la aplicación de verosimilitud perfil, aunque se puede notar un cambio respecto a su propio parámetro de dispersión: la densidad perfil de  $\xi$  está más concentrada que la original.

Por otro lado, es interesante analizar las distribuciones del cuantil bajo  $Q_{0,05}$  y el alto  $Q_{0,95}$ . Parecen tener una especie de simetría axial: el mínimo tiene una caída mucho más rápida a la derecha del estimador perfil máximo verosímil, parece estar mucho más acotado por arriba; el máximo presenta el comportamiento contrario.

Recordemos la interpretación del Teorema de la Envolvente, visto como Verosimilitud Perfil: nos permite conocer el efecto total que tiene un parámetro sobre la distribución, juntando el efecto directo con el indirecto que tiene a través del resto de los parámetros. Por lo tanto, se conoce con una mayor precisión la manera en que los parámetros de interés se comportan.

Al obtener los estimadores Perfil de  $Q_{0,05}$  y  $Q_{0,95}$ , estamos conociendo el efecto total de cada uno de ellos; gracias a ello, podríamos tomar mejores medidas precautorias para los eventos extremos que representan. Recordemos que los datos del REER se refieren a un índice de tasa de cambio general para México. Entonces, ahora contamos con estimaciones de las dos situaciones graves que podrían suscitarse en este aspecto:

**Subvaluación** ( $Q_{0,05}$ ) : si, como valor extremo alto, el Peso Mexicano estuviera gravemente subvaluado, entonces podría decirse que se estaría vendiendo mucho más barato de su valor real. Esto significa que habría un exceso de demanda por el Peso. Sería necesario prever por medio de, por ejemplo, políticas monetarias expansivas (a corto plazo), para evitar oportunidades de arbitraje.

**Sobrevaluación** ( $Q_{0,95}$ ) : una enorme sobrevaluación significaría que el Peso está siendo vendido a un precio mucho mayor a su valor real. Esto conllevaría un exceso de oferta de Pesos Mexicanos (y por lo tanto, un exceso de demanda de monedas extranjeras). Sería necesario prever con suficientes reservas

---

para poder cubrir esta situación.

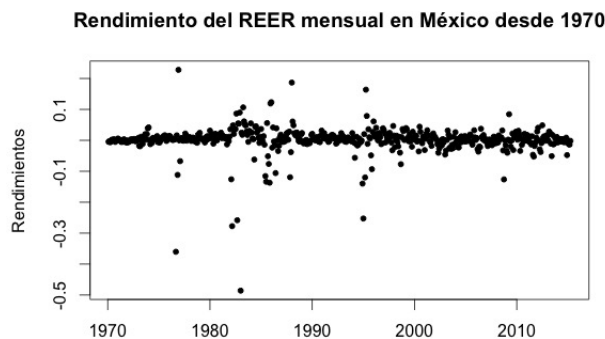
Por supuesto que no es necesario cubrir completamente los casos extremos (pues, por ejemplo, el valor del Peso tendría que tender a cero mientras su precio tiende a infinito para necesitar tener en reserva la totalidad de las monedas extranjeras) y, como fue mencionado anteriormente, un cuartil 0,95 puede ser interpretado como cubrir 19 de cada 20 observaciones de una variable sin contar la cola derecha, mientras que un cuartil 0,05 hace lo propio sin contar la cola izquierda. Las políticas fiscales y monetarias suelen cambiar a corto plazo, por lo tanto, tener cubierto poco más de un año y medio hacia adelante es bastante precautorio.

### 6.2.2. Enfoque DGVE - Rendimientos

El supuesto de independencia se puede cumplir más fácilmente si trabajamos con los rendimientos. La manera en que se calcularon los rendimientos fue:

$$r_t = \frac{REER_t - REER_{t-1}}{REER_{t-1}}$$

El histórico de rendimientos, desde 1970, se puede apreciar en la siguiente gráfica.

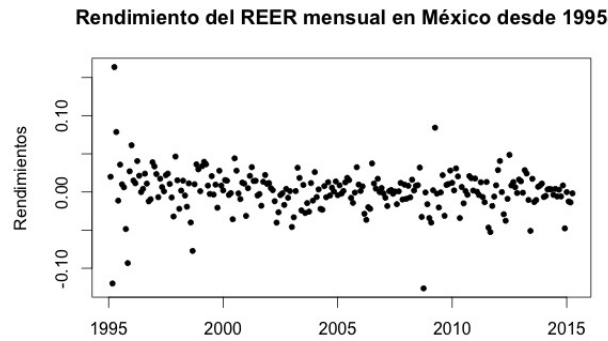


Es de esperarse que los rendimientos estén mayormente concentrados alrededor del cero; si oscilaran alrededor de un valor más alto, entonces el peso se apreciaría (o depreciaría) constantemente y surgiría una oportunidad de arbitraje. Se pueden apreciar también los eventos notables que se mencionaron en el análisis preliminar de los niveles del REER:

1. Rendimientos cercanos a cero, con muy poca variación, hasta 1982: este comportamiento obedece al tipo de cambio fijo que prevaleció hasta dicho año.

2. Rendimientos extremos en 1982: transición del tipo de cambio fijo a las bandas.
3. Rendimientos extremos en 1995: transición de las bandas al tipo de cambio flotante.
4. Rendimientos extremos en 2009: crisis económica internacional.

Si tomamos, al igual que en el análisis de los niveles, los datos a partir del 1995, entonces estaremos trabajando bajo el esquema de tipo de cambio flotante, lo cual obedece mejor a un modelo aleatorio.



