

GESTIÓN DEL RIESGO

Iker Caballero Bragagnini

Tabla de contenido

EL RIESGO EN PERSPECTIVA	3
EL RIESGO EN UNA INSTITUCIÓN FINANCIERA Y MÉTODOS DE VALORACIÓN	5
EL RIESGO Y LA DISTRIBUCIÓN DE PÉRDIDA	6
LA MEDICIÓN DEL RIESGO.....	13
LA DISTRIBUCIÓN NORMAL MULTIVARIANTE Y TESTS DE NORMALIDAD	29
LAS DISTRIBUCIONES NORMALES MIXTAS Y DISTRIBUCIONES GENERALIZADAS HIPERBÓLICAS	34
LAS DISTRIBUCIONES ESFÉRICAS Y ELÍPTICAS	ERROR! BOOKMARK NOT DEFINED.
LAS TÉCNICAS DE REDUCCIÓN DE DIMENSIONALIDAD	43
LA TEORÍA DEL VALOR EXTREMO: MÁXIMOS.....	49
LA TEORÍA DEL VALOR EXTREMO: EXCESOS DE UMBRALES.....	60
LA TEORÍA DEL VALOR EXTREMO: MODELOS DE PROCESOS PUNTUALES.....	ERROR! BOOKMARK NOT DEFINED.
LAS CÓPULAS	75
LAS CÓPULAS: CONCEPTOS DE DEPENDENCIA Y LAS MEDIDAS	83
LAS CÓPULAS: CÓPULAS NORMALES MIXTAS	ERROR! BOOKMARK NOT DEFINED.
LAS CÓPULAS: CÓPULAS ARQUIMEDEANAS.....	ERROR! BOOKMARK NOT DEFINED.
LAS CÓPULAS: AJUSTE A LOS DATOS	96
LAS MEDIDAS DE RIESGO COHERENTES Y CONVEXAS	ERROR! BOOKMARK NOT DEFINED.
LAS MEDIDAS DE RIESGO COHERENTES INVARIANTES A LEYES	ERROR! BOOKMARK NOT DEFINED.
LAS MEDIDAS DE RIESGO PARA CARTERAS LINEALES	ERROR! BOOKMARK NOT DEFINED.
LA AGREGACIÓN DEL RIESGO	ERROR! BOOKMARK NOT DEFINED.
LA ASIGNACIÓN DEL CAPITAL.....	ERROR! BOOKMARK NOT DEFINED.
EL RIESGO DE MERCADO: FACTORES DE RIESGO Y MAPEADO.....	ERROR! BOOKMARK NOT DEFINED.
EL RIESGO DE MERCADO: MEDICIÓN DEL RIESGO DE MERCADO	99
EL RIESGO DE MERCADO: <i>BACKTESTING</i>	ERROR! BOOKMARK NOT DEFINED.
EL RIESGO CREDITICIO Y LOS INSTRUMENTOS CREDITICIOS	104
LOS MODELOS ESTRUCTURALES DE INCUMPLIMIENTO: EL MODELO DE MERTON	123
LOS MODELOS ESTRUCTURALES DE INCUMPLIMIENTO: MODELOS PRÁCTICOS	123
LOS MODELOS DE TASA DE RIESGO PARA BONOS Y CDS.....	ERROR! BOOKMARK NOT DEFINED.
LA VALORACIÓN CON TASAS DE RIESGO ESTOCÁSTICAS.....	ERROR! BOOKMARK NOT DEFINED.
LOS MODELOS AFINES	ERROR! BOOKMARK NOT DEFINED.
LA GESTIÓN DEL RIESGO DE CARTERAS DE CRÉDITO: MODELOS DE UMBRAL	137
LA GESTIÓN DEL RIESGO DE CARTERAS DE CRÉDITO: MODELOS DE MEZCLA	ERROR! BOOKMARK NOT DEFINED.
LA GESTIÓN DEL RIESGO DE CARTERAS DE CRÉDITO: TEORÍA ASINTÓTICA PARA CARTERAS GRANDES	ERROR! BOOKMARK NOT DEFINED.
LA GESTIÓN DEL RIESGO DE CARTERAS DE CRÉDITO: MÉTODOS DE MONTE CARLO	ERROR! BOOKMARK NOT DEFINED.

LA GESTIÓN DEL RIESGO DE CARTERAS DE CRÉDITO: PROBLEMAS DEL MODELAJE DEL RIESGO... ERROR!
BOOKMARK NOT DEFINED.

LA GESTIÓN DEL RIESGO DE CARTERAS DE CRÉDITO: EL MODELO ASRF142

EL RIESGO DE CONCENTRACIÓN.....154

LA TEORÍA DEL VALOR EXTREMO: COLAS DE MODELOS ESPECÍFICOS ERROR! BOOKMARK NOT DEFINED.

LA TEORÍA DEL VALOR EXTREMO: MODELOS AUTOEXCITANTES ERROR! BOOKMARK NOT DEFINED.

LA TEORÍA DEL VALOR EXTREMO: MÁXIMOS MULTIVARIANTE ERROR! BOOKMARK NOT DEFINED.

LA TEORÍA DEL VALOR EXTREMO: EXCESOS DE UMBRALES MULTIVARIANTE ...ERROR! BOOKMARK NOT
DEFINED.

El riesgo en perspectiva

- Según el diccionario de Oxford, el riesgo se define como peligro o la posibilidad de malas consecuencias, pérdidas o exposición a la desgracia, pero para los riesgos financieros, estos se pueden intentar definir como cualquier evento o acción que afecte de manera adversa a la habilidad de conseguir los objetivos y ejecutar las estrategias de una empresa
 - El riesgo está relacionado con la incertidumbre y con la noción de aleatoriedad, de modo que se utiliza un espacio de probabilidad (Ω, \mathcal{F}, P)
 - Para poder modelar diversas situaciones financieras se puede definir una variable aleatoria que represente una posición arriesgada en un solo periodo en el espacio de probabilidad y estudiar su función de distribución de probabilidad acumulada
 - Muchas posiciones arriesgadas se denotan por un vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_d)$ y se puede introducir el tiempo mediante procesos estocásticos (\mathbf{X}_t)
 - En el contexto de las finanzas y los seguros existen diferentes tipos de riesgo que se pueden encontrar, aunque los límites de las categorías de riesgo pueden no estar muy definidos. Algunos de los cuales son los siguientes:
 - El riesgo de mercado es el riesgo de un cambio en el valor de la posición financiera debido a cambios en el valor de los componentes (activos) subyacentes de la que esta posición depende
 - El riesgo crediticio es el riesgo de no recibir los pagos prometidos de las inversiones pendientes (como préstamos y bonos) debido al incumplimiento del prestatario
 - El riesgo operacional es el riesgo de pérdidas resultantes de procesos internos inadecuados o fallidos, personas y sistemas o de eventos externos
 - El riesgo del modelo es el riesgo asociado con el uso de un modelo inapropiado o mal especificado para medir el riesgo
 - El riesgo de liquidez es el riesgo que nace de la falta de comerciabilidad de una inversión que no puede comprarse o venderse lo suficientemente rápido como para prevenir o minimizar una pérdida

- El riesgo de suscripción o *underwriting risk* es el riesgo inherente a las pólizas de seguro y servicios similares, tales como cambios en patrones de catástrofes naturales, demografía o comportamiento del consumidor
- Los conceptos y técnicas utilizadas suelen aplicarse a las primeras tres categorías de riesgo
 - Para poder manejar el riesgo de manera viable se tienen que utilizar un enfoque holístico (un enfoque integrado que toma todos los tipos de riesgo y sus interacciones en cuenta)
- Es necesario diferenciar entre la medición del riesgo y la gestión del riesgo
 - La medición del riesgo es esencialmente un problema estadístico que se basa en las observaciones históricas y, dado un modelo específico, una estimación estadística de la distribución del cambio de valor de una posición o uno de sus funcionales
 - Una buena medición del riesgo es un deber para las instituciones financieras ya que los clientes demandan información objetiva y detallada de los productos comprados y hay consecuencias legales de no proporcionar información adecuada. Por lo tanto, para cualquier producto vendido, una adecuada cuantificación de los riesgos subyacentes es necesaria para permitir al cliente decidir si un producto corresponde a su apetito al riesgo o no
 - La gestión del riesgo es una disciplina que se centra en como vivir con la posibilidad de que eventos futuros puedan causar efectos adversos. Usando experiencia, posición de mercado y estructura de capital, una institución financiera puede manejar riesgos empaquetándolos y transfiriéndolos a los mercados de manera personalizada
 - Por lo tanto, la gestión del riesgo está relacionada con preservar el flujo de beneficios y a técnicas que pueden definirse como el manejo de una institución financiera con tal de obtener un rendimiento de los fondos invertidos y mantener un cómodo superávit de activos más allá de sus pasivos
- A lo largo del tiempo ha habido desarrollos importantes para establece un marco regulatorio para la gestión del riesgo
 - La mayoría de la regulación se originó del comité de supervisión bancaria de Basel, el cual se estableció por los gobernadores de los bancos centrales del G-10 a finales del 1974

- Este no tiene ningún órgano supervisor supranacional, por lo que sus conclusiones no tienen fuerza legal, pero se formulan estándares de supervisión, guías y recomendaciones de buenas prácticas con la esperanza de que las autoridades individuales las implementen en su sistema nacional
- El primer acuerdo de Basel, conocido como Basel I, se realizó en 1988 y fue un paso importante para el estándar de capital mínimo internacional, enfocándose sobre todo en el riesgo crediticio (el más importante en su momento)
 - No obstante, este fue insatisfactorio en cuanto a su extensión, su medición del riesgo no diferenciada y el tratamiento de los derivados financieros
 - En 1993, el G-30 publicó un reportaje sobre los productos fuera del balance, como los derivados, de manera sistemática, y la industria vio la necesidad de la gestión del riesgo (dado que se necesita valoración de las posiciones de manera constante y se necesitaba un mejor entendimiento de posiciones involucrando los mismos subyacentes)
 - Consecuentemente, en 1996 se hizo una modificación al Basel I para incluir un modelo estandarizado del riesgo mercado y permitir el uso de modelos basados en la medida del VAR. No obstante, la implementación legal no ocurrió hasta el 2000

El riesgo en una institución financiera y métodos de valoración

- Los dos marcos regulatorios más importantes para la gestión del riesgo son el marco Basel, sobretodo enfocada a instituciones financieras como bancos, y el marco Solvency, sobretodo enfocada a instituciones aseguradoras
 - Una de las características clave del marco Basel es el concepto de los tres pilares
 - Bajo el primer pilar, los bancos requieren calcular un cargo mínimo de capital, llamado capital regulatorio, aunque especifican diferentes cargos dependiendo de si es riesgo de crédito, riesgo de mercado o riesgo operacional. La mayoría de bancos utilizan modelos internos basados en la metodología VaR para poder determinar el cargo de capital para el riesgo de mercado, mientras que se utilizan otros enfoques para el riesgo de crédito y el riesgo operacional

- El segundo pilar reconoce que cualquier enfoque cuantitativo para la gestión del riesgo tiene que estar incluido en una estructura de gobernanza corporativa funcional. Las buenas prácticas en gestión del riesgo imponen restricciones en la organización de la institución
- A través del segundo pilar, también llamado proceso de revisión de supervisión, los reguladores locales revisan las cuentas de las entidades, y además también expresa que los riesgos residuales cuantificables no incluidos en el primer pilar se tienen que considerar y se tienen que realizar *stress tests* para la adecuación de capital de un banco. El objetivo es asegurar que el banco mantiene capital en línea con su pérdida económica potencial verdadero, lo cual se denomina capital económico
- Finalmente, el tercer pilar busca establecer disciplina de mercado a través de una mejor publicación de medidas de riesgo y otra información relevante para la gestión del riesgo. En particular, los bancos tendrían que ofrecer mayor visibilidad en la adecuación de su capitalización

El riesgo y la distribución de pérdida

- La representación de la incertidumbre sobre los estados futuros del mundo se basa en un espacio de probabilidad (Ω, \mathcal{F}, P) y se considera una cartera de bonos y acciones, de derivados financieros, de préstamos arriesgados o la posición entera de una institución financiera en activos arriesgados
 - Se denota el valor de esta cartera en el momento s por $V(s)$ y se asume que $V(s)$ es una variable aleatoria observable en el momento s
 - Para un horizonte temporal determinado Δ , la pérdida $L_{[s, s+\Delta]}$ de la cartera en el periodo $[s, s + \Delta]$ es la siguiente:

$$L_{[s, s+\Delta]} \equiv -(V(s + \Delta) - V(s))$$
 - Aunque $L_{[s, s+\Delta]}$ se puede observar en el momento $s + \Delta$, normalmente se considera aleatoria desde el punto de vista de s
 - La distribución de $L_{[s, s+\Delta]}$ es la distribución de pérdidas, y se distingue la distribución condicional, la cual es la distribución de $L_{[s, s+\Delta]}$ dada toda la información disponible hasta e incluyendo el momento s , y la distribución incondicional
 - Los practicantes de la gestión del riesgo normalmente se centran en la distribución de pérdidas y ganancias o P&L, la cual es la distribución del

cambio de valor de la cartera $V(s + \Delta) - V(s)$ (de la variable aleatoria $-L_{[s,s+\Delta]}$)

- No obstante, en la gestión del riesgo uno se centra en la probabilidad de grandes pérdidas y, por tanto, con la cola superior de la distribución de pérdidas. Esto, a su vez, facilita la aplicación de técnicas de campos como actuariales y de estadística
- Como normalmente se utiliza un horizonte fijo Δ , es conveniente medir en unidades temporales de Δ e introducir notación de series temporales de $(Y_t)_{t \in \mathbb{N}}$ en donde $Y_t \equiv Y(t\Delta)$. Por lo tanto, la pérdida se puede escribir de la siguiente manera:

$$L_{t+1} \equiv L_{[t\Delta, (t+1)\Delta]} = -(V_{t+1} - V_t)$$

- Por ejemplo, en la gestión del riesgo en el mercado financiero se trabaja con el tiempo de calendario s se mide en años y los tipos de interés y la volatilidad se cotizan con una base anualizada. Se suelen utilizar pérdidas diarias $\Delta = 1/365$ o pérdidas de *trading days* $\Delta \approx 1/250$
- Por lo tanto, V_t y V_{t+1} representan el valor de la cartera en los días t y $t + 1$, respectivamente, y L_{t+1} representa la pérdida del día t al $t + 1$
- Una práctica estándar en la gestión del riesgo es la modelización del valor V_t como una función medible $f: \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathbb{R}$ del tiempo y de un vector aleatorio d -dimensional $\mathbf{Z}_t = (Z_{t1}, Z_{t2}, \dots, Z_{td})'$ de factores de riesgo, lo cual se denomina mapeado de riesgos

$$V_t = f(t, \mathbf{Z}_t)$$

- Los factores de riesgo de \mathbf{Z}_t se asumen observables en t (se conocen en t) y la selección de factores de riesgo y de f depende del problema de modelaje y de la cartera con la que se trabaja, pero se suelen usar precios logarítmicos, rendimientos y tipos de cambio logarítmicos
- Las series de cambios en los factores de riesgo $(\mathbf{X}_t)_{t \in \mathbb{N}}$ se define por $\mathbf{X}_t \equiv \mathbf{Z}_t - \mathbf{Z}_{t-1}$ y suelen ser los objetos de estudio más importantes en series temporales financieras. Usando el mapeado
- Utilizando el mapeado de riesgos, la función de pérdida de la cartera se puede representar de la siguiente manera:

$$L_{t+1} \equiv -(f(t+1, \mathbf{Z}_t + \mathbf{X}_{t+1}) - f(t, \mathbf{Z}_t))$$

- Como \mathbf{Z}_t es conocida en el momento t , la distribución de la pérdida se determina por la distribución del cambio en el factor de riesgo \mathbf{X}_{t+1} (el único componente aleatorio es este)
- El operador de pérdidas $l_{[t]}: \mathbb{R}^d \rightarrow \mathbb{R}$ mapea los cambios en el factor de riesgo a las pérdidas, y se define de la siguiente manera:

$$l_{[t]}(\mathbf{x}) \equiv -(f(t+1, \mathbf{Z}_t + \mathbf{x}) - f(t, \mathbf{Z}_t))$$

$$\Rightarrow L_{t+1} = l_{[t]}(\mathbf{X}_{t+1}) = -(f(t+1, \mathbf{Z}_t + \mathbf{X}_{t+1}) - f(t, \mathbf{Z}_t))$$

- Si f es diferenciable, se considera una aproximación lineal de primer orden L_{t+1}^Δ de la pérdida a partir del operador lineal de pérdida $l_{[t]}^\Delta$, que tienen la siguiente forma:

$$l_{[t]}^\Delta(\mathbf{x}) \equiv -\left(\frac{\partial}{\partial t}f(t, \mathbf{Z}_t) + \sum_{i=1}^d \frac{\partial}{\partial z_i}f(t, \mathbf{Z}_t)x_i\right)$$

$$\Rightarrow L_{t+1}^\Delta \equiv -\left(\frac{\partial}{\partial t}f(t, \mathbf{Z}_t) + \sum_{i=1}^d \frac{\partial}{\partial z_i}f(t, \mathbf{Z}_t)X_{t+1,i}\right)$$

- El operador de pérdida lineal que corresponde a la aproximación es conveniente porque permite representar la pérdida como una función lineal de los cambios en el factor de riesgo. Esta fórmula funciona mejor cuando los cambios en los factores de riesgo son pequeños o si la cartera es casi lineal en los factores de riesgo (si la función tiene segundas derivadas pequeñas)
- Los resultados anteriores se han desarrollado asumiendo de manera implícita las unidades temporales y la composición de la cartera
 - Al desarrollar las fórmulas anteriores se ha asumido que el tiempo se media en unidades del horizonte temporal Δ , pero la convención del mercado suele ser medir el tiempo en años
 - Con tal de ser consistente con la convención del mercado, se consideran mapeados de la forma $g(s, \mathbf{Z})$ donde el tiempo $s = \Delta t$ se mide en años. Por lo tanto, se dan las siguientes igualdades:

$$V_t = f(t, \mathbf{Z}_t) = g(t\Delta, \mathbf{Z}_t)$$

$$\Rightarrow L_{t+1} = -\left(g((t+1)\Delta, \mathbf{Z}_t + \mathbf{X}_{t+1}) - g(t\Delta, \mathbf{Z}_t)\right)$$

$$\Rightarrow L_{t+1} = -\left(\frac{\partial}{\partial s} g(t\Delta, \mathbf{Z}_t)\Delta + \sum_{i=1}^d \frac{\partial}{\partial z_i} g(t\Delta, \mathbf{Z}_t) X_{t+1,i}\right)$$

- El término $g(t\Delta, \mathbf{Z}_t)\Delta$ es un término pequeño cuando se trabaja en un horizonte temporal pequeño Δ , por lo que en la práctica se puede eliminar de la ecuación
- La definición dada de las pérdidas de una cartera asume implícitamente que la composición de la cartera se mantiene intacta a lo largo del horizonte Δ
 - Aunque esto no sea importante en las pérdidas diarias, si que es relevante y no realista para horizontes temporales largos
- Para poder trabajar con pérdidas en muchos periodos de tiempo, lo más simple es agregar los datos, por lo que se tiene que transformar el proceso y resulta en la siguiente función de pérdida:

$$\begin{aligned} L_{t+h}^{(h)} &= -(V_{t+h} - V_t) = -(f(t+h, \mathbf{Z}_{t+h}) - f(t, \mathbf{Z}_t)) = \\ &= -(f(t+h, \mathbf{Z}_t + \mathbf{X}_{t+1} + \dots + \mathbf{X}_{t+h}) - f(t, \mathbf{Z}_t)) \equiv l_{[t]}^{(h)} \left(\sum_{i=1}^h \mathbf{X}_{t+i} \right) \end{aligned}$$

- Este método hace que se pierdan observaciones, y aunque se pueden hacer cálculos cuando se sobreponen datos, esto hace que se añada dependencia entre observaciones
- El operador puede ser linealizado de la siguiente manera:

$$L_{t+h}^{(h)\Delta} = l_{[t]}^{(h)\Delta} \left(\sum_{i=1}^h \mathbf{X}_{t+i} \right) = \sum_{i=1}^h \mathbf{b}_t' \mathbf{X}_{t+i}$$

where \mathbf{b}_t is any vector

- La distribución de la función de pérdida linealizada para factores de riesgo no correlacionados \mathbf{X} , por tanto, es la siguiente:

assuming $\mathbf{X}_{t+i} \sim N(\boldsymbol{\mu}, \Sigma)$:

$$E \left(\sum_{i=1}^h \mathbf{b}_t' \mathbf{X}_{t+i} \right) = h \mathbf{b}_t' \boldsymbol{\mu} \quad \text{Var} \left(\sum_{i=1}^h \mathbf{b}_t' \mathbf{X}_{t+i} \right) = h \mathbf{b}_t' \Sigma \mathbf{b}_t$$

$$\Rightarrow L_{t+h}^{(h)\Delta} \sim N(\mathbf{b}_t' h \boldsymbol{\mu}, h \mathbf{b}_t' \Sigma \mathbf{b}_t)$$

$$\Rightarrow \frac{L_{t+h}^{(h)\Delta} - h \mathbf{b}_t' \boldsymbol{\mu}}{\sqrt{h \mathbf{b}_t' \Sigma \mathbf{b}_t}} \sim N(\mathbf{0}, \mathbf{1}_{d \times d})$$

- Como se ha mencionado, a veces uno está interesado en la distribución incondicional o en la condicional, dado que ambas son relevantes para la gestión del riesgo, pero es necesaria su distinción

- Se supone que los cambios en los factores de riesgo forman una serie temporal con una distribución estacionaria F_X en \mathbb{R}^d (la distribución no varía por desplazamientos del tiempo)

- Fijando el momento actual t y denotando como \mathcal{F}_t el espacio de eventos que representa la información pública disponible en el momento t , se puede denotar como $F_{X_{t+1}|\mathcal{F}_t}$ a la distribución condicional de X_{t+1} dada la información actual \mathcal{F}_t
- Normalmente $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$, que es el espacio de eventos generado por los valores presentes y pasados de los cambios en los factores de riesgo, normalmente llamado historial o *history* de X hasta e incluyendo t
- En la mayoría de modelos de gestión del riesgo se cumple que $F_{X_{t+1}|\mathcal{F}_t} \neq F_X$, pero si $(X_t)_{t \in \mathbb{N}}$ es una serie independiente e idénticamente distribuida, entonces $F_{X_{t+1}|\mathcal{F}_t} = F_X$

- Fijando el operador $l_{[t]}$ correspondiente a la cartera, la distribución condicional de pérdidas $F_{L_{t+1}|\mathcal{F}_t}$ se define como la distribución del operador $l_{[t]}$ bajo $F_{X_{t+1}|\mathcal{F}_t}$

$$F_{L_{t+1}|\mathcal{F}_t} = P(l_{[t]}(X_{t+1}) \leq l | \mathcal{F}_t) = P(L_{t+1} \leq l | \mathcal{F}_t) \text{ for } l \in \mathbb{R}$$

- Esta distribución es especialmente útil para la gestión del riesgo de mercado, y las técnicas de gestión utilizadas con la distribución condicional se denominan gestión de riesgo dinámica o condicional
- En cambio, la distribución incondicional de pérdidas $F_{L_{t+1}}$ se define como la distribución de $l_{[t]}$ bajo la distribución estacionaria F_X de los cambios en los factores de riesgo, y esta da la distribución de pérdidas en la cartera si se considera un cambio de los factores de riesgo genérico X con la X_1, X_2, \dots, X_t
- Esta distribución es especialmente útil para medir las pérdidas en un horizonte temporal relativamente largo (como en la

gestión del riesgo crediticio y los seguros) y las técnicas de gestión utilizadas con la distribución incondicional se denominan gestión de riesgo estática

- Existen algunos ejemplos importantes en el área del riesgo de mercado y de crédito que permiten tener un marco de análisis
 - Considerando una cartera de d acciones y denotando por λ_i el número de acciones de la acción i en la cartera en el momento t y por $(S_{i,t})_{t \in \mathbb{N}}$ el proceso de una acción i , se puede utilizar el logaritmo de los precios como factores de riesgo (de modo que $Z_{i,t} \equiv \ln S_{i,t}$)

- Por lo tanto, el cambio en los factores de riesgo será $X_{i,t+1} = \ln S_{i,t+1} - \ln S_{i,t}$ y el valor de la cartera en el momento t es el siguiente:

$$V_t = \sum_{i=1}^d \lambda_i \exp(Z_{i,t}) = \sum_{i=1}^d \lambda_i S_{i,t}$$

- En consecuencia, se pueden obtener expresiones para L_{t+1} y L_{t+1}^Δ , en donde $w_{i,t} = (\lambda_i S_{i,t})/V_t$ da la proporción del valor de la cartera invertida en la acción i

$$L_{t+1} = -(V_{t+1} - V_t) = -\sum_{i=1}^d \lambda_i S_{i,t} (\exp(X_{i,t+1}) - 1)$$

$$L_{t+1}^\Delta = -\sum_{i=1}^d \lambda_i S_{i,t} X_{i,t+1} = -V_t \sum_{i=1}^d w_{i,t} X_{i,t+1}$$

- El operador lineal correspondiente se define por la siguiente expresión:

$$l_{[t]}^\Delta(\mathbf{x}) \equiv -V_t \mathbf{w}_t' \mathbf{x} = -V_t \sum_{i=1}^d w_{i,t} x_i$$

- Suponiendo que el vector aleatorio \mathbf{X} sigue una distribución con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas y covarianzas Σ , se pueden usar las reglas generales de la esperanza y la varianza para obtener los primeros dos momentos del operador lineal:

$$E(l_{[t]}^\Delta(\mathbf{X})) = E(-V_t \mathbf{w}_t' \mathbf{X}) = -V_t \mathbf{w}_t' E(\mathbf{X}) = -V_t \mathbf{w}_t' \boldsymbol{\mu}$$

$$\begin{aligned} \text{Cov}\left(l_{[t]}^{\Delta}(X)\right) &= \text{Cov}(-V_t \mathbf{w}_t' X) = -V_t^2 \mathbf{w}_t' \text{Cov}(X) \mathbf{w}_t = \\ &= -V_t^2 \mathbf{w}_t' \Sigma \mathbf{w}_t \end{aligned}$$

- Si se utiliza la distribución condicional $F_{X_{t+1}|\mathcal{F}_t}$, entonces los resultados anteriores se mantienen, pero se cambian los vectores de media y la matriz de varianza y covarianza por $\boldsymbol{\mu}_t$ y Σ_t respectivamente. Estos permiten obtener los momentos de la distribución condicional e incondicional de las pérdidas
- Considerando una cartera de opciones de compra europeas sobre una acción sin dividendos con precio S y que tienen un vencimiento T y precio de ejercicio K . Para la valoración, se utiliza la fórmula de Black-Scholes-Merton para opciones de compra europeas para la valoración de la cartera

$$C^{BS}(t, S; r, \sigma, K, T) = S\Phi(d_1) - Ke^{r(T-t)}\Phi(d_2)$$

$$\text{where } d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}} \quad \& \quad d_2 = d_1 - \sigma\sqrt{T-t}$$

- Esta expresión se mide en años porque sigue la convención del mercado, por lo que $\Delta = 1/250$
- Un factor de riesgo es el precio logaritmico del activo subyacente, y aunque en la fórmula se asume que r y σ son constantes, estas varían y se suele escoger un vector de factores de riesgo $\mathbf{Z}_t = (\ln S_t, r_t, \sigma_t)'$. Bajo la suposición de que el valor de la cartera $V_t = C^{BS}(t\Delta, S_t; r_t, \sigma_t, K, T)$, se puede ver como los cambios de los factores de riesgo permiten definir la función de pérdida lineal:

$$V_t = C^{BS}(t\Delta, S_t; r_t, \sigma_t, K, T) = S\Phi(d_1) - Ke^{r(T-t\Delta)}\Phi(d_2)$$

$$\Rightarrow \mathbf{X}_{t+1} = (\ln(S_{t+1}) - \ln(S_t), r_{t+1} - r_t, \sigma_{t+1} - \sigma_t)'$$

$$\Rightarrow L_{t+1}^{\Delta} = -\left(\frac{\partial C^{BS}}{\partial t}\Delta + \frac{\partial C^{BS}}{\partial S}S_t X_{t+1} + \frac{\partial C^{BS}}{\partial r}X_{t+1,2} + \frac{\partial C^{BS}}{\partial \sigma}X_{t+1,3}\right)$$

- Estas derivadas son las griegas, las cuales tienen un rol prominente en la gestión del riesgo en una cartera de derivados, y estas son las siguientes:

$$\text{Theta} = \frac{\partial C^{BS}}{\partial t} \quad \text{Delta} = \frac{\partial C^{BS}}{\partial S}$$

$$Rho = \frac{\partial C^{BS}}{\partial r} \quad Vega = \frac{\partial C^{BS}}{\partial \sigma}$$

- La pérdida lineal puede ser una aproximación pobre para la pérdida verdadera, ya que el valor de una cartera de derivados suele ser altamente no lineal en los factores de riesgos, haciendo que se hagan aproximaciones de mayor orden (aproximaciones lineales de segundo orden como la aproximación delta-gamma)

La medición del riesgo

- En la práctica, la medición del riesgo se usa para una variedad de razones y hay diferentes enfoques para poder medir el riesgo de las instituciones financieras
 - Las razones más importantes para las que se mide el riesgo son para la determinación del capital riesgo y la adecuación del capital, para usarla como herramienta de gestión y para determinar las primas de seguros
 - Una de las principales funciones de la gestión del riesgo en el sector financiero es determinar la cantidad de capital que una institución financiera necesita mantener como un recurso contra pérdidas futuras inesperadas en su cartera con tal de satisfacer a los reguladores. Debido a esto, un problema asociado es el de determinar requerimientos de margen apropiados para inversores que comercian en bolsa, que suele hacer mediante una *clearing house*
 - Las medidas de riesgo normalmente se usan como una herramienta de gestión para limitar la cantidad de riesgo que una unidad dentro de la empresa debería tomar
 - Las primas de seguros compensan a las aseguradoras por soportar el riesgo de los reclamos asegurados. El tamaño de la compensación se puede ver como una medida del riesgo de estos reclamos
 - Los enfoques existentes se pueden clasificar en cuatro categorías: el enfoque de monto nocional, las medidas a sensibilidad de factores, medidas basadas en la distribución de pérdidas y las medidas basadas en escenarios
 - En el enfoque de monto nocional, el riesgo de la cartera se define como la suma de los montos nocionales de los activos individuales en la cartera, donde cada monto puede estar ponderado por un factor que representa una evaluación del riesgo de la clase al que el activo pertenece. La ventaja de este enfoque es su aparente simplicidad, pero también tiene

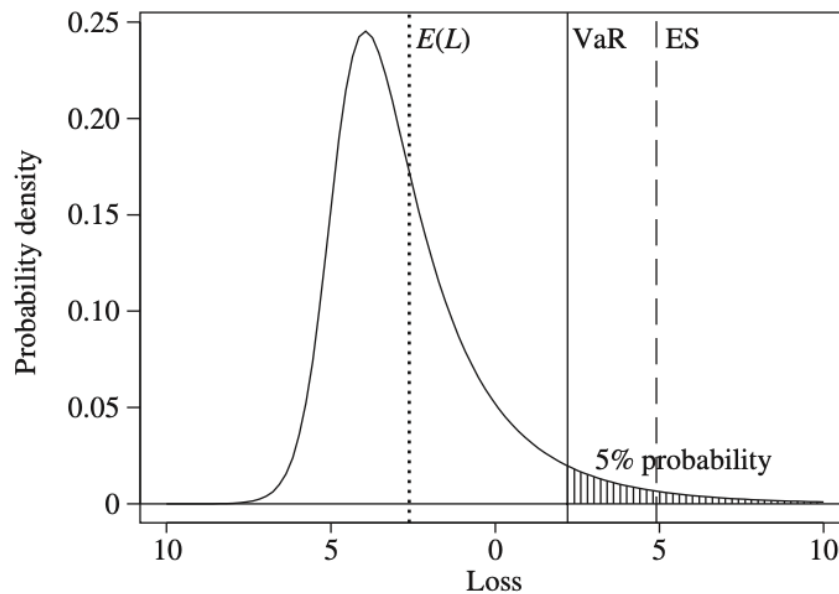
problemas, tales como no diferenciar entre posiciones largas o cortas (no se hace un neto), no reflejar los beneficios de la diversificación en el riesgo total de la cartera o no funcionar correctamente con carteras con derivados (el valor económico del subyacente puede diferir mucho del valor de la posición en el derivado)