Aplicando la misma metodología que el inciso anterior, obtenemos:

### Datos completos

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	0.0702	0.1173	0.8644
$\sigma$	0.0149	0.0213	0.0277
$\mu$	0.0181	0.0258	0.03352

Es de esperarse que la media (el parámetro de localización) sea bastante positivo y que las colas sean pesadas, es decir, que haya valores extremos bastante grandes. El control gubernamental del tipo de cambio forzaba al peso a ser vendido más caro de su valor real (visto de otra manera, forzaba a los dólares a ser vendidos más baratos de lo que realmente valían), por lo tanto, matemáticamente existiría siempre una oportunidad de arbitraje. Por supuesto, esta oportunidad fue controlada por medio de la limitación de dólares que podían ser comprados.

---

### Datos desde 1995

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.155	0.1149	0.3848
$\sigma$	0.0125	0.0202	0.0278
$\mu$	0.0169	0.0276	0.0383

Es interesante ver la diferencia entre los parámetros. Tanto  $\mu$  como  $\sigma$  sufren un cambio muy pequeño; sin embargo, el parámetro  $\xi$  cambia radicalmente. Con los datos completos,  $\xi$  sugiere fuertemente un modelo Fréchet, cargado hacia la derecha: es decir, los puntos extremos se encuentran más bien en el lado positivo.

Hacemos la misma sustitución que en la sección anterior, para poder tener al cuantil de interés como variable, obtenemos:

### Datos completos

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	0.0702	0.1173	0.8644
$\sigma$	0.0149	0.0213	0.0277
$Q_{0,05}$	-0.0289	-0.002605	0.0246
$Q_{0,95}$	0.0262	0.0348	0.0490

### Datos desde 1995

Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.155	0.1149	0.3848
$\sigma$	0.0125	0.0202	0.0278
$Q_{0,05}$	-0.0054	0.0097	0.0286
$Q_{0,95}$	0.0358	0.0491	0.0548

---

## Verosimilitud Perfil

Dado que la información necesaria para determinar el tipo de acciones que se tendrían que tomar depende de qué tanto podría sobrevaluarse el peso, nuestros parámetros de interés siguen siendo los mismos. Se utilizará el mismo algoritmo que en la sección anterior. Los resultados que se obtienen son:

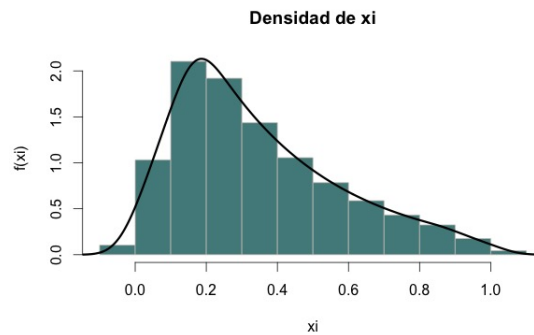
### Datos completos

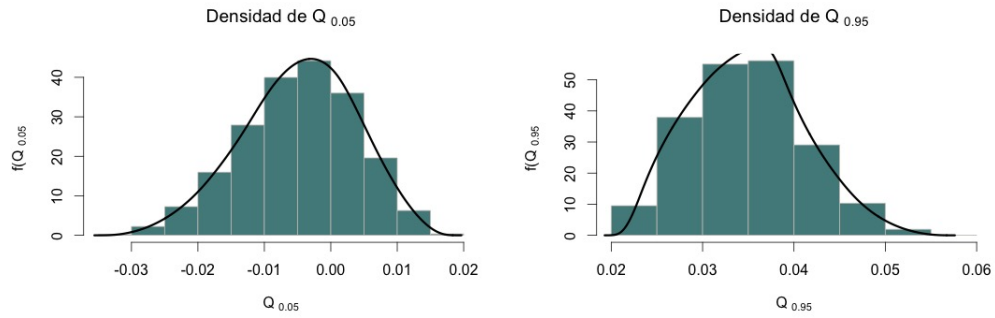
Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	0.0514	0.1148	0.8117
$Q_{0,05}$	-0.0217	-0.0026	0.0196
$Q_{0,95}$	0.0288	0.0384	0.04803

### Datos desde 1995

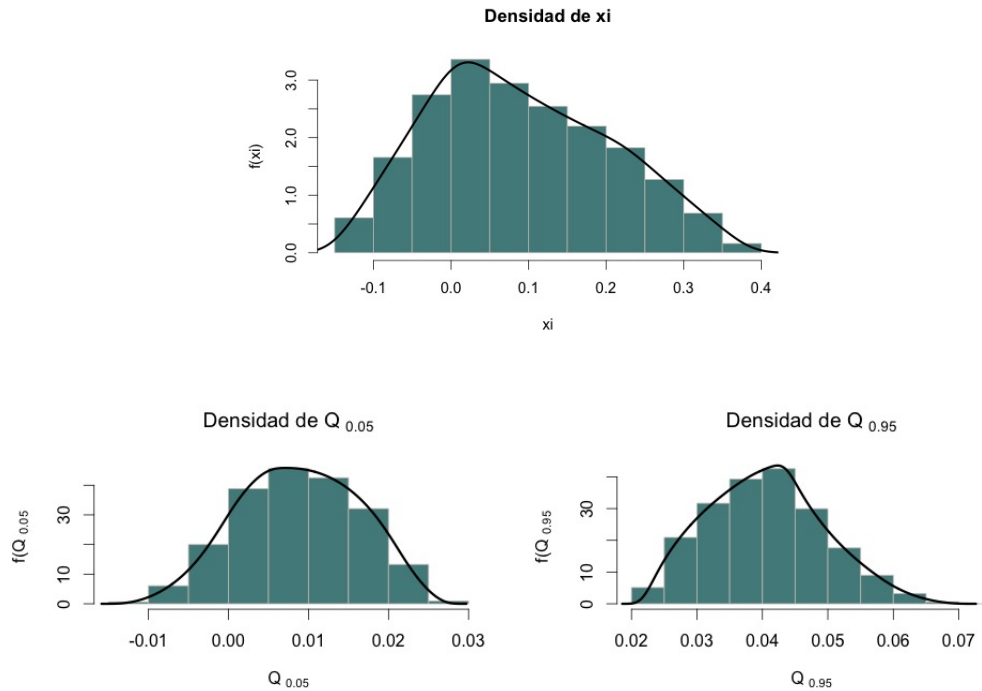
Parámetro	Lím Inf	EMV	Lím Sup
$\xi$	-0.1648	0.1190	0.3931
$Q_{0,05}$	-0.0044	0.0099	0.0283
$Q_{0,95}$	0.0320	0.0456	0.0591

Las distribuciones perfil quedan representadas de la siguiente manera:





Y para los datos desde 1995:



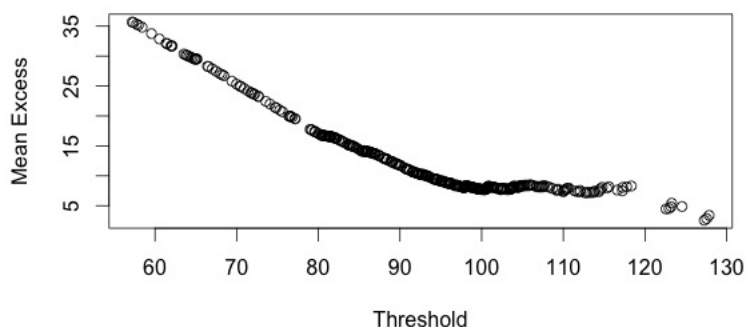
Es interesante notar que, en ambos conjuntos de datos, el cuartil bajo incluye valores negativos. Recordemos que estamos haciendo el análisis para los máximos. Estos resultados podrían interpretarse como que, dentro de las observaciones extremas de máximos, incluso habría depreciación. Con los datos desde 1995, el parámetro  $\xi$  disminuye, en incluso deja de sugerir fuertemente un modelo Fréchet, a contener en su intervalo tanto al cero como a los negativos, lo cual indica que la cola derecha pasa a ser considerablemente más corta. Esto tiene sentido, pues al dejar que el mercado decida (tipo de cambio flotante), en vez de manipulación humana (por medio del Banco), se espera que haya menos tendencia hacia un *lado de la*



---

*balanza*, precisamente para prevenir arbitraje.

### 6.2.3. Enfoque DPG



Para la selección del umbral, se recurriría de nuevo a la gráfica de medias en exceso. Se buscaría un punto en el cual la gráfica dejara de tener un comportamiento lineal. Se puede observar que el comportamiento de la gráfica cambia evidentemente a partir de  $u = 100$ ; si se fijara este umbral, se trabajaría más de una cuarta parte de los datos. Con  $u = 105,28$ , sobreviven alrededor de 15 % de los datos (60 observaciones). Solamente 10 % de los datos (40 observaciones) sobreviven si se impone  $u = 109,25$ . Es importante recordar que el modelo DPG funciona “mejor” con umbrales más altos, pero esto evidentemente significa que se trabaja con menos excesos, es decir, menos datos.

Sin embargo, los datos de REER no se apegan a un modelo Pareto. Intuitivamente, esto tiene sentido. Recordemos la ley de van der Wijk: la media en exceso crece de manera lineal al aumentar el umbral. En el caso de índices relacionados a tasas de cambio, tal comportamiento es poco factible: las fluctuaciones extremas son controladas rápidamente por la sencilla ley de la oferta y demanda. Si una moneda está sobrevalorada en exceso, la cantidad demandada será mayor, lo cual disminuirá la cantidad disponible, con lo que el precio (la tasa de cambio) volverá a subir. Esto quiere decir que cualquier comportamiento extremo está, de cierta manera, limitado (lo cual es coherente con una distribución de colas ligeras), y, por lo tanto, incompatible con una DPG.

---

### 6.3. Interpretación

El dominio de atracción del modelo Weibull comprende distribuciones cuyo dominio tiene una cota superior. Esto tiene sentido, pues el REER, al ser un promedio ponderado de variables finitas (dado que cualquier precio es finito, así que un nivel de precios también debe ser finito), está acotado, aunque el valor de la cota no sea conocido.

Por otro lado, las distribuciones pertenecientes al dominio Weibull tienen colas ligeras. Es por esto, también, que el modelo Pareto Generalizado no resultó verosímil bajo la luz de los datos. Recordemos la Ley de van der Wijk [26], que dice que el exceso sobre el umbral  $u$  es proporcional al crecimiento de  $u$  mismo en los modelos Pareto con sus generalizaciones; una distribución con colas ligeras no puede cumplir esta condición. Esto tiene sentido también, por el comportamiento de la variable: si una moneda está sobrevalorada, no caerá en un círculo vicioso de sobrevalorarse cada vez más y en mayor magnitud.

Se obtuvieron intervalos de confianza de los parámetros de interés más estrechos, en general. Esto podría ser consecuencia del mismo Método de Verosimilitud Perfil. Además, el parámetro de forma  $\xi$  aparentemente mantuvo su tipo de distribución, lo cual podría indicar que hay poca dependencia entre nuestro parámetro elegido de ruido  $\sigma$  y él. Por otro lado, esta estabilidad sugiere que el modelo elegido, Weibull, era verdaderamente adecuado para los datos, a pesar de que hubiera (pocos) valores mayores a cero en el intervalo original.