- Las medidas a sensibilidad de factores dan el cambio en el valor de la cartera para un cambio predeterminado en uno de los factores de riesgo subyacentes (normalmente en forma de derivadas). Mientras que estas medidas aportan información sobre la robustez de la cartera para ciertos eventos definidos, no pueden medir el riesgo total de la cartera y se pueden crear problemas en la agregación de riesgos (no se puede agregar la sensibilidad con respecto a cambios de diferentes factores y no se pueden agregar las medidas entre mercados)
- La mayoría de medidas del riesgo son cantidades estadísticas describiendo la distribución de pérdidas condicional o incondicional sobre un periodo de tiempo determinado. Aunque sea problemático solo usar un estadístico, el uso de varios para describir la función de distribución es adecuado porque las pérdidas son el objeto central de la gestión del riesgo y es natural usar medidas basadas en su distribución; el concepto de distribución de pérdidas tiene sentido en todos los niveles de agregación de la cartera; se reflejan los cambios netos y los efectos de diversificación; y estas distribuciones se pueden comparar entre carteras
- Los dos problemas más grandes al trabajar con distribuciones de pérdidas son que los estimadores se basan en datos pasados y que hay un problema práctico a la hora de estimar precisamente la distribución de pérdidas (por ejemplo, por usar suposiciones no factibles). No obstante, esto no indica que no se deban de usar estas medidas, sino que se deben complementar con mejores datos y con mejores métodos estadísticos
- En el enfoque basado en escenarios uno considera un número de posibles cambios futuros en un factor de riesgo, y el riesgo de la cartera se mide como la máxima pérdida de la cartera bajo todos los escenarios (aunque se puede ponderar). Fijando un vector $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ de cambios en los factores de riesgo, un vector $\mathbf{w} = (w_1, w_2, \dots, w_n)' \in [0,1]^n$ de ponderaciones y considerando una cartera de activos arriesgados, el riesgo de la cartera se mide de la siguiente forma:

$$\psi_{[X, \mathbf{w}]} \equiv \max\{w_1 l_{[t]}(x_1), w_2 l_{[t]}(x_2), \dots, w_n l_{[t]}(x_n)\}$$

- El valor en riesgo o *value-at-risk* (VaR) es probablemente la medida más usada por instituciones financieras y está incluida en el marco del Basel II, por lo que merece discutir su definición y sus usos de manera extensa
 - Considerando una cartera de activos arriesgados en un horizonte temporal fijo Δ y siendo $F_L(l) = P(L \leq l)$ la función de distribución correspondiente a la distribución de pérdidas, se quiere definir un estadístico basado en F_L que mida la severidad del riesgo de mantener la cartera durante el periodo Δ
 - No se distingue entre L y L^Δ o entre la distribución condicional e incondicional, dado que se asume que la elección se ha hecho fuera del análisis y F_L representa la distribución de interés
 - Un candidato obvio es la pérdida máxima posible, la cual se define como $\inf\{l \in \mathbb{R} : F_L(l) = 1\}$ (el menor valor l que hace que $F_L(l) = 1$, la cual sería la pérdida máxima). No obstante, en la mayoría de modelos el soporte de F_L no está acotado, por lo que la pérdida máxima es infinita, y no se tiene en cuenta la información de las probabilidades en F_L
 - El VaR es una extensión de la pérdida máxima que tiene en cuenta estas cosas, dado que se basa en la idea de reemplazar la pérdida máxima por la pérdida máxima que no es excedida por una alta probabilidad dada (nivel de confianza)
 - Dado un nivel de confianza $\alpha \in (0,1)$, el VaR de una cartera a un nivel de confianza α es el número l más pequeño tal que la probabilidad de que la pérdida L exceda l no es mayor a $1 - \alpha$

$$VaR_\alpha = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}$$



- Por lo tanto, el VaR no es más que un cuartil de la distribución de pérdida y se puede interpretar como la cantidad que como mínimo se pierde con una probabilidad de $1 - \alpha$
- Los valores típicos de α suelen ser $\alpha = 0.95$ o $\alpha = 0.99$ y el horizonte temporal suele ser 1 día o 10 días si se gestiona el riesgo de mercado, y un año si se gestiona riesgo operacional. En este caso, α no se interpreta como un nivel de significación, sino como un nivel de confianza
- Por definición, el VaR a un nivel de confianza α no da información sobre la severidad de las pérdidas que ocurren con probabilidad menor a $1 - \alpha$ (a la derecha del cuartil), lo cual es una desventaja al ser una medida de riesgo y no permite decir nada sobre esas posibles pérdidas
- Denotando μ como la media de la distribución de pérdidas, hay veces que se utiliza el estadístico $VaR_{\alpha}^{mean} \equiv VaR_{\alpha} - \mu$ para la adecuación del capital
 - Debido a que se le extrae la media de la distribución, el VaR_{α}^{mean} de diferentes distribuciones de pérdida se pueden comparar
 - Si el horizonte temporal Δ es un día, al VaR_{α}^{mean} se le puede llamar ganancias diarias en riesgo
 - La distinción entre VaR_{α}^{mean} y VaR_{α} suelen ser pequeñas cuando se gestiona riesgo de mercado, en donde el horizonte temporal es corto y μ es cercana a cero

- No obstante, es relevante en la gestión del riesgo crediticio porque el horizonte temporal es largo. Particularmente, se usa en la valoración de préstamos para determinar el capital económico necesario para soportar pérdidas inesperados en la cartera de préstamos
- Se puede obtener el valor del VaR_α a partir de la definición del inverso generalizado y la función cuartil

- Dado una función creciente $T: \mathbb{R} \rightarrow \mathbb{R}$, el inverso generalizado de T está definido por $T^\leftarrow(y) \equiv \inf\{x \in \mathbb{R} : T(x) \geq y\}$, usando la convención que la ínfima de un conjunto vacío es ∞
- Dada una función de distribución F , el inverso generalizado de F (F^\leftarrow) se denomina función cuartil de F . Para $\alpha \in (0,1)$, el α -cuartil de F es dado por la siguiente expresión:

$$q_\alpha(F) \equiv F^\leftarrow(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

- Si F es una función estrictamente creciente y continua, entonces $q_\alpha(F) \equiv F^{-1}(\alpha)$ (se cumple que $T(x) = y$). Para variables discretas, este no sería el caso (la función de distribución es discontinua) y no se cumpliría que $T(x) = y$
- Un punto $x_0 \in \mathbb{R}$ es un α -cuartil de F si, y solo si, se satisface que $F(x_0) \geq \alpha$ y $F(x) < \alpha$ para toda $x < x_0$
- Suponiendo que la distribución de pérdidas F_L es normal con media μ y varianza σ^2 y fijando una $\alpha \in (0,1)$, entonces se pueden obtener el VaR_α y el VaR_α^{mean} con las siguientes fórmulas:

$$VaR_\alpha = \mu + \sigma\Phi^{-1}(\alpha) \quad VaR_\alpha^{mean} = \sigma\Phi^{-1}(\alpha)$$

for $L \sim N(\mu, \sigma^2)$ and $\Phi = \text{std. normal dist.}$

- En este caso $\Phi^{-1}(\alpha)$ es la función α -cuartil $q_\alpha(F)$ de Φ , la cual se define de la siguiente manera:

$$\Phi^{-1}(\alpha) = \inf\{x \in \mathbb{R} : \Phi(x) \geq \alpha\}$$

- La demostración de estos resultados se basa en que F_L es una función estrictamente creciente, por lo que se tiene que dar que $F_L(VaR_\alpha) = \alpha$ y $F_L(VaR_\alpha^{mean}) = \alpha$ (gracias a que $F_L(x_0) \geq \alpha$ y $F_L(x) < \alpha$ para toda $x < x_0$ si x_0 es el α -cuartil) y se obtiene el siguiente resultado:

$$P(L \leq VaR_\alpha) = P\left(\frac{L - \mu}{\sigma} \leq \Phi^{-1}(\alpha)\right) = \Phi(\Phi^{-1}(\alpha)) = \alpha$$

$$\Rightarrow P(L \leq VaR_\alpha) = P(L \leq \mu + \sigma\Phi^{-1}(\alpha))$$

$$\Rightarrow VaR_\alpha = \mu + \sigma\Phi^{-1}(\alpha)$$

$$P(L \leq VaR_\alpha^{mean}) = P\left(\frac{L}{\sigma} \leq \Phi^{-1}(\alpha)\right) = \Phi(\Phi^{-1}(\alpha)) = \alpha$$

$$\Rightarrow P(L \leq VaR_\alpha^{mean}) = P(L \leq \sigma\Phi^{-1}(\alpha))$$

$$\Rightarrow VaR_\alpha^{mean} = \sigma\Phi^{-1}(\alpha)$$

- Se puede obtener un resultado similar para familias de escala-localización, de modo que si se supone que la distribución de pérdidas F_L es t-Student con media μ , parámetro σ y ν grados de libertad, fijando una $\alpha \in (0,1)$ se puede obtener el VaR_α con la siguiente fórmula:

$$VaR_\alpha = \mu + t_\nu^{-1}(\alpha) \sqrt{\sigma^2 \frac{\nu}{\nu-2}} \quad VaR_\alpha^{mean} = t_\nu^{-1}(\alpha) \sqrt{\sigma^2 \frac{\nu}{\nu-2}}$$

for $L \sim t(\nu, \mu, \sigma^2)$ and $t_\nu = \text{std. } t - \text{Student with } \nu \text{ df}$

- En este caso, el primer momento equivale al primer momento de la distribución normal, pero el segundo momento centrado no lo hace, de modo que se puede encontrar una expresión para este en función de σ y viceversa (que sirve para estimar con la varianza muestral):

$$E(L) = \mu$$

$$Var(L) = \frac{\nu\sigma^2}{\nu-2} \text{ for } \nu > 2 \Rightarrow \sigma^2 = \frac{\nu-2}{\nu} Var(L) \text{ for } \nu > 2$$

- Además, también se puede obtener un resultado para la distribución log-normal
 - Para una distribución log-normal con media μ y con varianza σ^2 , fijando una $\alpha \in (0,1)$ se puede obtener el VaR_α con la siguiente fórmula:

$$VaR_\alpha = \exp(\mu + \sigma\Phi^{-1}(\alpha))$$

$$VaR_\alpha^{mean} = \exp(\mu + \sigma\Phi^{-1}(\alpha)) - \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

for $\ln L \sim N(\mu, \sigma^2)$ and $\Phi = \text{std. normal dist.}$

- El VaR tiene problemas relacionados con la selección de parámetros, el riesgo de modelo y la liquidez del mercado
 - Al trabajar con el VaR, los parámetros α y Δt se tienen que escoger, pero al no haber valores óptimos únicos, hay consideraciones que pueden influenciar la selección de los reguladores o de los constructores del modelo
 - El horizonte Δt tiene que reflejar el tiempo que se piensa mantener la cartera (relacionado con restricciones legales y consideraciones de liquidez), y muchas veces no hay más remedio que escoger un horizonte temporal apropiado para el mercado en donde se desarrollan su actividad principal
 - Hay consideraciones prácticas para escoger un horizonte temporal corto: la suposición de que la composición de la cartera se mantiene constante solo es válida para horizontes temporales cortos y la calibración y comprobación de modelos estadísticos para cambios históricos de los factores de riesgo son más fáciles si Δt (suele significar que se tienen más datos disponibles)
 - Para el nivel de confianza α tampoco hay un único valor óptimo: para banca se suele utilizar 95% pero el Basel recomienda el uso del 99% para la adecuación del capital y el *Solvency II* utiliza un 99.5%. No obstante, el *backtesting* de estos modelos se suele llevar a cabo con niveles de confianza más bajos y un horizonte temporal más corto
 - En la práctica, se les da una interpretación literal a los números del VaR, de modo que si $VaR_\alpha = l$, entonces hay una probabilidad de $1 - \alpha$ que la pérdida en la posición sea mayor a l
 - No obstante, esta interpretación puede ser confusa porque ignora el error de estimación, el riesgo del modelo y el riesgo de liquidez del mercado
 - El riesgo del modelo es el riesgo de que el modelo que se usa para la distribución de pérdidas esté mal especificado, el cual es más pronunciado para niveles de confianza muy elevados. En cambio, el riesgo de liquidez se refiere a que cualquier intento de liquidar una posición grande puede mover el precio de manera no conveniente, incrementando las pérdidas

- Las medidas de riesgo basadas en cuartiles se usan en muchos cálculos de capital en riesgo en la práctica, tales como en el cálculo de capital regulatorio o para el requerimiento del capital de solvencia establecido por el *Solvency II*

- El VaR es una medida de riesgo que se aplica para calcular el número de cargos por capital regulatorio para la cartera de negociación o *trading book* de un banco

- Bajo el enfoque de modelaje interno, un banco calcula el VaR diario para la distribución de pérdidas posibles para un horizonte temporal de diez días, basándose en datos recientes sobre los cambios en los factores de riesgo suponiendo que la composición de la cartera se mantiene fija en todo el periodo
- Aunque los detalles exactos pueden variar dependiendo de la regulación nacional, el cargo de capital en el día t normalmente se calcula acorde a la siguiente fórmula:

$$RC^t = \max \left\{ VaR_{0.99}^{t,10}, \frac{k}{60} \sum_{i=1}^{60} VaR_{0.99}^{t-i+1,10} \right\} \text{ for } k \in [3,4]$$

where $VaR_{0.99}^{j,10} = VaR$ for day j with $\alpha = 0.99$ & $\Delta = 10$ days

- La media de los valores VaR de los últimos 60 días tiende a suavizar los cambios en el cargo por capital regulatorio en el tiempo a no ser que el número más reciente $VaR_{0.99}^{t,10}$ sea muy grande, por lo que se escogería este
- Se pueden aplicar un número de cargos de capital adicionales al RC^t , tales como un cargo por VaR bajo estrés financiero o *stressed VaR*, un cargo por riesgo incremental y otros diseñados para tener en cuenta riesgos específicos debido a movimientos de precio idiosincráticos no explicados por factores de riesgo de mercado generales
 - El cargo por *stressed VaR* se calcula de manera similar al RC^t pero se toman datos de una ventana histórica en la que los mercados eran particularmente volátiles
 - El cargo por riesgo incremental es una estimación del cuartil al 99.9% de la distribución de pérdidas anuales debido a impagos y rebajas de la calificación de instrumentos crediticios en el *trading book* (excepto titularizaciones)
- Una definición informal del requerimiento de capital de solvencia es que este es el nivel de capital que permite a una aseguradora cumplir

sus obligaciones en un horizonte temporal de un año con un nivel de confianza alto α

- Considerando el balance de una aseguradora y asumiendo que su capital viene dado por $V_t = A_t - B_t$ (la diferencia entre el valor de activos y de obligaciones), el asegurador quiere estar seguro de que será solvente durante un año con una probabilidad alta α
- Dado su balance actual y su modelo de negocio, se intenta determinar la cantidad mínima de capital extra x_0 que tendría que obtener en el momento actual t para invertirla en activos sin riesgo y así ser solvente en un año con probabilidad α . Formalmente, se intenta determinar lo siguiente:

$$x_0 = \inf \{x : P(V_{t+1} + x(1 + r_{t,1}) \geq 0) = \alpha\}$$

where $r_{t,1}$ = risk – free rate for 1 year investment at t

- Un cálculo simple permite ver como el capital se puede determinar el requerimiento de capital de solvencia $V_t + x_0$ (corregido para el monto de capital adicional x_0) a partir de la función de cuartil de la distribución de $(V_t - V_{t+1})/(1 + r_{t,i})$ (esta cantidad representa una distribución de pérdida que tiene en cuenta el valor temporal del dinero a través del descuento):

$$\begin{aligned} x_0 &= \inf \{x : P(V_{t+1} + x(1 + r_{t,1}) \geq 0) = \alpha\} = \\ &= \inf \{x : P(-V_{t+1} \leq x(1 + r_{t,1})) = \alpha\} = \\ &= \inf \left\{ x : P\left(\frac{V_t - V_{t+1}}{(1 + r_{t,1})} \leq x + V_t\right) = \alpha \right\} \\ &\Rightarrow V_t + x_0 = q_\alpha \left(\frac{V_t - V_{t+1}}{1 + r_{t,1}} \right) \end{aligned}$$

- Si x_0 es negativo, entonces la compañía está bien capitalizada en el momento t y el dinero se podría sacar de la compañía. Para este tipo de compañías, la cuantía $-x_0 = V_t - q_\alpha \left(\frac{V_t - V_{t+1}}{1 + r_{t,1}} \right)$ (los fondos propios menos el requerimiento de solvencia de capital) se denomina exceso de capital
- A parte del VaR, existen otras medidas relacionadas que se basan en el análisis de la cola de la distribución, tales como el VaR de cola, el déficit esperado, u otros

- Los momentos parciales son medidas de riesgo basadas en la parte inferior y superior de la distribución