Notablemente, los resultados varían en gran proporción entre el análisis con los datos completos y los tomados a partir de 1995. Antes de la liberación del tipo de cambio como flotante, se mantenía en niveles constantes y muy controlados durante ciertos periodos de tiempo, para después explotar hacia niveles extremos. Sin embargo, dado que estos cambios fueron debidos a un control externo de los tipos de cambio, no son comparables directamente en un modelo que tiene como hipótesis la aleatoriedad de los datos. Esto se refleja claramente en el comportamiento de las distribuciones de los parámetros: con datos completos, controlados, los estimadores tienen una menor varianza pero indican que los cuantiles son más extremos, es decir, más riesgo. Con los datos sin control externo, se observa menores eventos extremos, y por lo tanto, estimadores de ambos  $VaR$  menos alejados de la media, pero las varianzas de los parámetros se vuelven más grandes.

Se observaron resultados parecidos al analizar los rendimientos: con los datos reducidos, se obtuvo incluso una posibilidad de subvaluación en el comportamiento de los máximos, lo cual era completamente

---

inverosímil bajo los datos completos. La libertad del tipo de cambio provocó que el valor falso - incluso se podría decir *inflado* - del Peso, causado por el control gubernamental, desapareciera, y así los valores del REER fueran más relacionados a su poder adquisitivo real.

---

## Capítulo 7

# Conclusiones

En esta tesis se trabajó con datos económico-financieros de un solo país, sin embargo, este método no opera bajo hipótesis relacionadas con las variables internas del REER; la ponderación de las monedas extranjeras, así como los tipos de cambio de cada uno de ellas que se consideran para determinar el nivel del índice. Esto quiere decir que el Método de Verosimilitud Perfil se podría aplicar de la misma manera presentada en esta tesis no solamente al REER de otro país, sino incluso a cualquier otro modelo en el que sean de interés solamente un subconjunto de los parámetros y se conozca el comportamiento de todos.

Se obtuvieron resultados coherentes con la teoría: al determinar a  $\sigma$  como parámetro de ruido, y no obtener estimación directa de ella, se obtuvo información acerca de los parámetros de interés (cuando  $\sigma$ , ruido, está en su óptimo) que, además de mostrar el comportamiento directo de ellos, parece ser más exacta, pues se obtuvieron intervalos de confianza más estrechos.

Se comparó el comportamiento de los parámetros tomando los datos desde 1970 contra desde 1995. Esta comparación es relevante, pues entre 1970 y 1995, el tipo de cambio estaba controlado por el gobierno: en un primer periodo, con una tasa de cambio fija para el dólar; en el segundo, con bandas de tolerancia para tal tipo de cambio. Los resultados apuntan a que el control conlleva a menor varianza para los estimadores de los parámetros, pero dichos estimadores muestran un comportamiento más extremo del índice. Por otro lado, con los datos en la época en la que se les dejó fluctuar, se observa un comportamiento menos extremo pero menos certidumbre acerca de los estimadores.

---

También se analizó el comportamiento de los *rendimientos*, los cuales se ajustan mucho mejor al supuesto de independencia. No solamente surgieron resultados relativamente esperados, como el hecho de que la media de los rendimientos es bastante menor si no se toman en cuenta todos los datos del período en el cual el tipo de cambio no era flotante; o la forma extraña de la función de distribución, con una subida estrepitosa en 0; también salieron a la luz resultados interesantes, como la factibilidad de valores extremos (máximos) que involucran rendimientos negativos, lo cual significaría que la moneda estaría debajo de su valor, las exportaciones se volverían más baratas y la economía tendría un estímulo positivo repentino.

## 7.1. La situación actual

Recordemos que en esta tesis, se trabajó con un subconjunto de los datos, para que fuera factible el supuesto de que los datos eran observaciones de una variable aleatoria.

Sin embargo, los acontecimientos recientes, relacionados a la devaluación del yuan chino y la apreciación del dólar, afectan este supuesto. El primero, causado por intervención gubernamental, parece ser una estrategia para acelerar el crecimiento de la economía china. Recordemos uno de los preceptos que se utilizaron para el análisis: si el tipo de cambio de una economía local baja, entonces sus productos se vuelven más baratos para exportar. El segundo, causado por el mercado, obedece mayormente por la disminución del precio de materias primas y por la expectativa ante la respuesta de la Reserva Federal de EUA a la recuperación de la economía de ese país después de la última crisis. Esto ha significado una depreciación fuera de lo normal.

El Banco de México puede intervenir en esta situación, por medio de subastas de dólares por ejemplo. Esta política monetaria es efectiva si los cambios son transitorios (lo cual sucedería si los cambios en el tipo de cambio fueran completamente aleatorios), pero si son de una naturaleza más permanente, sería más efectivo focalizar los esfuerzos en crecimiento económico, pues se ha visto ya que forzar al tipo de cambio a un nivel resulta contraproducente. Es decir, si el efecto parece ser más bien permanente, lo cual es sugerido por el análisis de rendimientos, sería mejor que el Banco de México tomara medidas para crecimiento acelerado de la economía, en vez de dirigidas hacia las reservas para defender el tipo de cambio.

Queda para futuros trabajos la aplicación del Método de Verosimilitud Perfil en inferencia sobre muestras con datos faltantes o censurados, modelos con un número de parámetros demasiado grande comparado con el número de observaciones con las que se cuenta para obtener sus estimadores y comparación con otras

---

variantes de Máxima Verosimilitud, como Marginal o Condicional, entre otros muchos.