- La mayoría de la literatura sobre gestión del riesgo se preocupa del riesgo inherente en la cola inferior de la distribución de pérdidas y ganancias, por lo que se suelen usar los momentos parciales inferiores para medir este riesgo
- Al utilizar la distribución de pérdida, uno se preocupa por los momentos parciales superiores de la distribución. Dado un exponente $k \geq 0$ y un punto de referencia q , el momento parcial superior $UPM(k, q)$ se define de la siguiente manera:

$$UPM(k, q) = \int_q^{\infty} (l - q)^k dF_L(l) \in [0, \infty)$$

- Algunas combinaciones de k y q tienen una interpretación especial: para $k = 0$ se obtiene $P(L \geq q)$; para $k = 1$ se obtiene $E[(L - q)I_{\{L \geq q\}}]$ y para $k = 2$ y $q = E(L)$ se obtiene la semivarianza superior de L . Cuanto más grande sea el valor escogido de k , más conservadora será la medida de riesgo porque se da más peso a desviaciones grandes con respecto al punto de referencia q
- El déficit esperado o *expected shortfall* (ES) es una medida de riesgo muy relacionada al VaR (por lo que también se puede llamar *tail VaR*). Para una pérdida L con $E(|L|) < \infty$ y una función de distribución F_L , el ES en un nivel de confianza $\alpha \in (0, 1)$ se define de la siguiente manera:

$$ES_{\alpha} = \frac{1}{1 - \alpha} \int_{\alpha}^1 q_u(F_L) du = \frac{1}{1 - \alpha} \int_{\alpha}^1 VaR_u(L) du$$

- La condición de que $E(|L|) < \infty$ asegura que la integral está bien definida y que existe. Además, se divide entre $1 - \alpha$ porque esta es la probabilidad que hay a la derecha del cuartil $q_{\alpha}(F_L)$, de modo que la media se tiene que compensar

$$E(L) = \int_0^1 q_u(F_L) du \Rightarrow ES_{\alpha}(L) = \frac{1}{P(L > q_{\alpha}(F_L))} \int_{\alpha}^1 q_u(F_L) du$$

- En vez de fijar un nivel de confianza particular, se hace una media de los valores de VaR para todos los niveles $u \geq \alpha$ y así tener más información sobre la cola de la distribución de pérdida. El ES_{α} solo depende de la distribución de L y $ES_{\alpha} \geq VaR_{\alpha}$ (por definición)

- Como el ES_α se puede entender como la media de todas las pérdidas mayores o iguales al VaR_α , es sensible a la severidad de las pérdidas que exceden el VaR_α , lo cual es una ventaja frente a esta misma medida. Cuanto más grande son las colas, mayores son estas medidas
- Derivando resultados asintóticos para la *ratio* ES_α/VaR_α cuando $\alpha \rightarrow 1$, se puede ver como, para distribuciones que no sean normales y tengan más densidad en las colas, la diferencia entre el ES_α y el VaR_α es más pronunciada (y el ES_α es siempre más alto):

$$\lim_{\alpha \rightarrow 1} \frac{ES_\alpha}{VaR_\alpha} = 1 \quad \text{for } L \sim N(\mu, \sigma^2)$$

$$\lim_{\alpha \rightarrow 1} \frac{ES_\alpha}{VaR_\alpha} = \frac{v}{v-1} = \frac{1}{1 - \frac{1}{v}} > 1 \quad \text{for } L \sim t(v, \mu, \sigma^2)$$

- Si la distribución de pérdida es continua en el VaR_α , se puede derivar una expresión aún más intuitiva. Para una pérdida integrable L con una función de distribución continua F_L y para cualquier $\alpha \in (0,1)$, se cumple la siguiente igualdad:

$$ES_\alpha = \frac{E(L I_{\{L \geq q_\alpha(L)\}})}{1 - \alpha} = \frac{E(L I_{\{L \geq q_\alpha(L)\}})}{P(L \geq q_\alpha(L))} = E(L | L \geq VaR_\alpha) = CTE_\alpha$$

$$\text{where } I_{\{L \geq q_\alpha(L)\}} = \begin{cases} 1 & \text{if } L \geq q_\alpha(L) \\ 0 & \text{if } L < q_\alpha(L) \end{cases}$$

- A la esperanza $E(L | L \geq VaR_\alpha)$ se le suele llamar *conditional tail expectation* (CTE), y la igualdad obtenida expresa que el ES_α es equivalente al CTE_α para distribuciones de pérdida continuas
- Denotando una variable aleatoria uniforme por $U \sim U(0,1)$, se sabe que la variable aleatoria $F_L^\leftarrow(U)$ tiene una función de distribución $F_L = q(F)$, por lo que se puede demostrar que $E(L | L \geq q_\alpha(F)) = \int_\alpha^\infty F_L^\leftarrow(u) du$ mediante las siguientes equivalencias:

$$E(L I_{\{L \geq q_\alpha(L)\}}) = E(F_L^\leftarrow(U) I_{\{F_L^\leftarrow(U) \geq F_L^\leftarrow(\alpha)\}}) = E(F_L^\leftarrow(U) I_{\{U \geq \alpha\}})$$

as F_L^\leftarrow is strictly increasing

$$\Rightarrow \frac{E(F_L^\leftarrow(U) I_{\{U \geq \alpha\}})}{P(U \geq \alpha)} = \frac{E(F_L^\leftarrow(U) I_{\{U \geq \alpha\}})}{1 - \alpha} = E(F_L^\leftarrow(U) | U \geq \alpha)$$

- A partir del lema anterior se pueden derivar las expresiones del ES_α en el caso de una distribución normal y de una distribución t-Student

- En el caso de una distribución normal con media μ y varianza σ^2 , el ES_α se puede expresar de la siguiente manera:

$$ES_\alpha = \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha}$$

- Esto se puede demostrar a partir de los resultados anteriores sobre el VaR de una distribución de las mismas características:

$$ES_\alpha = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_u(L) du$$

$$\Rightarrow ES_\alpha = \mu + \sigma E\left(\frac{L - \mu}{\sigma} \middle| \frac{L - \mu}{\sigma} \geq q_\alpha\left(\frac{L - \mu}{\sigma}\right)\right) = \mu + \sigma ES_\alpha\left(\frac{L - \mu}{\sigma}\right)$$

$$ES_\alpha\left(\frac{L - \mu}{\sigma}\right) = \frac{1}{1 - \alpha} \int_{\Phi^{-1}(\alpha)}^\alpha l \phi(l) dl = \frac{1}{1 - \alpha} [-\phi(l)]_{\Phi^{-1}(\alpha)}^\infty = \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha}$$

$$\Rightarrow ES_\alpha = \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha}$$

- En el caso de una distribución t-Student con media μ , parámetro σ y ν grados de libertad (y función de densidad g_ν), el ES_α se puede expresar de la siguiente manera:

$$ES_\alpha(L) = \mu + \sigma \frac{g_\nu(t_\nu^{-1}(\alpha))}{1 - \alpha} \left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu - 1} \right)$$

- En el caso de una distribución log-normal con media μ y varianza σ^2 , el ES_α se puede expresar de la siguiente manera:

$$ES_\alpha(L) = \frac{\Phi(\sigma - \Phi^{-1}(\alpha))}{1 - \alpha} \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

- En algunos libros se hace una distinción entre el ES_α y lo que se denomina VaR_α de cola o *tail VaR*, el $CVaR_\alpha$ o VaR_α condicional y el \overline{ES}_α (denotado así para evitar confusiones con la definición previa)

- Dado un nivel de confianza $\alpha \in (0,1)$ y una variable aleatoria L , se define el \overline{ES}_α de la siguiente manera:

$$\overline{ES}_\alpha = E\left[(L - VaR_\alpha(L))_+\right]$$

- Dado un nivel de confianza $\alpha \in (0,1)$ y una variable aleatoria L , se define el $TVaR_\alpha$ de la siguiente manera:

$$TVaR_\alpha = ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 VaR_u(L) du$$

- Dado un nivel de confianza $\alpha \in (0,1)$ y una variable aleatoria L , se define el $CVaR_\alpha$ de la siguiente manera:

$$CVaR_\alpha = E[L - VaR_\alpha | L > VaR_\alpha(L)]$$

- A partir de estas definiciones alternativas, es posible definir las siguientes relaciones entre medidas del riesgo:

$$CTE_\alpha = VaR_\alpha + \frac{1}{1-F(VaR_\alpha)} \overline{ES}_\alpha(L)$$

$$CVaR_\alpha = CTE_\alpha - VaR_\alpha = \frac{\overline{ES}_\alpha(L)}{1-F(VaR_\alpha)}$$

$$TVaR_\alpha = VaR_\alpha + \frac{1}{1-\alpha} \overline{ES}_\alpha(L) = CTE_\alpha + \left(\frac{1}{1-\alpha} - \frac{1}{1-F(VaR_\alpha)} \right) \overline{ES}_\alpha(L)$$

- Otras métricas estadísticas de la distribución de pérdidas como los momentos estadísticos son usadas frecuentemente como medidas de riesgo en la gestión del riesgo en instituciones financieras y aseguradoras
 - La esperanza de la distribución es una medida natural del riesgo ya que permite ver cuáles serán las pérdidas de media en la cartera

$$E(L) = \int_{-\infty}^{\infty} l P(L \leq l) dl$$

- No obstante, esta medida no permite analizar nada sobre la variabilidad en la distribución
- Históricamente, la varianza de la distribución de pérdidas y ganancias (o la desviación estándar) ha sido la medida de riesgo dominante en las finanzas, mayoritariamente debido a la teoría de cartera moderna de Markowitz
 - La varianza es un concepto bien entendido y es fácil de utilizar analíticamente, pero esta medida tiene dos problemas
 - Para trabajar con esta medida se tiene que asumir que el segundo momento de la distribución de pérdida existe, lo cual

no es problemático para la mayoría de distribuciones de rendimientos, pero lo puede ser para algunas áreas de seguros no-vida o de pérdidas operacionales

- En el lado conceptual, como no hay distinción entre desviaciones positivas y negativas de la media, es una buena medida de riesgo para distribuciones aproximadamente simétricas con respecto a la media, pero cuando las distribuciones están muy sesgadas, esta medida no es buena
- El tercer y el cuarto momento de la diferencia entre la variable aleatoria y su media también son unas medidas de riesgo que permiten analizar las colas
 - El tercer momento estadístico, la asimetría, permite medir qué tan asimétrica es la distribución. En relación a la distribución de pérdidas, normalmente la asimetría es positiva, por lo que la cola derecha de la distribución tiene más probabilidad acumulada y eso hace que pérdidas muy grandes sean más probables

$$Assimetry = \frac{E \left[(L_{t+1} - E(L_{t+1}))^3 \right]}{(\sqrt{Var(L_{t+1})})^3}$$

- El cuarto momento estadístico, la curtosis, permite medir qué tan extremos son los valores atípicos de la distribución (el grosor de las colas). Si la curtosis es positiva, los datos muestran valores atípicos más extremos que la distribución normal (colas gruesas), mientras que una negativa indica valores atípicos menos extremos (colas cortas)

$$Kurtosis = \frac{E \left[(L_{t+1} - E(L_{t+1}))^4 \right]}{Var(L_{t+1})^2}$$

- Es deseable que las medidas de riesgo sean buenas, en el sentido que cumplen una lista de propiedades, las cuales se basan en el principio de coherencia (propuesto por Artzner) y de convexidad (propuesto por Föllmer y Schied)
 - Suponiendo que las medidas de riesgo pertenecen a un espacio lineal de variables aleatorias \mathcal{M} (asumiendo que incluyen constantes), hay dos posibles interpretaciones para los elementos de \mathcal{M} : se pueden considerar como los futuros valores netos de los activos en una cartera o posición (denotados por V) o pueden representar pérdidas, definidas por $L = -(V - V_0)$

- Por lo tanto, hay dos nociones posibles de medidas de riesgo en \mathcal{M} , ya que se puede ver la medida de riesgo como la cantidad de capital adicional que se necesita para una posición con valor neto futuro V con tal de que sea aceptable por el regulador o gestor (denotada por $\tilde{q}(V)$) o como la cantidad total de capital necesaria para respaldar una posición con una pérdida L (denotada por $\varrho(L)$)
- Ambas nociones de las medidas de riesgo están relacionadas, ya que el capital total es igual al capital disponible más el capital social. Por lo tanto, se relacionan por la siguiente ecuación:

$$\varrho(L) = V_0 + \tilde{q}(V)$$

- Aunque las nociones estén relacionadas, tienen un efecto en como los axiomas se presentan y por eso se prefiere el uso de una medida de riesgo $\varrho: L \rightarrow \varrho(L)$
- El axioma de monotonidad expresa que para $L_1, L_2 \in \mathcal{M}$ tal que $L_1 \leq L_2$ de manera casi segura, se tiene que $\varrho(L_1) \leq \varrho(L_2)$
 - Desde un punto de vista económico, el axioma dice que las posiciones con mayores pérdidas en cualquier estado del mundo requieren más capital riesgo, por lo que posiciones con $\varrho(L) \leq 0$ no requieren capital
- El axioma de invariación por translación expresa que para toda $L \in \mathcal{M}$ y cualquier $l \in \mathbb{R}$, se tiene que $\varrho(L + l) \leq \varrho(L) + l$
 - Este axioma significa que añadir o sustraer una cantidad determinística l a una posición con una pérdida L altera el requerimiento de capital por exactamente esa cuantía
 - En términos de la noción alternativa de medida de riesgo, se puede ver que $\tilde{q}(V + k) = \tilde{q}(V) - k$ para toda $k \in \mathbb{R}$ y por tanto $\tilde{q}(V + \tilde{q}(V)) = 0$ (no se necesita inyectar más dinero)
- El axioma de subaditividad expresa que para $L_1, L_2 \in \mathcal{M}$ se tiene que $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$
 - La subaditividad refleja la idea de que el riesgo se puede reducir diversificando. Además, si el regulador utiliza una medida no subaditiva para determinar el capital regulatorio de una institución financiera, esta tiene un incentivo legal de dividirse en varias subsidiarias con tal de reducir sus requerimientos de capital

- La subaditividad hace que la descentralización de los sistemas de gestión del riesgo sea posible, dado que, si se usa una medida subaditiva ϱ y se requiere que $L = L_1 + L_2 \leq M$, solo es necesario escoger dos cotas M_1 y M_2 tal que $\varrho(L_i) \leq M_i$ y así hacer que $\varrho(L) = \varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2) \leq M_1 + M_2 \leq M$
- El axioma de homogeneidad positiva expresa que para toda $L \in \mathcal{M}$ y cualquier $\lambda > 0$, se tiene que $\varrho(\lambda L) \leq \lambda \varrho(L)$
 - Si se cumple la subaditividad, se puede ver que, para $n \in \mathbb{N}$, $\varrho(nL) = \varrho(\sum_{i=1}^n L_i) \leq n\varrho(L)$. Como no son pérdidas que derivan de diferentes instrumentos y no hay efectos de diversificación, esta desigualdad tiene que ser una igualdad
 - La subaditividad junto a la homogeneidad positiva implican que la medida de riesgo ϱ es convexa en \mathcal{M}
- Una medida de riesgo ϱ cuyo dominio incluye el cono convexo \mathcal{M} es llamada coherente (en \mathcal{M}) si satisface el axioma de monotonicidad, de invariación por translación, de subaditividad y de homogeneidad positiva
 - El axioma de homogeneidad positiva es criticado porque se ha sugerido que para valores grandes de λ , $\varrho(\lambda L) > \lambda \varrho(L)$ para penalizar por la concentración de riesgo y el riesgo de liquidez de las posiciones grandes. No obstante, esto es imposible con una medida de riesgo subaditiva
 - Esto ha llevado al estudio de una clase más grande de medidas de riesgo, llamadas medidas de riesgo convexas. En esta clase, las condiciones de homogeneidad positiva y subaditividad se relajan y solo se requiere una propiedad de convexidad
- Una medida de riesgo ϱ en \mathcal{M} se denomina medida de riesgo convexa en \mathcal{M} si satisface el axioma de monotonicidad, el de invarianza por translación y el de convexidad
 - El axioma de convexidad expresa que para toda $L_1, L_2 \in \mathcal{M}$ y toda $\lambda \in [0,1]$ se tiene que $\varrho(\lambda L_1 + (1 - \lambda)L_2) \leq \lambda \varrho(L_1) + (1 - \lambda)\varrho(L_2)$. La justificación de la convexidad es la idea de los efectos de la diversificación para reducir el riesgo
 - Una medida de riesgo coherente es siempre convexa, pero lo converso no es verdad, dado que dentro de las medidas de riesgo convexas es posible encontrar medidas que penalizan la concentración de riesgo (de modo que $\varrho(\lambda L) \geq \varrho(L)$ para toda $\lambda > 1$)

- Para medidas del riesgo positiva, la subaditividad y la convexidad son axiomas equivalentes