---

---

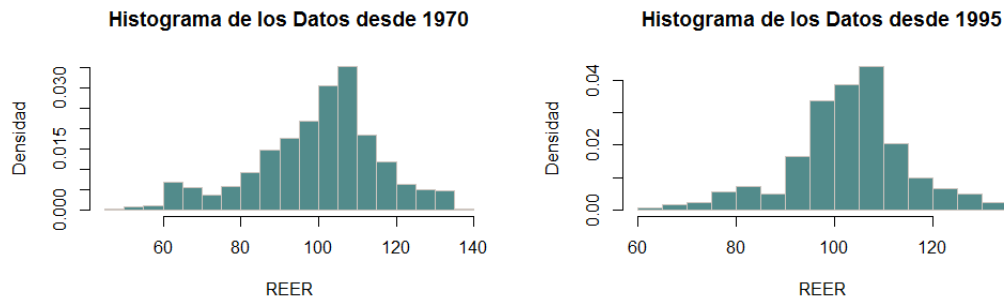
## Apéndice A

# Adecuación del modelo

El primer paso para encontrar un modelo adecuado para un conjunto de datos es, por supuesto, familiarizarse con ellos. No basta solamente con saber las restricciones teóricas (como cotas, o incluso discontinuidades). La metodología estadística en general nos permite apreciar propiedades importantes que proporcionan pistas acerca de qué modelos pueden ajustar mejor.

### A.1. Niveles

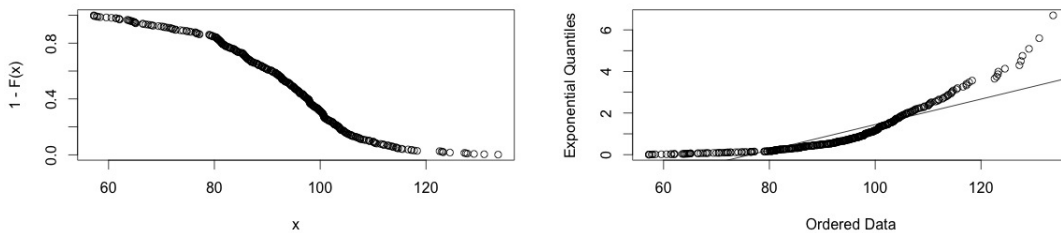
Los datos de interés en esta tesis, los niveles mensuales del REER durante 25 años, se distribuyen de la siguiente manera:



Es evidente que los datos carecen de simetría. Además, se puede ver un primer indicio de la pesadez de las colas: la sección a la izquierda del intervalo con más densidad contiene mucha más concentración de datos que

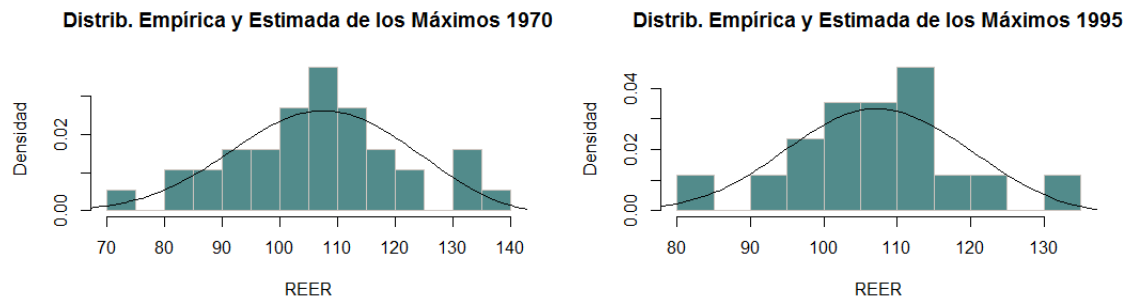


la derecha. Esto, aunado a la restricción de positividad de los datos, descarta cualquier modelo simétrico. A continuación se muestran varias gráficas de los datos mensuales del REER en México, completos: la función de supervivencia empírica, y la gráfica cuantil-cuantil contra el mejor modelo exponencial ajustable a los datos por medio de máxima verosimilitud. Sería factible, e incluso parsimonioso, ajustar este modelo, dado que se trata de datos estrictamente positivos y sin una cota superior teórica.



Sin embargo, se puede notar, con una simple inspección visual, que la forma de la cola de la distribución empírica difiere en gran medida de una exponencial, e incluso sugiere que su cola es más ligera. Esto nos da un primer indicio de qué modelos podremos usar: una distribución de colas ligeras pertenecerá al dominio de atracción de la distribución Weibull. Por otro lado, tenemos una primera sospecha de que el modelo DPG (Distribución Pareto Generalizada) podría no ser adecuado para estos datos, ya que este comprende distribuciones de colas pesadas, pues incluso el modelo de cola más ligera perteneciente a esta generalización, el exponencial, está siendo descartado por el QQ plot.

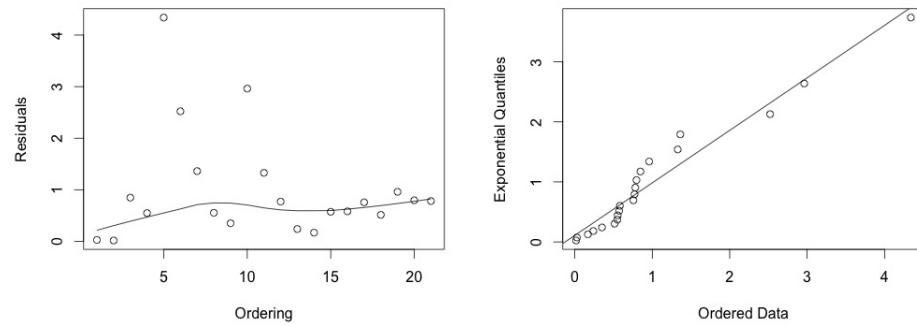
A continuación se muestran los histogramas de los máximos (método *Block Maxima*) con el modelo Weibull evaluado en los estimadores de Máxima Verosimilitud.