La distribución normal multivariante y contrastes de normalidad

- Un modelo estocástico para un vector o variable aleatoria puede ser entendido como un modo de describir el comportamiento de un vector de componentes a través de las probabilidades, la dependencia y su estructura de correlación. Una de

- Siendo $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)'$ un vector de variables aleatorias estándares independientes (de modo que $Z \sim N(0,1)$), A una matriz $d \times k$ con entradas reales y $\boldsymbol{\mu} \in \mathbb{R}^d$, un vector de variables aleatorias $\mathbf{X} = (X_1, X_2, \dots, X_d)'$ sigue una distribución normal multivariante si se cumple la siguiente relación:

$$\mathbf{X} \sim \boldsymbol{\mu} + A\mathbf{Z}$$

- Debido a las propiedades de la esperanza y de la covarianza aplicada a vectores, se puede ver como el valor esperado es $E(\mathbf{X}) = \boldsymbol{\mu}$ y la matriz de varianzas y covarianzas es $Var(\mathbf{X}) = \Sigma = AA' \geq 0$ (es semidefinida positiva)

$$E(\mathbf{X}) = E(\boldsymbol{\mu}) + AE(Z) = \boldsymbol{\mu}$$

$$Cov(\mathbf{X}) = E[(A\mathbf{Z})(A\mathbf{Z})'] = AE(\mathbf{Z}\mathbf{Z}')A' = AA' = \Sigma$$

- La función característica de la distribución normal multivariante muestra como la distribución se caracteriza por el vector de medias $\boldsymbol{\mu}$ y la matriz de varianzas y covarianzas, por lo que se utiliza la notación $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$

$$\phi_X(\mathbf{t}) = E(e^{it'X}) = \exp\left(it'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right) \text{ for } \mathbf{t} \in \mathbb{R}^d$$

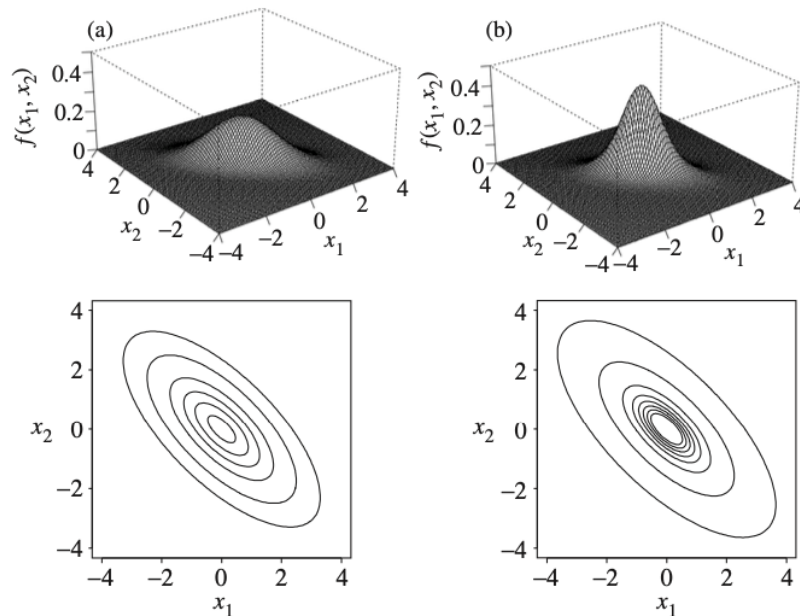
- Los componentes de \mathbf{X} son mutuamente independientes si, y solo si, Σ es diagonal
- La definición es esencialmente una metodología de simulación para distribuciones normales multivariantes. Si se quisiera generar un vector \mathbf{X} con distribución $N_d(\boldsymbol{\mu}, \Sigma)$, donde Σ es una matriz definida positiva, se utilizaría el siguiente algoritmo:
 - Primero, se tiene que realizar una descomposición de Cholesky de Σ para obtener el factor $\Sigma^{1/2} = A$. La descomposición se basa en dividir la matriz Σ en $\Sigma = AA'$, donde A es una matriz

triangular inferior con elementos positivos en su diagonal y A es su transpuesta (dado que Σ tiene entradas reales)

- Después, se genera un vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)'$ de variables normales estándar independientes
 - Finalmente, se fija $\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{Z} = \boldsymbol{\mu} + \mathbf{AZ}$ y se obtiene el vector
- Cuando el rango de la matriz de varianzas y covarianzas es completo (por lo que $\text{rank}(\Sigma) = d \leq k$), esta matriz es invertible, positiva definida y \mathbf{X} sigue una función de distribución continua con la siguiente función de densidad:

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \text{ for } \mathbf{x} \in \mathbb{R}^d$$

- La forma de la función de densidad claramente indica que los puntos con la misma densidad están en los elipsoides determinados por las ecuaciones de la forma $(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c$ para $c > 0$. En dos dimensiones, estos elipsoides serían elipses y su gráfico sería como los siguientes:



- Si se toman combinaciones lineales de vectores aleatorios que siguen una distribución normal multivariante, entonces estas también siguen una distribución normal multivariante. Siendo $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, B una matriz cualquiera $k \times d$ con entradas reales y \mathbf{b} es un vector real $k \times 1$, se obtiene el siguiente resultado:

$$B\mathbf{X} + \mathbf{b} \sim N_k(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B')$$

- Este resultado se puede demostrar a través de la función característica de la variable aleatoria
- Como un caso especial, si \mathbf{a} es un vector real $d \times 1$, entonces se obtiene el siguiente resultado:

$$\mathbf{a}'\mathbf{X} \sim N_k(\mathbf{a}'\boldsymbol{\mu} + \mathbf{b}, \mathbf{a}'\Sigma\mathbf{a})$$

- Usando las funciones características, se puede demostrar que \mathbf{X} sigue una distribución normal multivariante si, y solo si, $\mathbf{a}'\mathbf{X}$ sigue una distribución normal univariante para todo vector $\mathbf{a} \in \mathbb{R}^d - \{\mathbf{0}\}$
- Está claro que las distribuciones marginales de las variables aleatorias del vector aleatorio son normales univariantes

- De manera más general, usando $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$, un vector columna $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$ y una matriz Σ cuadrada, la distribución de la combinación lineal implica que las distribuciones marginales de \mathbf{X}_1 y \mathbf{X}_2 son también normales multivariantes y están dadas por $\mathbf{X}_1 \sim N_k(\boldsymbol{\mu}_1, \Sigma_{11})$ y $\mathbf{X}_2 \sim N_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22})$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

- Asumiendo que Σ es una matriz semidefinida positiva, las distribuciones condicionales de \mathbf{X}_2 dada \mathbf{X}_1 y la de \mathbf{X}_1 dada \mathbf{X}_2 son distribuciones normales multivariantes

$$\mathbf{X}_2|\mathbf{X}_1 \sim N_{d-k}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1}) \quad \mathbf{X}_1|\mathbf{X}_2 \sim N_k(\boldsymbol{\mu}_{1.2}, \Sigma_{11.2})$$

- En este caso, la media condicional y la matriz de covarianza para cada una de las distribuciones se expresan de la siguiente manera:

$$\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \Sigma_{22}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \quad \boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}_1 + \Sigma_{11}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

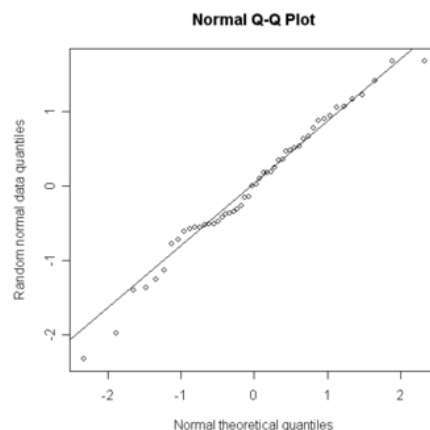
$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \quad \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

- Si $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ con una matriz Σ definida positiva, entonces la forma cuadrática $(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})$ sigue una distribución chi cuadrada con d grados de libertad

$$(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_d^2$$

- Esto se puede ver sabiendo que $\mathbf{Z} = \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim N_d(\mathbf{0}, I_d)$ y $(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}'\mathbf{Z} \sim \chi_d^2$

- Esta propiedad es importante a la hora de comprobar la normalidad multivariante
- Si \mathbf{X} e \mathbf{Y} son vectores aleatorios d -dimensionales independientes satisfaciendo $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ y $\mathbf{Y} \sim N_d(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$, entonces se da que la suma $\mathbf{X} + \mathbf{Y} \sim N_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma + \tilde{\Sigma})$
 - Para comprobarlo, se puede realizar el producto de las funciones características de cada uno de los vectores
- A partir de las nociones básicas de la distribución normal multivariante, es posible considerar el problema de comprobar si los datos $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ son observaciones de una distribución normal multivariante
 - Si $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ son variables normales independientes e idénticamente distribuidas, entonces para $1 \leq j \leq d$, la muestra univariante $X_{1,j}, X_{2,j}, \dots, X_{n,j}$ consistiendo en observaciones del componente j deben ser independientes e idénticamente distribuidas como una normal univariante
 - En general, cualquier muestra construida de una combinación lineal de datos de la forma $\mathbf{a}'\mathbf{X}_1, \mathbf{a}'\mathbf{X}_2, \dots, \mathbf{a}'\mathbf{X}_n$ tiene que tener variables independientes e idénticamente distribuidas como una normal univariante
 - Esto se puede ver a través de un gráfico Q-Q (cuartiles-cuartiles) contra una distribución normal estándar de referencia (en el eje horizontal). De este modo, se puede comprobar la normalidad si los cuartiles de los datos empíricos (eje vertical) coinciden con los de la distribución normal multivariante (en la línea recta)



- También se puede contrastar la normalidad a través de tests numéricos sobre normalidad
- Para comprobar normalidad multivariante no es suficiente que los márgenes univariantes de la distribución sean normales (que las variables se distribuyan bajo una normal univariante), por lo que se debe ser capaz de comprobar la normalidad multivariante y una manera de contrastar esto es explotar el hecho de que se cumple $(\mathbf{X} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_d^2$

- Suponiendo que se estima $\boldsymbol{\mu}$ y Σ usando los estimadores estándar, se construyen los siguientes datos:

$$\{D_i^2 = (\mathbf{X} - \bar{\mathbf{X}})' S^{-1} (\mathbf{X} - \bar{\mathbf{X}}) : i = 1, 2, \dots, n\}$$

- Debido a que las estimaciones del vector de medias y de la matriz de varianzas y covarianzas se usan en la construcción de D_i^2 , estos datos no son independientes, aunque las variables originales lo fueran. Además, la distribución marginal de D_i^2 bajo la hipótesis nula no es exactamente una chi cuadrada, si no que es una beta escalonada (aunque tiende asintóticamente a una chi cuadrada)

$$n(n-1)^{-2} D_i^2 \sim \text{Beta}\left(\frac{1}{2}d, \frac{1}{2}(n-d-1)\right)$$

- Por lo tanto, se espera que $D_1^2, D_2^2, \dots, D_n^2$ se comporten más o menos como variables independientes de una distribución χ_d^2 , por lo que se pueden construir gráficos Q-Q contra esta distribución (o contra la beta escalada)
- También es posible realizar tests numéricos de normalidad multivariante basados en medidas multivariantes de asimetría y curtosis, las cuales se definen de la siguiente manera:

$$b_d = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n D_{ij}^3 \quad k_d = \frac{1}{n} \sum_{j=1}^n D_i^4$$

- En este caso, D_i se define como antes y se conoce como la distancia de Mahalanobis entre \mathbf{X} y $\bar{\mathbf{X}}$, mientras que $D_{ij} = (\mathbf{X} - \bar{\mathbf{X}})' S^{-1} (\mathbf{X} - \bar{\mathbf{X}})$ se conoce como el ángulo de Mahalanobis entre $\mathbf{X}_i - \bar{\mathbf{X}}$ y $\mathbf{X}_j - \bar{\mathbf{X}}$
- El test de Mardia consiste en contrastar que, bajo la hipótesis nula de normalidad multivariante, los siguientes estadísticos siguen las siguientes distribuciones asintóticas:

$$\frac{1}{6}nb_d \sim \chi^2_{d(d+1)(d+2)/6} \quad \frac{k_d - d(d+2)}{\sqrt{8d(d+2)/n}} \sim N(0,1)$$

- El test de Jarque-Bera consiste en contrastar que, bajo la hipótesis nula, el siguiente estadístico sigue una distribución chi cuadrada con dos grados de libertad:

$$\sqrt{b} \equiv \frac{\left(\frac{1}{n}\right) \sum_{i=1}^n (X_{ij} - \bar{X}_j)^3}{\left(\frac{1}{n}\right) \left(\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2\right)^{\frac{3}{2}}} \quad k \equiv \frac{\left(\frac{1}{n}\right) \sum_{i=1}^n (X_{ij} - \bar{X}_j)^4}{\left(\frac{1}{n}\right) \left(\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2\right)^2}$$

$$\frac{1}{6}n \left(b + \frac{1}{4}(k-3)^2 \right) \sim \chi^2_2$$

Las distribuciones normales mixtas y distribuciones generalizadas hiperbólicas

- Es posible generalizar esta distribución normal multivariante para obtener distribuciones mixtas de normal multivariante. La idea crucial es la introducción de aleatoriedad en la primera matriz de varianzas y covarianzas y del vector de medias a través de una mezcla con una variable positiva W
 - El vector aleatorio \mathbf{X} tiene una distribución mixta de normal multivariante si este se define de la siguiente manera:

$$\mathbf{X} \sim \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$$

- En este caso, el vector \mathbf{Z} sigue una distribución normal multivariante estándar $N_k(\mathbf{0}, \mathbf{I}_k)$, $W \geq 0$ es una variable no negativa con valores escalares que es independiente del vector \mathbf{Z} , y \mathbf{A} es una matriz $d \times k$ con valores reales y $\boldsymbol{\mu}$ es el vector de medias constantes
- Estas distribuciones se conocen como mezclas de varianzas porque si se condiciona el vector aleatorio a W , se puede comprobar que $w \equiv \mathbf{X} | W$ sigue una distribución normal multivariante $N_d(\boldsymbol{\mu}, w\Sigma)$, donde $\Sigma = \mathbf{A}\mathbf{A}'$ y donde la varianza queda condicionada a w
- La distribución de \mathbf{X} se puede interpretar como una distribución compuesta construida tomando un conjunto de distribuciones normales multivariantes con el mismo vector de medias y con la misma matriz de varianzas y covarianzas multiplicada por la constante w

- El caso interesante es donde la matriz A tiene rango $\text{rank}(A) = d \leq k$ (rango completo) y Σ tiene rango completo y es positiva definida. Esto permite obtener una mezcla de varianza normal no singular
- La distribución mixta se construye obteniendo muestras aleatorias de este conjunto de distribuciones normales multivariantes acorde a un conjunto de ponderaciones determinadas por la distribución de W , por lo que la distribución resultante no es una distribución normal multivariante (si no se condiciona a un valor)
- En el contexto de modelaje de rendimientos de factores de riesgo, la distribución de mezcla W se puede interpretar como el *shock* que surge de nueva información y de impactos de volatilidad en todos los factores de riesgo
- Suponiendo que W tiene una expectación finita, se pueden obtener la esperanza y la matriz de varianzas y covarianzas para \mathbf{X}

$$E(\mathbf{X}) = E(\boldsymbol{\mu} + \sqrt{W}\mathbf{AZ}) = \boldsymbol{\mu} + E(\sqrt{W})A E(\mathbf{Z}) = \boldsymbol{\mu}$$

$$\begin{aligned} \text{Cov}(\mathbf{X}) &= E[(\sqrt{W}\mathbf{AZ})(\sqrt{W}\mathbf{AZ})'] = E[\sqrt{W}\mathbf{AZ}\mathbf{Z}'A'\sqrt{W}] = \\ &= E[\sqrt{W}\sqrt{W}\mathbf{AZ}\mathbf{Z}'A'] = E(W)A E(\mathbf{Z}\mathbf{Z}')A' = E(W)AA' = E(W)\Sigma \end{aligned}$$

- Normalmente $\boldsymbol{\mu}$ se denomina vector de localización y Σ se denomina matriz de dispersión de la distribución
- La matriz Σ es la matriz de varianzas y covarianzas de \mathbf{AZ} es la matriz de varianzas y covarianzas de \mathbf{X} si $E(W) = 1$, y $\boldsymbol{\mu}$ es solo el vector de medias cuando $E(W)$ está definido, lo cual requiere que $E(\sqrt{W}) < \infty$
- Estas distribuciones proporcionan un buen ejemplo de que la falta de correlación no necesariamente implica independencia de los componentes de \mathbf{X}
 - Suponiendo que (X_1, X_2) tienen una distribución mixta de normal multivariante con $A = I_{2 \times 2}$ y $E(W) < \infty$ tal que $\text{Cov}(X_1, X_2) = 0$, entonces X_1 y X_2 son independientes si, y solo si, W es casi seguro constante (las variables se distribuyen normalmente, por ejemplo)