Para tener una mayor comprensión acerca de qué tan bien ajusta el modelo propuesto a los datos, se puede ver el comportamiento de los residuales. A continuación se muestra la distribución de los errores y

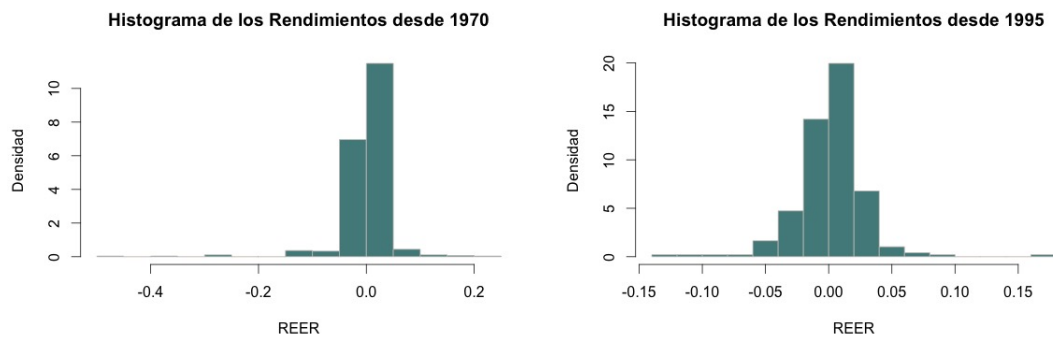
---

el QQ plot respecto al modelo Weibull máximo verosímil, para el set completo de datos.

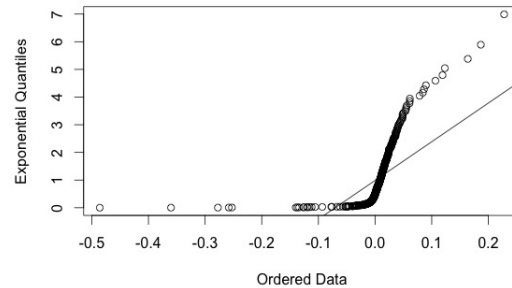
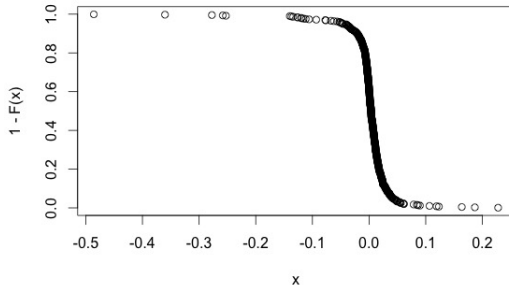


## A.2. Rendimientos

Los rendimientos mensuales del REER durante 25 años, se distribuyen de la siguiente manera:



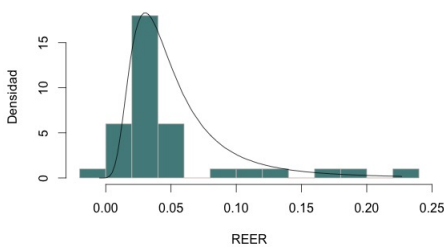
Se puede notar, desde la función de distribución empírica, que hay algo fuera de lo común en los datos. Hay una cantidad inusual de datos concentrados en el cero, y los valores positivos inmediatos. También resalta que la cola izquierda es mucho más larga que la derecha.



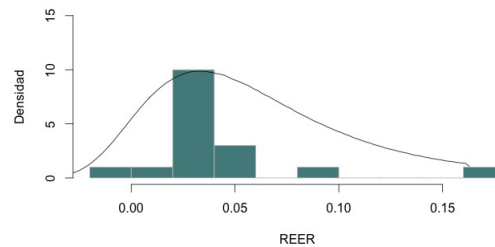
Sin embargo, se puede notar, con una simple inspección visual, que la forma de la cola de la distribución empírica difiere en gran medida de una exponencial, e incluso sugiere que su cola es más ligera. Sin embargo, hay un gran indicio de que se podría aplicar una ley de potencias (modelo Pareto Generalizado) con un umbral en 0.

A continuación se muestran los histogramas de los máximos (método *Block Maxima*) con el modelo Weibull evaluado en los estimadores de Máxima Verosimilitud.

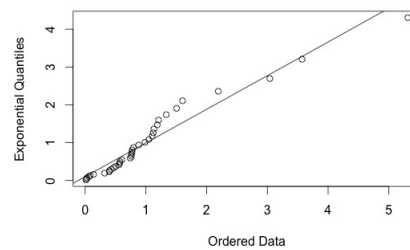
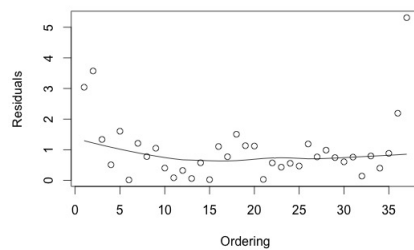
D. Empírica y Estimada de los Máx. de los Rendimientos 1970



D. Empírica y Estimada de los Máx. de los Rendimientos 1995



Para tener una mayor comprensión acerca de qué tan bien ajusta el modelo propuesto a los datos, se puede ver el comportamiento de los residuales. A continuación se muestra la distribución de los errores y el QQ plot respecto al modelo Fréchet máximo verosímil, para el set completo de datos.