- Si W es casi seguro una constante, entonces X_1 y X_2 siguen una distribución bivalente normal y son independientes. Conversamente, si X_1 y X_2 son independientes, se tiene que $E(|X_1||X_2|) = E(|X_1|)E(|X_2|)$

$$\begin{aligned} E(|X_1||X_2|) &= E(W|Z_1||Z_2|) = E(W)E(|X_1|)E(|X_2|) \geq \\ &\geq E(\sqrt{W})E(|Z_1|)E(\sqrt{W})E(|Z_2|) = E(|X_1|)E(|X_2|) \end{aligned}$$

- Solo es posible obtener una igualdad en el medio si W es una constante
- Usando la ecuación característica vista para las normales multivariantes, se puede calcular la función característica de una mezcla de varianza normal:

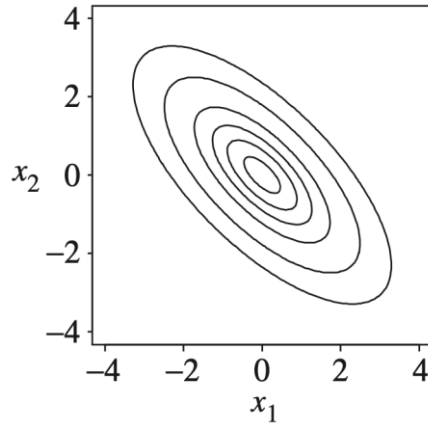
$$\phi(\mathbf{X}) = E[E(e^{it'\mathbf{X}}|W)] = E\left(e^{it'\boldsymbol{\mu} - \frac{1}{2}W\mathbf{t}'\Sigma\mathbf{t}}\right) = e^{it'\boldsymbol{\mu}} - \hat{H}\left(\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right)$$

$$\text{where } \hat{H}(x) = \int_0^\infty e^{-xv} dH(v)$$

- A partir de este resultado, la notación para denotar a una distribución mixta de normal multivariante es $\mathbf{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{H})$
- Asumiendo que Σ es definida positiva y que la distribución de W no tiene un punto de masa en cero, se puede derivar la función de densidad conjunta de esta distribución

$$\begin{aligned} f(\mathbf{x}) &= \int f_{\mathbf{X}|W}(\mathbf{x}|w) dH(w) = \\ &= \int \frac{w^{-\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})}{2w}\right] dH(w) \end{aligned}$$

- En este caos, $f_{\mathbf{X}|W}(\mathbf{x}|w)$ es la función de densidad conjunta gaussiana condicional a W y la función de densidad conjunta de \mathbf{x} se expresa en términos de la integral de Lebesgue-Stieltjes
- Cuando H tiene densidad h , esta función de densidad se convierte en la integral de Riemann de $\int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}|w)h(w) dw$. Todas estas densidades dependen de \mathbf{x} solo a través del término cuadrático $(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})$, de modo que son las densidades de las distribuciones elípticas



- Si $\mathbf{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{H})$ y $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$, donde $\mathbf{B} \in \mathbb{R}^{k \times d}$ y $\mathbf{b} \in \mathbb{R}^k$, entonces $\mathbf{Y} \sim M_k(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\Sigma\mathbf{B}', \hat{H})$

- La función característica mostrada anteriormente se puede utilizar para demostrar el siguiente resultado:

$$\phi(\mathbf{X}) = E(e^{it'(B\mathbf{X}+\mathbf{b})}) = e^{it'\mathbf{b}}\phi(B\mathbf{X}) = e^{it'(B\boldsymbol{\mu}+\mathbf{b})}\hat{H}\left(\frac{1}{2}\mathbf{t}'B\Sigma B'\mathbf{t}\right)$$

- Como se puede ver, la subclase de distribuciones de mezcla especificada por \hat{H} es cerrado bajo transformaciones lineales (no cambia)
- A partir de las distribuciones mixtas de normal multivariante, se pueden obtener diferentes distribuciones importantes al escoger una forma para la variable W
 - Si se toma W como una variable con una distribución gamma inversa $W \sim Ig\left(\frac{1}{2}\nu, \frac{1}{2}\nu\right)$ (que es equivalente a $\nu/W \sim \chi_\nu^2$), entonces \mathbf{X} tiene una distribución t multivariante con ν grados de libertad
 - La función de densidad para la variable W de mezcla es la siguiente:

$$f(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-(\alpha+1)} e^{-\frac{\beta}{w}} = \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu}{2w}}$$

$$\text{where } \alpha = \frac{\nu}{2} \text{ \& } \beta = \frac{\nu}{2}$$

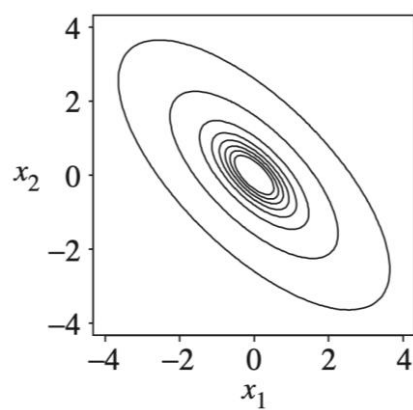
- La notación para esta distribución con ν grados de libertad es $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$ (sin que Σ sea la matriz de varianzas y covarianzas en esta definición). Debido a que $E(W) = \nu/(\nu -$

2) y a que $Cov(W) = [\nu/(\nu - 2)]\Sigma$, la matriz de varianzas y covarianzas de esta distribución solo está definida para $\nu > 2$

- La densidad de esta distribución se puede calcular a través de la función de densidad derivada anteriormente para la mixta de normal multivariante

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{1}{2}(\nu + d)\right)}{\Gamma\left(\frac{\nu}{2}\right) (\pi\nu)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \left[1 + \frac{(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{\nu} \right]^{-\frac{(\nu+d)}{2}}$$

- El locus de puntos con igual densidad es un elipsoide con la ecuación $(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c$ para alguna $c > 0$. Como se puede ver, a comparación de la normal multivariante, los niveles de densidad incrementan más rápido en el centro de la distribución y decaen más lentamente cuanto más se alejan del centro



- A comparación de la normal multivariante, esta distribución tiene las colas más anchas y, por tanto, una tendencia más pronunciada a generar valores extremos
- Todas las distribuciones anteriormente vistas tienen simetría elíptica, lo cual puede simplificar mucho el modelo para los datos de los rendimientos de los factores de riesgo reales, por lo que se pueden ver otros modelos para introducir simetría
 - La simetría elíptica implica que las distribuciones marginales unidimensionales son rígidamente simétricas, lo cual contradice la observación frecuente de los rendimientos que los rendimientos negativos tienen colas más anchas que los positivos
 - Los modelos que introducen asimetría a la clase de mezcla de normales lo hacen a través de mezclar las distribuciones normales con diferentes medias y diferentes varianzas

- El vector aleatorio \mathbf{X} tiene una distribución normal media-varianza multivariante si se define de la siguiente manera:

$$\mathbf{X} \sim \mathbf{m}(W) + \sqrt{W}\mathbf{A}\mathbf{Z}$$

- En este caso, el vector \mathbf{Z} sigue una distribución normal multivariante estándar $N_k(\mathbf{0}, \mathbf{I}_k)$, $W \geq 0$ es una variable no negativa con valores escalares que es independiente del vector \mathbf{Z} , \mathbf{A} es una matriz $d \times k$ con valores reales y $\mathbf{m}: [0, \infty) \rightarrow \mathbb{R}^d$ es una función medible
- En este caso, si se condiciona el vector aleatorio a W , se puede comprobar que $w \equiv \mathbf{X} | W$ sigue una distribución normal multivariante $N_d(\mathbf{m}(w), w\Sigma)$, donde $\Sigma = \mathbf{A}\mathbf{A}'$ y, debido a que ahora hay mezcla tanto en la media como en la varianza, estas se denominan distribuciones de mezcla de media-varianza de normales
- Una posible especificación concreta para la función $\mathbf{m}(W)$ suele ser $\mathbf{m}(W) = \boldsymbol{\mu} + W\boldsymbol{\gamma}$ cuando $\boldsymbol{\mu}$ y $\boldsymbol{\gamma}$ son vectores de parámetros en \mathbb{R}^d
 - El vector $\boldsymbol{\gamma}$ se puede interpretar como el parámetro de asimetría, de modo que si $\boldsymbol{\gamma} = \mathbf{0}$, la distribución es simétrica alrededor de $\boldsymbol{\mu}$, y si $\boldsymbol{\gamma} \neq \mathbf{0}$, entonces la distribución es asimétrica alrededor de $\boldsymbol{\mu}$
 - Debido a que $E(\mathbf{X}|W) = \boldsymbol{\mu} + W\boldsymbol{\gamma}$ y $Cov(\mathbf{X}|W) = W\Sigma$, se puede ver que la esperanza y la matriz de varianzas y covarianzas de \mathbf{X} vienen dadas por las siguientes expresiones cuando la variable de mezcla W tiene varianza finita:

$$E(\mathbf{X}) = E[E(\mathbf{X}|W)] = E(\boldsymbol{\mu} + W\boldsymbol{\gamma}) = \boldsymbol{\mu} + E(W)\boldsymbol{\gamma}$$

$$Cov(\mathbf{X}) = E[Cov(\mathbf{X}|W)] + Cov[E(\mathbf{X}|W)] =$$

$$= E(W)\Sigma + Cov[(\boldsymbol{\mu} + W\boldsymbol{\gamma})'(\boldsymbol{\mu} + W\boldsymbol{\gamma})] =$$

$$= E(W)\Sigma + Cov[\boldsymbol{\gamma}'W'W\boldsymbol{\gamma}] = E(W)\Sigma + Var(W)\boldsymbol{\gamma}\boldsymbol{\gamma}'$$

- Como se puede ver, los parámetros $\boldsymbol{\mu}$ y Σ no son, en general, el vector de medias y la matriz de varianzas y covarianzas, respectivamente, de \mathbf{X} (o un múltiplo). Esto solo es el caso si $\boldsymbol{\gamma} = \mathbf{0}$, de modo que la distribución es una normal de mezcla de varianza y las fórmulas anteriormente vistas aplican

- Una clase especial de distribuciones hiperbólicas generalizadas (GH) es la familia de mezclas de normal multivariante conocidas como las distribuciones hiperbólicas generalizadas simétricas

- Esta familia de distribuciones se obtiene tomando W como una variable aleatoria que sigue una distribución gaussiana inversa generalizada (GIG) $W \sim N^-(\lambda, \chi, \psi)$

- Los parámetros λ , χ y ψ determinan la forma de la distribución o el peso de las colas y del centro de la distribución. En general, mientras más grandes sean estos parámetros, más cerca estarán de la normal
- La función de densidad para la variable W de mezcla es la siguiente:

$$f(w) = \frac{\chi^{-\lambda}(\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})} w^{-\lambda-1} e^{-\frac{1}{2}(\chi w^{-1} + \psi w)} \quad \text{when } w > 0$$

- Los parámetros satisfacen las siguientes condiciones para los posibles valores de λ :

$$\lambda < 0 \Rightarrow \chi > 0 \text{ \& } \psi \geq 0$$

$$\lambda = 0 \Rightarrow \chi > 0 \text{ \& } \psi > 0$$

$$\lambda > 0 \Rightarrow \chi \geq 0 \text{ \& } \psi > 0$$

- Usando la densidad de la mezcla de normal multivariante, se puede obtener la función de densidad conjunta para esta familia de distribuciones:

$$f(\mathbf{x}) = \frac{(\sqrt{\chi\psi})^{-\lambda} \psi^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} K_\lambda(\sqrt{\chi\psi})} \frac{K_{\lambda - (\frac{d}{2})} \sqrt{[\chi + (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})] \psi}}{[\sqrt{[\chi + (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})] \psi}]^{\frac{d}{2} - \lambda}}$$

$K_\lambda = \text{Bessel function of third kind}$

- Esta distribución es un caso especial ($\boldsymbol{\gamma} = \mathbf{0}$) de una familia más general de distribuciones hiperbólicas generalizadas multivariante, la cual puede ser obtenida como una mezcla de media-varianza de normales, las cuales no son necesariamente distribuciones elípticas

$$f(\mathbf{x}) = c \frac{K_{\lambda-(d/2)} \sqrt{[\chi + (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})] (\psi + \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma})}}{[\sqrt{[\chi + (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})] (\psi + \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma})}]^{\frac{d}{2}-\lambda}}$$

$$\text{where } c \equiv \frac{(\sqrt{\chi\psi})^{-\lambda} \psi^{\lambda} (\psi + \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma})^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} K_{\lambda}(\sqrt{\chi\psi})}$$

- En este caso general, la distribución no es elíptica y tiene márgenes (distribuciones marginales) asimétricos
- Se suele adoptar la notación $\mathbf{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ para la GH
- El vector de medias y la matriz de varianzas y covarianzas se puede calcular fácilmente a través de las fórmulas anteriormente vistas para mezclas de media-varianza de normales utilizando la información sobre la distribución GIG y sus momentos

- Además, la función de la distribución GH se puede calcular a través de la función característica anteriormente vista

$$\phi(\mathbf{X}) = E(e^{it'\mathbf{X}}) = e^{it'\mathbf{b}} \phi(B\mathbf{X}) = e^{it'\boldsymbol{\mu}} \hat{H}\left(\frac{1}{2} \mathbf{t}' \Sigma \mathbf{t} - it' \boldsymbol{\gamma}\right)$$

- Si $\mathbf{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ y $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, donde $B \in \mathbb{R}^{k \times d}$ y $\mathbf{b} \in \mathbb{R}^k$, entonces $\mathbf{Y} \sim GH_d(\lambda, \chi, \psi, B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B', B\boldsymbol{\gamma})$
- Usando la función característica de esta distribución, se puede obtener el siguiente resultado:

$$\phi(\mathbf{Y}) = E(e^{it'(B\mathbf{X}+\mathbf{b})}) = e^{it'(B\boldsymbol{\mu}+\mathbf{b})} \hat{H}\left(\frac{1}{2} \mathbf{t}' B\Sigma B' \mathbf{t} - it' B\boldsymbol{\gamma}\right)$$

- Por lo tanto, la clase de distribuciones GH son cerrados bajo operaciones lineales y los parámetros heredados de la distribución de mezcla GIG no cambian bajo operaciones lineales
- Esto causa que los márgenes de \mathbf{X} sean fáciles de calcular, ya que estos se distribuirán como $\mathbf{X}_i \sim GH_1(\lambda, \chi, \psi, \boldsymbol{\mu}_i, \Sigma_{ii}, \boldsymbol{\gamma}_i)$
- La distribución de mezcla GIG es muy flexible y contiene varios tipos de distribuciones concretas, las cuales son casos especiales de la GH
- Si $\lambda > 0$ y $\chi = 0$, se obtiene una distribución gamma, y si $\lambda < 0$ y $\psi = 0$, se obtiene una distribución gamma inversa. En estos casos la función de densidad anterior se puede interpretar como

el límite cuando $\chi \rightarrow 0$ y $\psi \rightarrow 0$ (respectivamente), de modo que son casos límites especiales

- La distribución de mezcla gamma resulta en distribuciones de Laplace o los llamados modelos de varianza-gamma simétrica (VG), mientras que la mezcla de gamma inversa resulta en una distribución t multivariante con $\lambda = -\nu/2$ y $\chi = \nu$
- Si $\lambda = (d + 1)/2$, la distribución es una distribución hiperbólica multivariante (pero no generalizada) cuyas marginales univariantes son distribuciones hiperbólicas
- Si $\lambda = -1/2$, la distribución que se obtiene es la inversa gaussiana normal simétrica (NIG), mientras que si $\lambda = 1$ la distribución es una GH multivariante cuyas marginales univariantes son distribuciones hiperbólicas
- Hay varias parametrizaciones alternativas para la distribución GH en la literatura, y es muy común encontrar esta distribución en su forma reparametrizada. Las parametrizaciones más comunes son la parametrización $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, la $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$ y la $(\lambda, \alpha, \mu, \Delta, \delta, \beta)$