---

## Apéndice B

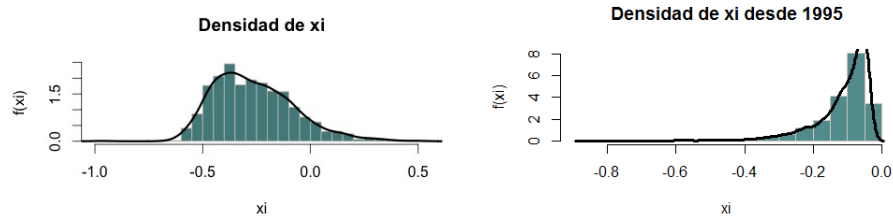
# Comportamiento de los parámetros para la DGVE

Para la DGVE<sup>1</sup>, la paquetería *evir* nos arroja los estimadores máximo verosímiles de los parámetros junto con sus respectivas desviaciones estándar. Maximizar la función numéricamente nos permite tener una comprensión más profunda de tales estimadores. Se pueden consultar estos resultados en el capítulo 6.

A continuación se presentan las densidades de los parámetros de la DGVE, obteniendo su maximización por medio de la función *optim*, utilizando el procedimiento original de máxima verosimilitud. Se tratan de resultados de inferencia basadas en los *Block Maxima* de los datos elegidos.

### Distribución del parámetro de forma $\xi$

Se puede notar que la distribución de este parámetro no tiene una forma simétrica.



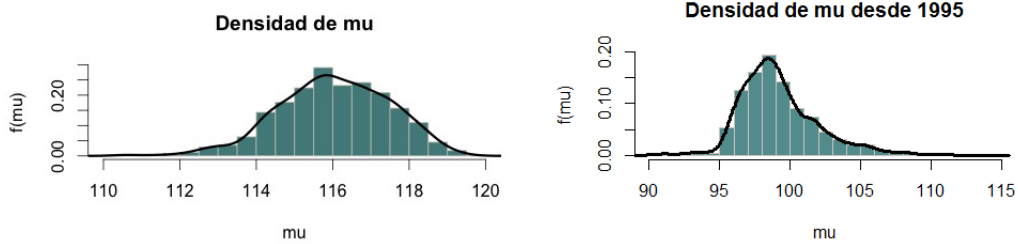
---

<sup>1</sup>Distribución Generalizada de Valores Extremos

---

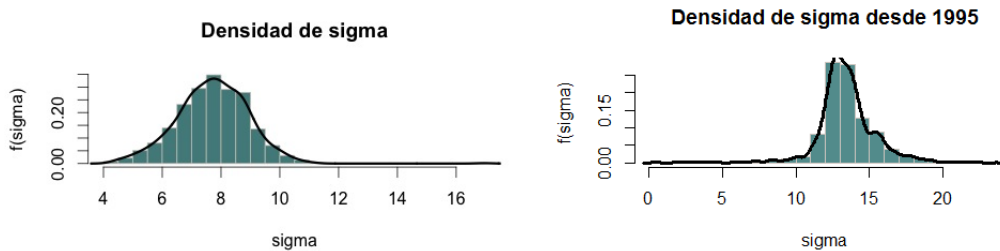
## Distribución del parámetro de localización $\mu$

Una de las facilidades acerca de los parámetros de localización, es que en muchos casos tienen distribuciones simétricas. En específico, bajo la hipótesis de DGVE y los datos del REER completos, la forma de la densidad sugiere simetría. Al considerar los datos reducidos, pero realmente estables y no bajo un control externo, tal simetría se pierde un poco.



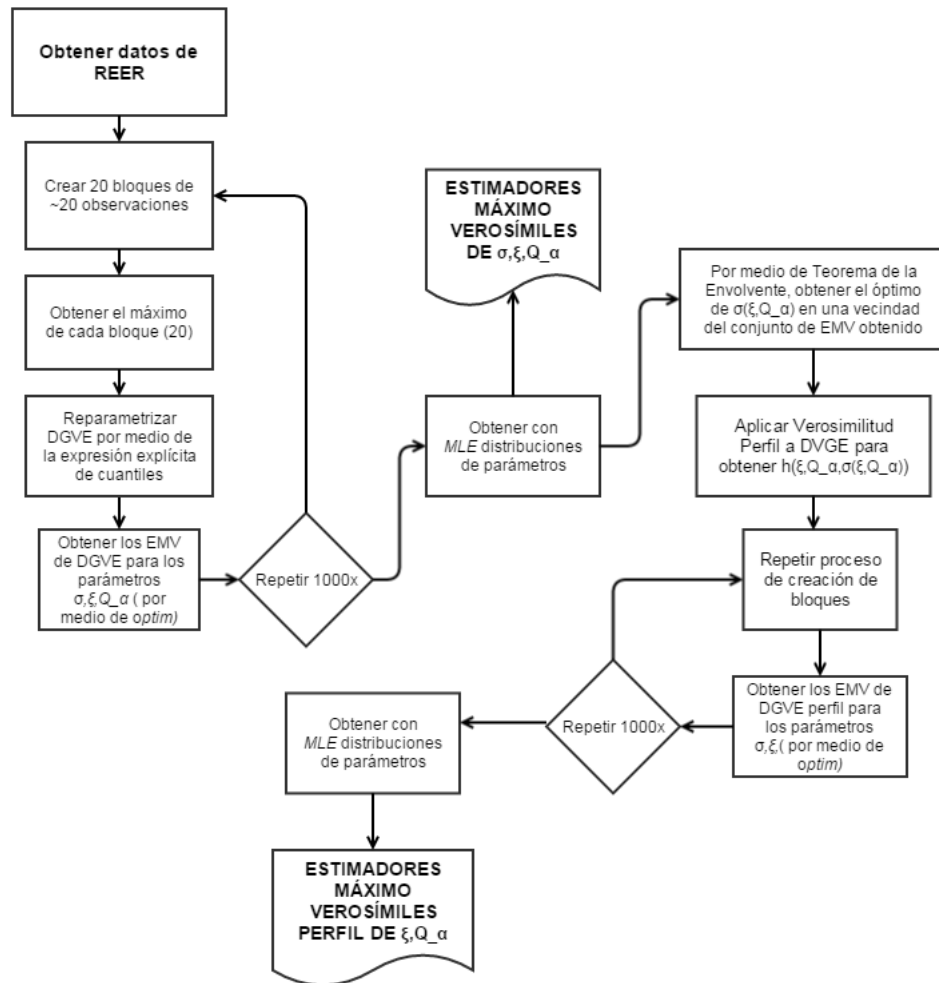
## Distribución del parámetro de escala $\sigma$

La primera, y más importante, restricción del parámetro  $\sigma$ , es que sea positivo. En comparación con la densidad de  $\mu$ , que podría sugerir normalidad, la forma de la densidad de  $\sigma$  sugiere una distribución  $\Xi^2$ , lo cual también es común en parámetros de escala. Además, los datos desde 1995 muestran una distribución más puntiaguda.



## Algoritmo

A continuación se presenta el algoritmo seguido para obtener los resultados, tanto de Máxima Verosimilitud Original como Perfil.



---

---

# Bibliografía

- [1] David Albouy. *Economics 101 Section Notes*. Berkeley, 2005.
- [2] Sergey Aleksashenko. Devaluación del 2014 ¿qué pasa con el rublo y cuándo terminará? (artículo en ruso). *Consultora de Negocios Rusia*, pág. <http://www.rbcdaily.ru/economy/562949992932863>, 12/11/2014.
- [3] A. Bolívar, E. Díaz-Francés, J. Ortega, y E. Vilchis. Profile Likelihood Intervals for Quantiles in Extreme Value Distributions. 2010.
- [4] José Juan Cáceres Hernández. *Conceptos Básicos de Estadística para Ciencias Sociales*. Delta Publicaciones, 2007.
- [5] Pasquale Cirillo. Are your data really pareto distributed? *Econophysics and Networks Across Scales, workshop at Lorentz Center Leiden*, 2013.
- [6] Richard Courant y Fritz John. *Introduction to Calculus and Analysis, Vol 2*. Springer, 1988.
- [7] T. Csendes. Nonlinear parameter estimation by global optimization—efficiency and reliability. *Acta Cybernetica 8 (4): 361–370*, págs. 361–370, 1988.
- [8] Morris H. DeGroot. *Probabilidad y Estadística*. Addison-Wesley Iberoamericana, Segunda Edición, 302 – 303, 1989.
- [9] Klueppelberg C. Embrechts, P. y T Mikosch. *Modelling Extremal Events. Vol. 33 of Applications in Mathematics*. Springer-Verlag, 1997.
- [10] Manfred Gilli y Evis Këllezi. An application of extreme value theory for measuring financial risk. *Computational Economics*, 27:1–23, 2006.



- 
- [11] Robert Hogg y Allen Craig. *Introduction to Mathematical Statistics*. Prentice Hall, 1995.
  - [12] Philippe Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 2006.
  - [13] J. G. Kalbfleisch. *Probability and Statistical Inference, Volume 2*. Springer, 1985.
  - [14] José Arturo Montoya Laos. *Tesis Doctoral: La verosimilitud perfil en la Inferencia Estadística*. CIMAT, Centro de Investigación en Matemáticas, A.C., Guanajuato, Gto. México, <http://probayestadistica.cimat.mx/sites/default/files/PDFs/TE%20269%20MONTTOYA%20LAOS.pdf>, 2008.
  - [15] Alexander McNeil, Rudiger Frey, y Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance, 1967.
  - [16] Alexander McNeil, Rüdiger Frey, y Paul Embrechts. *Quantitative Risk Management: Concepts Techniques and Tools*. Princeton University Press, 2005.
  - [17] Raymond F. Mikesell. *The Bretton Woods Debates: A Memoir*. Essays in International Finance 192 (Princeton: International Finance Section, Department of Economics, Princeton University, 1994.
  - [18] Jorge Nocedal y Stephen J. Wright. *Numerical Optimization*. Springer-Verlag, 2006.
  - [19] D. B. Owen y T. A. Hua. *Tables of Confidence Limits on the Tail Area of the Normal Distribution*. Communication and Statistics, Part B—Simulation and Computation, 6, 285–311, 1977.
  - [20] Yudi Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Clarendon Press - Oxford, 2001.
  - [21] K. Pearson. *Contributions to the Mathematical Theory of Evolution*. Phil.Trans., 1894.
  - [22] John W. Pratt. *F. Y. Edgeworth and R. A. Fisher on the efficiency of maximum likelihood estimation*. The Annals of Statistics 4 (3): 501–514, 1976.
  - [23] David E. Sprott. *Statistical Inference in Science*. Springer, 2000.
  - [24] Stephen M. Stigler. *Gauss and the Invention of Least Squares*. The Annals of Statistics 9 (3): 465–474, 1981.
  - [25] Leon Lasdon John C. Plummer Fred Glover James Kelly Ugray, Zsolt y Rafael Marti. *INFORMS Journal on Computing, Vol. 19, No. 3*, pág. 328–340.

- 
- [26] J. van der Wijk. Inkomens en vermogensverdeling (income and power distribution). *Publication of the Nederlandsch Economisch Instituut*, 26, 1939.
- [27] Hal R. Varian. *Microeconomic Analysis*. Norton, 1992.
- [28] Zhiwei Zhang. Profile likelihood and incomplete data. *International Statistical Review*, 78(1):102–116, 2010. ISSN 1751-5823.
- [29] Victoria Zinde-Walsh y John W. Galbraith. A test of singularity for distribution functions. *CIRANO*, 2011.