	$(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ -Parametrization					
	λ	χ	ψ	μ	Σ	γ
ghyp	$\lambda \in \mathbb{R}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
hyp	$\lambda = \frac{d+1}{2}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
NIG	$\lambda = -\frac{1}{2}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
t	$\lambda < 0$	$\chi > 0$	$\psi = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
VG	$\lambda > 0$	$\chi = 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
	$(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$ -Parametrization					
	λ	$\bar{\alpha}$	μ	Σ	γ	
ghyp	$\lambda \in \mathbb{R}$	$\bar{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$	
hyp	$\lambda = \frac{d+1}{2}$	$\bar{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$	
NIG	$\lambda = -\frac{1}{2}$	$\bar{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$	
t	$\lambda = -\frac{\nu}{2} < -1$	$\bar{\alpha} = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$	
VG	$\lambda > 0$	$\bar{\alpha} = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$	
	$(\lambda, \alpha, \mu, \Delta, \delta, \beta)$ -Parametrization					
	λ	α	δ	μ	Δ	β
ghyp	$\lambda \in \mathbb{R}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x' \Delta x > 0\}$
hyp	$\lambda = \frac{d+1}{2}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x' \Delta x > 0\}$
NIG	$\lambda = -\frac{1}{2}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x' \Delta x > 0\}$
t	$\lambda < 0$	$\alpha = \sqrt{\beta' \Delta \beta}$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \mathbb{R}^d$
VG	$\lambda > 0$	$\alpha > 0$	$\delta = 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x' \Delta x > 0\}$

- La parametrización $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ se obtiene de la distribución de mezcla de media-varianza de normal cuando $W \sim N^-(\lambda, \chi, \psi)$, la cual presenta problemas de identificación
 - Debido a que los vectores $X \sim GH_d(\lambda, \chi/k, k\psi, \mu, k\Sigma, k\gamma)$ y $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ son idénticas para cualquier $k > 0$, ocurre un problema de identificación cuando se intentan estimar los parámetros en la práctica

- Esto se puede resolver restringiendo el determinante de Σ a un valor particular cuando se ajusta el modelo (normalmente a 1). Esta restricción tiene un efecto en los valores de χ y ψ que se estiman, pero no tienen un efecto en el valor de su producto $\chi\psi$, el cual es un buen parámetro resumen de la distribución
- Para eliminar el grado de libertad de manera más elegante que fijando el determinante de Σ , se puede requerir que el valor esperado de la variable W que sigue una GIG sea 1, derivando a la parametrización $(\lambda, \bar{\alpha}, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$
 - Esto hace que la interpretación de $\boldsymbol{\gamma}$ sea más fácil y que, además, el proceso de ajuste sea más rápido
 - Como el valor esperado de W existe mientras $\psi > 0$, si $\psi \rightarrow 0$, la distribución GIG se aproxima a la distribución de gamma inversa, para el cual la esperanza solo existe si $\boldsymbol{\gamma} < -\mathbf{1}$. De este modo, si $E(W) = 1$, se puede obtener ψ y χ a partir de $\bar{\alpha}$ y K_λ

$$E(W) = \sqrt{\frac{\chi}{\psi} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}} = 1$$

$$\text{If } \bar{\alpha} \equiv \sqrt{\chi\psi} \Rightarrow \psi = \bar{\alpha} \frac{K_{\lambda+1}(\bar{\alpha})}{K_\lambda(\bar{\alpha})} \text{ \& } \chi = \frac{\bar{\alpha}^2}{\psi} = \bar{\alpha} \frac{K_\lambda(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})}$$

- La desventaja de esta parametrización es que no existe en el caso en que $\bar{\alpha} = 0$ y $\lambda \in [-1, 0]$, lo que corresponde a la distribución t -Student con varianza inexistente
- En la introducción de la distribución GH se utilizó la parametrización $(\lambda, \alpha, \boldsymbol{\mu}, \Delta, \delta, \boldsymbol{\beta})$ para el caso multivariante
 - Igual que en la parametrización $(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$, hay un problema de identificación que se puede resolver restringiendo el determinante de Δ a 1
 - La representación de la mezcla perteneciente a esta parametrización es la siguiente:

$$\mathbf{X}|W = w \sim N_d(\boldsymbol{\mu} + w\boldsymbol{\beta}\Delta, w\Delta)$$

$$\text{where } W \sim N^-(\lambda, \delta^2, \alpha^2 - \boldsymbol{\beta}'\Delta\boldsymbol{\beta})$$

Las técnicas de reducción de dimensionalidad

- Usando un modelo factorial, se intenta explicar la aleatoriedad de los componentes de un vector \mathbf{X} d -dimensional en términos de un conjunto más pequeño de factores comunes
 - El vector aleatorio \mathbf{X} sigue un modelo factorial de p -factores si este se puede descomponer de la siguiente manera:

$$\mathbf{X} = \mathbf{a} + B\mathbf{F} + \boldsymbol{\varepsilon}$$

- El vector $\mathbf{F} = (F_1, F_2, \dots, F_p)'$ es un vector aleatorio de factores comunes con $p < d$ y una matriz de varianzas y covarianzas que es definida positiva
- El vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d)'$ es un vector aleatorio de términos de error idiosincráticos, los cuales no están correlacionados y tienen una media nula. Además, la covarianza entre \mathbf{F} y $\boldsymbol{\varepsilon}$ es nula

$$E[(\varepsilon_i - E(\varepsilon_i))(\varepsilon_j - E(\varepsilon_j))'] = E(\varepsilon_i \varepsilon_j') = 0$$

$$Cov(\mathbf{F}, \boldsymbol{\varepsilon}) = E[(\mathbf{F} - E(\mathbf{F}))(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}))'] = E[(\mathbf{F} - E(\mathbf{F}))\boldsymbol{\varepsilon}'] = 0$$

- La matriz B es una matriz de tamaño $d \times p$ de coeficientes reales y el vector $\mathbf{a} = (a_1, a_2, \dots, a_d)'$ es un vector de constantes
- Las condiciones de correlación nula anteriores no expresan independencia, de modo que esta no es necesaria para el modelo
- A partir de las suposiciones básicas del modelo, es posible obtener una estructura específica para la matriz de varianzas y covarianzas Σ de \mathbf{X}
 - Denotando la matriz de varianzas y covarianzas de \mathbf{F} por Ω y la de $\boldsymbol{\varepsilon}$ por la matriz diagonal Υ , se puede obtener la siguiente matriz:

$$\Sigma = Cov(\mathbf{X}) = E[(\mathbf{a} + B\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{a} + B\mathbf{F} + \boldsymbol{\varepsilon})'] = B\Omega B' + \Upsilon$$

- Si el modelo de factores es cierto, los factores comunes siempre se pueden transformar de modo que se tenga media nula y que sean ortogonales (estandarización). Fijando los factores como $\mathbf{F}^* = \Omega^{1/2}[\mathbf{F} - E(\mathbf{F})]$ y los coeficientes como $B^* = \Omega^{1/2}B$, se puede obtener una representación del modelo estandarizado como $\mathbf{X} = \boldsymbol{\mu} + B^*\mathbf{F}^* + \boldsymbol{\varepsilon}$, donde $\boldsymbol{\mu} = E(\mathbf{X})$ y la matriz de varianzas y covarianzas viene dada por la siguiente expresión:

$$\Sigma = Cov(\mathbf{X}) = Cov(\boldsymbol{\mu} + \mathbf{B}^* \mathbf{F}^* + \boldsymbol{\varepsilon}) = \mathbf{B}^* \mathbf{B}^{*'} + \mathbf{Y}$$

$$as \ \Omega = I_{p \times p}$$

- De manera converso, se puede ver como cuando un vector aleatorio \mathbf{X} tiene una matriz de varianzas y covarianzas que satisface $\Sigma = \mathbf{B}\mathbf{B}' + \mathbf{Y}$ para alguna $\mathbf{B} \in \mathbb{R}^{d \times p}$ con rango $rank(\mathbf{B}) = p < d$ y matriz diagonal \mathbf{Y} , entonces \mathbf{X} tiene una representación como modelo factorial para un vector \mathbf{F} p -dimensional y un vector $\boldsymbol{\varepsilon}$ d -dimensional
- Para poder comprobar que existe una descomposición en términos similar a la anterior a partir de la esperanza de un vector aleatorio \mathbf{X} y de su matriz de varianzas y covarianzas, se puede utilizar el siguiente procedimiento:

- El primer paso consiste en hacer una descomposición de la matriz de varianzas y covarianzas a través del cálculo de sus vectores propios y de sus valores propios. Para poder obtener los valores y vectores propios, se resuelven las siguientes ecuaciones:

$$\det(\lambda \mathbf{I} - \mathbf{B}) = 0 \quad (\lambda \mathbf{I} - \mathbf{B})\mathbf{x} = 0$$

- Una vez obtenidos, la matriz \mathbf{B} estará formada de los primeros vectores propios y la matriz Ω será una matriz diagonal con los primeros valores propios

$$\mathbf{B} = \begin{bmatrix} \mathbf{x}'_1 \\ \dots \\ \mathbf{x}'_d \end{bmatrix} = \begin{bmatrix} x_1^1 & \dots & x_1^r \\ \vdots & \ddots & \vdots \\ x_d^1 & \dots & x_d^r \end{bmatrix} \quad \Omega = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_r \end{bmatrix}$$

$r = \text{first } r \text{ eigenvalues}$

- A partir de estas matrices, se puede obtener la matriz de varianzas y covarianzas de para los factores \mathbf{F}

$$\mathbf{B}\Omega^{1/2} = \begin{bmatrix} x_1^1 & \dots & x_1^r \\ \vdots & \ddots & \vdots \\ x_d^1 & \dots & x_d^r \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\lambda_r} \end{bmatrix} = \begin{bmatrix} x_1^1 \sqrt{\lambda_1} & \dots & x_1^r \sqrt{\lambda_r} \\ \vdots & \ddots & \vdots \\ x_d^1 \sqrt{\lambda_1} & \dots & x_d^r \sqrt{\lambda_r} \end{bmatrix}$$

$$\mathbf{B}\Omega\mathbf{B}' = \begin{bmatrix} x_1^1 \sqrt{\lambda_1} & \dots & x_1^r \sqrt{\lambda_r} \\ \vdots & \ddots & \vdots \\ x_d^1 \sqrt{\lambda_1} & \dots & x_d^r \sqrt{\lambda_r} \end{bmatrix} \begin{bmatrix} x_1^1 \sqrt{\lambda_1} & \dots & x_1^r \sqrt{\lambda_r} \\ \vdots & \ddots & \vdots \\ x_d^1 \sqrt{\lambda_1} & \dots & x_d^r \sqrt{\lambda_r} \end{bmatrix}$$

- Finalmente, se puede obtener la matriz Υ a partir de diferencia entre la matriz Σ y $B\Omega B'$, y la matriz de correlaciones de los errores R_ε

$$\Upsilon = \Sigma - B\Omega B'$$

$$R_\varepsilon = (\Delta Y)^{-1} \Upsilon (\Delta Y)^{-1} \quad \text{where } \Delta Y \equiv \text{diag}(Y)^{1/2}$$

- Uno de los usos más importantes de los modelos factoriales es el análisis de la matriz de correlaciones, ya permite analizar las relaciones entre diferentes activos a través de los factores de riesgo comunes
- Suponiendo que \mathbf{X} es un vector aleatorio con variables estandarizadas y una matriz de equicorrelación (la correlación entre cada par de componentes es igual a $\rho > 0$), entonces la matriz de varianzas y covarianzas se puede expresar de la siguiente manera:

$$\Sigma = \rho J_{d \times d} + (1 - \rho) I_{d \times d}$$

- En este caso, $J_{d \times d}$ es una matriz cuadrada de unos y $I_{d \times d}$ es la matriz identidad
 - Por lo tanto, la matriz Σ se puede descomponer como $\Sigma = BB' + \Upsilon$ para una matriz $B = \sqrt{\rho} \mathbf{1}_d$ de tamaño $d \times 1$
- Con tal de encontrar una descomposición factorial de \mathbf{X} , se puede utilizar cualquier variable estandarizada Y que sea independiente de \mathbf{X} y definir un solo factor común F (que tendrá varianza unitaria y esperanza nula por construcción) y errores ε de la siguiente manera:

$$F = \frac{\sqrt{\rho}}{1 - \rho(d-1)} \sum_{j=1}^d X_j + Y \sqrt{\frac{1 - \rho}{1 - \rho(d-1)}}$$

$$\varepsilon_j = X_j - \sqrt{\rho} F$$

- Por lo tanto, se tiene una descomposición factorial estandarizada $\mathbf{X} = BF + \varepsilon$, en donde $\text{Cov}(F, \varepsilon) = 0$ para cualquier j y $\text{Cov}(\varepsilon_j, \varepsilon_k) = 0$ cuando $j \neq k$, y un vector aleatorio \mathbf{X} con matriz equicorrelacionada se puede entender como un modelo factorial con un solo factor
- Como se puede tomar cualquier Y , los factores y los errores en la descomposición no son únicos. Considerando el caso en el que \mathbf{X} es un vector aleatorio con variables normales estándar y Y es una variable normal estándar, se puede obtener la siguiente representación:

- El factor común sigue una distribución normal estándar y el vector de errores sigue una distribución normal multivariante y se cumple $Cov(F, \varepsilon) = 0$ para cualquier j y $Cov(\varepsilon_j, \varepsilon_k) = 0$ cuando $j \neq k$
- En este caso, la varianza de F es ρ y la varianza de ε_j es $1 - \rho$, por lo que se puede escribir el modelo factorial para una matriz equicorrelacionada de la siguiente manera:

$$X_j = \sqrt{\rho}F + \sqrt{1 - \rho}Z_j \quad \text{for } j = 1, 2, \dots, d$$

$Z = \text{independent Gaussian variable}$

- Esto se utiliza mucho en modelos de carteras homogéneas de crédito
- Si se supone que el vector aleatorio \mathbf{X} representa el rendimiento de d compañías diferentes, se puede construir una variable $Z_{(d)} = (1/d) \sum_{j=1}^d X_j$ que represente el rendimiento de una cartera con la misma inversión para cada compañía. Usando el modelo factorial, se puede obtener la siguiente descomposición:

$$Z_{(d)} = (1/d) \sum_{j=1}^d X_j = \frac{1}{d} \mathbf{1}' B F + \frac{1}{d} \mathbf{1}' \varepsilon = \sqrt{\rho} F + \frac{1}{d} \sum_{j=1}^d \varepsilon_j$$

- El riesgo en el primer término de la ecuación no se ve afectado por el aumento de tamaño de la cartera d , pero el riesgo idiosincrático si se ve afectado, por lo que se podría reducir el riesgo idiosincrático aumentando el tamaño de la cartera (si se midiera el riesgo a partir de la varianza) y se demuestra que el factor sistemático es aquel que mayor riesgo aporta en una cartera grande

$$\begin{aligned} Var(Z_{(d)}) &= Var(\sqrt{\rho}F) + Var\left(\frac{1}{d} \sum_{j=1}^d \varepsilon_j\right) = \\ &= \rho + \frac{d}{d^2} (1 - \rho) = \rho + \frac{1 - \rho}{d} \end{aligned}$$

$$\Rightarrow \lim_{d \rightarrow \infty} Var(Z_{(d)}) = \lim_{d \rightarrow \infty} \rho + \frac{1 - \rho}{d} = \rho$$

- Asumiendo que se tienen datos $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\} \in \mathbb{R}^d$ sobre los rendimientos asociados con muchos riesgos, el modelo factorial supone que cada observación $\mathbf{X}_t = (\mathbf{X}_{t1}, \mathbf{X}_{t2}, \dots, \mathbf{X}_{td})'$ proviene de algunos factores $\mathbf{F}_t =$

$(F_{t1}, F_{t2}, \dots, F_{tp})'$ sumados a errores $\varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2}, \dots, \varepsilon_{tp})'$ y se requiere hacer una estimación del modelo

- En el caso en el que los factores sean observables, los vectores X_t y F_t son vectores medibles para $t = 1, 2, \dots, n$ y se puede hacer una estimación a través de técnicas de regresión univariantes:

- Esta técnica consiste en estimar d modelos de regresión con los p factores en cada una de las regresiones. Estas estimaciones se hacen a través de mínimos cuadrados

$$X_{tj} = a_j + b'_j F_t + \varepsilon_{tj} \text{ for } t = 1, 2, \dots, n$$

- En notación matricial, la obtención de los parámetros relevantes se facilita, dado que se incluye el término constante en los cálculos

$$X_j = \begin{pmatrix} X_{1j} \\ \vdots \\ X_{nj} \end{pmatrix} \quad b_{j2} = \begin{pmatrix} a_j \\ b'_j \end{pmatrix} \quad F = \begin{pmatrix} 1 & F'_1 \\ \vdots & \vdots \\ 1 & F'_d \end{pmatrix}$$

$$\hat{b}_{j2} = (F'F)^{-1}F'X_j \text{ for } \forall j = 1, 2, \dots, d$$

- Para obtener la matriz de varianzas y covarianzas se utilizan los estimadores econométricos más comunes. También es posible obtenerla a través de la matriz \hat{B} , conformada de tantas filas como factores haya y con tantas columnas como variables a estimar a través de los modelos
- Para poder medir el ajuste del modelo, se utilizan las métricas más comunes, como el coeficiente de determinación u otros
- Otra técnica para poder estimar el modelo en el caso en el que los factores sean observables es a través de una regresión multivariante
 - Este tipo de estimación funciona de manera equivalente a la anteriormente vista, aumentando las dimensiones de las matrices

$$X = \begin{pmatrix} X'_1 \\ \vdots \\ X'_n \end{pmatrix}_{n \times d} \quad B_2 = \begin{pmatrix} a' \\ B' \end{pmatrix}_{(p+1) \times 1}$$

$$F = \begin{pmatrix} 1 & F'_1 \\ \vdots & \vdots \\ 1 & F'_d \end{pmatrix}_{n \times (p+1)} \quad E = \begin{pmatrix} \varepsilon'_1 \\ \vdots \\ \varepsilon'_n \end{pmatrix}_{n \times d}$$

$$\hat{B}_2 = (F'F)^{-1}F'X$$

La teoría del valor extremo: máximos

- La teoría del valor extremo es una rama de la probabilidad que se centra en las leyes de límites para valores extremos en muestras grandes
 - La teoría contiene resultados importantes que describen el comportamiento de máximos y mínimos, estadísticos de orden alto y valores muestrales que exceden umbrales altos
 - El interés de esta teoría se centra en la aplicación de los resultados para desarrollar modelos de comportamiento extremo de factores de riesgo financiero
 - En particular, uno se centra en el desarrollo de modelos para la cola de la distribución de los cambios en los factores de riesgo
 - Hay dos tipos de modelos principales para valores extremos: los modelos de máximos de bloque y de excesos de umbral
 - Los modelos de máximos de bloque se centran en modelar las observaciones más grandes de muestras grandes con observaciones idénticamente distribuidas
 - Los modelos de excesos de umbral modelan observaciones que sobrepasan un cierto umbral, y son considerados más útiles para la práctica por su uso eficiente de las observaciones extremas
- Considerando una secuencia de variables aleatorias independientes e idénticamente distribuidas $(X_t) \in \mathbb{R}$ representando pérdidas financieras, se pueden establecer resultados de convergencia de sumas útiles para la generalización de la distribución de valores extremos (GEV)
 - El papel de la distribución GEV es análoga a la de la normal para el teorema del límite central
 - La teoría de los valores extremos se centra en la distribución de los máximos normalizados. El máximo de n variables aleatorias independientes e idénticamente distribuidas se denota por M_n , llamado máximo del n -bloque
 - Las únicas distribuciones asintóticas no degeneradas (distribuciones no concentradas en un solo punto) para un máximo normalizado cuando n tiende a infinito son las de la familia de distribuciones GEV

- La función de distribución de la distribución GEV estándar es la siguiente:

$$H_{\xi}(x) = \begin{cases} \exp\left(-(1 + \xi x)^{-\frac{1}{\xi}}\right) & \text{for } \xi \neq 0 \\ \exp(-e^{-x}) & \text{for } \xi = 0 \end{cases}$$

where $1 + \xi x > 0$

- Una familia triparamétrica es obtenida definiendo $H_{\xi,\mu,\sigma}(x) \equiv H_{\xi,\mu,\sigma}((x - \mu)/\sigma)$ para un parámetro de localización $\mu \in \mathbb{R}$ y un parámetro de escala σ

$$H_{\xi}(x) = \begin{cases} \frac{1}{\sigma} \left(1 + \xi \frac{(x - \mu)}{\sigma}\right)^{-\frac{1}{\xi}-1} \exp\left(-\left(1 + \xi \frac{(x - \mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right) & \text{for } \xi \neq 0 \\ \exp(-x) \exp(-e^{-x}) & \text{for } \xi = 0 \end{cases}$$

where $1 + \xi x > 0$

- El parámetro ξ se conoce como el parámetro de forma de la distribución, y H_{ξ} define el tipo de distribución (distribuciones con localización y escala definidas)
- Esta distribución permite especificar tres tipos de distribución según ξ . Si $\xi > 0$, entonces la distribución es una distribución de Fréchet; si $\xi = 0$, la distribución es una Gumbel, y si $\xi < 0$, la distribución es una Weibull

$$\xi < 0 \Rightarrow x < \mu - \frac{\sigma}{\xi} \quad (\text{Weibull})$$

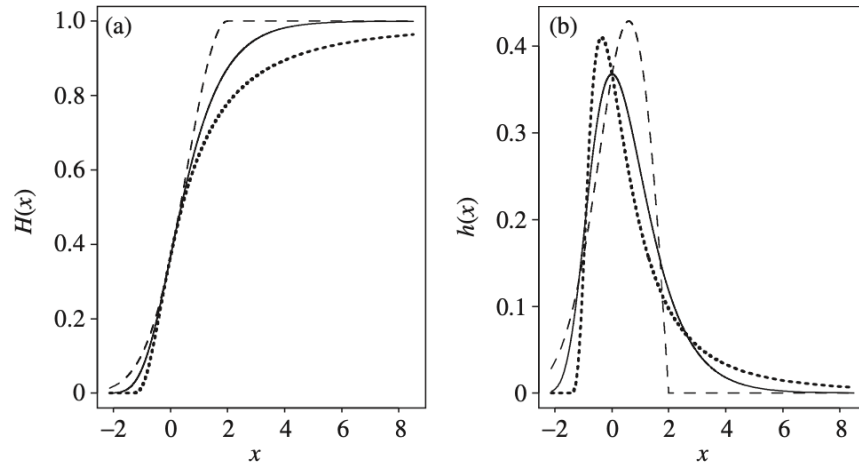
$$\xi = 0 \Rightarrow x \in \mathbb{R} \quad (\text{Gumbel})$$

$$\xi > 0 \Rightarrow x > \mu - \frac{\sigma}{\xi} \quad (\text{Fréchet})$$

- Para una x fija, el límite de H_{ξ} cuando ξ tiende a cero es H_0 (para ambos lados), de modo que la parametrización es continua en ξ y eso facilita el uso de esta distribución

$$\lim_{\xi \rightarrow 0} H_{\xi} = H_0$$

- La función de distribución y de densidad de la distribución GEV se puede representar a través de los siguientes gráficos y para diferentes valores de ξ :



————— = *Gumbel*
 = *Fréchet*
 - - - = *Weibull*

- La distribución Weibull tiene colas cortas con un punto asintótico derecho finito. Este punto se denota por $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$
- Las distribuciones de Fréchet y Gumbel tienen puntos asintótico derechos infinitos, pero la reducción de la pendiente de la cola de la distribución de Fréchet es más lenta que la de Gumbel
- Suponiendo que los máximos M_n de variables independientes e idénticamente distribuidas convergen en distribución cuando $n \rightarrow \infty$ bajo una normalización apropiada, se observa que la convergencia implica que existen una secuencia de constantes reales (d_n) y (c_n) (donde $c_n > 0$ para toda n) tal que se da la siguiente igualdad para una función de distribución $H(x)$ no degenerada:

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n \rightarrow \infty} [F(c_n x + d_n)]^n = H(x)$$

- Por resultados de probabilidad básica, se puede ver como la función de distribución de probabilidad de un máximo es $[F(\cdot)]^n$ para x_1, \dots, x_n independientes e idénticamente distribuidas

$$P(\max\{x_1, \dots, x_n\} \leq k) = P(x_1 \leq k, x_2 \leq k, \dots, x_n \leq k)$$

- Si esta igualdad se mantiene para una función de distribución H , entonces F está en el dominio de máxima atracción de H , denotado por $F \in MDA(H)$
- Si la convergencia de los máximos normalizados se da, el tipo de distribución asintótica (especificada por ξ) está únicamente

determinada, aunque la localización y la escala dependen exactamente de las secuencias normalizadoras escogidas. No obstante, siempre es posible escoger secuencias para que la distribución asintótica se obtenga en su forma estándar

- Si $F \in MDA(H)$ para una función de distribución no degenerada H , entonces H tiene que ser una distribución del tipo H_ξ (una distribución de valor extremo generalizada)
 - De este modo, si se cumple la convergencia anterior, entonces F pertenece a la familia de distribuciones de valor extremo generalizado
 - La función de distribución exponencial, la de la distribución de Pareto y la log-normal son distribuciones de valor extremo generalizadas dado que se pueden escoger unas secuencias normalizantes concretas

$$F(x) = 1 - e^{-\lambda x} \text{ for } x \geq 0 \Rightarrow F^n(c_n x + d_n) = \left(1 - \frac{e^{-x}}{n}\right)^n$$

$$F(x) = 1 - \left(\frac{\kappa}{\kappa + x}\right)^\alpha \text{ for } \alpha, \kappa > 0 \text{ \& } x \geq 0$$

$$\Rightarrow F^n(c_n x + d_n) = \left(1 - \frac{1}{n} \left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right)^n$$

$$F(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \text{ for } x \geq 0$$

- La teoría asintótica de convergencia de los máximos lleva consigo el comportamiento de los mínimos utilizando la siguiente identidad:

$$\min(X_1, X_2, \dots, X_n) = -\max(-X_1, -X_2, \dots, -X_n) = -M_n^*$$

- Los mínimos normalizados de muestras independientes e idénticamente distribuidas con función de distribución F convergen en distribución si la función de distribución $\tilde{F} = 1 - F(-x)$ está en el dominio de máxima atracción de una distribución de valor extremo

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n^* - d_n}{c_n} \leq x\right) = H_\xi(-x)$$

$$\Rightarrow \lim_{n \rightarrow \infty} P\left(\frac{M_n^* + d_n}{c_n} \leq x\right) = 1 - H_\xi(-x)$$

- Por lo tanto, distribuciones asintóticas para los mínimos son distribuciones del tipo $1 - H_\xi(-x)$
- Para la mayoría de aplicaciones es suficiente con notar que esencialmente todas las distribuciones continuas comunes de estadística o de ciencias actuariales están en $MDA(H_\xi)$ para algún valor de ξ . Por lo tanto, se pueden considerar que distribuciones subyacentes llevan a unos límites concretos para los máximos (Fréchet, Gumbel y Weibull)

- Las distribuciones que permiten obtener el límite de Fréchet $H_\xi(x)$ para $\xi > 0$ tienen una caracterización particular involucrando funciones que varían lentamente o que varían regularmente

- Una función positiva medible de Lebesgue L sobre $(0, \infty)$ varía lentamente en ∞ (escrito como $L \in \mathcal{R}_0$) si se cumple la siguiente condición:

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1 \quad \text{for } t > 0$$

- Una función positiva medible de Lebesgue h sobre $(0, \infty)$ varía regularmente en ∞ con índice $\rho \in \mathbb{R}$ si se cumple la siguiente condición:

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\rho \quad \text{for } t > 0$$

- Las funciones que varían lentamente son funciones que, en comparación con funciones potenciales, cambian relativamente lento para una x (como el logaritmo $L(x) = \ln(x)$). Las funciones que varían regularmente son funciones que se pueden representar como funciones potenciales multiplicadas por funciones que varían lentamente ($h(x) = x^\rho L(x)$ para alguna $L \in \mathcal{R}_0$)

- Para $\xi > 0$, se cumple que $F \in MDA(H_\xi) \Leftrightarrow 1 - F(x) = x^{-\frac{1}{\xi}} L(x)$ para alguna función $L \in \mathcal{R}_0$

- Esto significa que las distribuciones que derivan al caso de Fréchet son distribuciones con colas que varían regularmente con un índice negativo de variación. Su cola decae esencialmente como una función potencial, y una tasa de caída $\alpha = 1/\xi$ se suele llamar como el índice de cola de la distribución

- Una consecuencia de este teorema es que el punto extremo derecho de cualquier distribución en el MDA de Fréchet satisface $x_F = \infty$
- Los máximos normalizados de variables distribuidas como una Pareto convergen a una distribución de Fréchet. La cola de una función de distribución de Pareto se puede escribir como $1 - F(x) = x^{-\alpha}L(x)$ cuando $L(x) = (\kappa^{-1} + x^{-1})^{-\alpha}$ es una función que varía lentamente, dado que, cuando $x \rightarrow \infty$, $L(x) \rightarrow \kappa^\alpha$
- Estas distribuciones son las distribuciones más estudiadas de la teoría del valor extremo, y son de particular interés para aplicaciones financieras porque tienen colas gruesas con momentos infinitos
 - Si X es una variable aleatoria no negativa cuya función de distribución es F y X es un elemento de $MDA(H_\xi)$ para $\xi > 0$, entonces puede demostrarse que $E(X^k) = \infty$ para $k > 1/\xi$
 - Si, para una $\varepsilon > 0$ pequeña, la distribución está en $MDA(H_{(1/2)+\varepsilon})$, la distribución tiene varianza infinita, mientras que si $MDA(H_{(1/4)+\varepsilon})$, entonces es una distribución con cuarto momento infinito
 - A parte de la distribución de Pareto, otras distribuciones que llevan al límite de Fréchet son la distribución de Fréchet misma, la gamma inversa, la t-Student, la gamma logarítmica, la F de Fisher-Snedecor y la distribución de Burr
- La caracterización de las distribuciones en la clase de Gumbel es más complicada que en la de Fréchet. No obstante, hay mucha variedad en las colas de las distribuciones en esta clase (por ejemplo, la distribución normal y la lognormal pertenecen a esta clase)
 - Una variable aleatoria positiva con una función de distribución en $MDA(H_0)$ tiene momentos finitos de cualquier orden positivo (de modo que $E(X^k) < \infty$ para toda $k > 0$). Además, los puntos extremos derechos de las distribuciones en esta clase satisfacen que $x_F \leq \infty$, por lo que $x_F = \infty$ es posible
 - En el modelaje financiero normalmente se asume erróneamente que los modelos más interesantes son las distribuciones con colas potenciales de la clase de Fréchet, pero la clase de Gumbel es también interesante porque contiene muchas distribuciones con colas mucho más gruesas que las de la normal, aunque no sean colas potenciales que varían regularmente (distribución hiperbólica e hiperbólica generalizada, excepto por la t-Student)

- Otras distribuciones en $MDA(H_0)$ incluyen la gamma, la chi cuadrado, la Weibull estándar (diferente de la distribución en la GEV), la Gumbel misma y la otra
- El caso Weibull es el menos interesante para el modelaje financiero, como mínimo en el área de riesgo de mercado, dado que las distribuciones en esta clase tienen puntos extremos derechos finitos
 - Aunque la mayoría de pérdidas financieras potenciales son, en la práctica, están acotadas, se tiende a favorecer modelos con soporte infinito para el modelaje de pérdidas. Una excepción puede ser en el área de riesgo de crédito, en donde las distribuciones en el intervalo unitario son muy útiles
 - Para $\xi > 0$, se cumple que $F \in MDA(H_\xi) \Leftrightarrow 1 - F(x_F - x^{-1}) = x^{\frac{1}{\xi}} L(x)$ para alguna función $L \in \mathcal{R}_0$
 - Se puede demostrar que la distribución beta con la densidad $f_{\alpha, \beta}$ está en $MDA(H_{-1/\beta})$, junto al caso especial $\alpha = \beta = 1$ (la distribución uniforme)
- La teoría estándar de las secciones anteriores se centre en los máximos de secuencias de variables independientes e idénticamente distribuidas. Con las series temporales financieras en mente, uno puede revisar la teoría para máximos de series temporales estrictamente estacionarias y encontrar las distribuciones límite que aplican
 - En esta sección $(X_i)_{i \in \mathbb{Z}}$ denotan que las series temporales estrictamente estacionarias con función de distribución estacionaria F , y $(\tilde{X}_i)_{i \in \mathbb{Z}}$ denota el proceso iid asociado (un proceso de ruido blanco estricto con la misma función de distribución)
 - Definiendo los máximos como $M_n = \max\{X_1, X_2, \dots, X_n\}$ y $\tilde{M}_n = \max\{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n\}$, para muchos procesos $(X_i)_{i \in \mathbb{N}}$ se puede demostrar que existe un número real $\theta \in (0, 1]$ tal que se cumple la siguiente condición para una $H(x)$ no degenerada:

$$\lim_{n \rightarrow \infty} P\left(\frac{\tilde{M}_n - d_n}{c_n} \leq x\right) = H(x) \Leftrightarrow \lim_{n \rightarrow \infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = H^\theta(x)$$
 - Para este tipo de procesos, el valor θ se conoce como índice extremo del proceso (no es el índice de cola de las distribuciones de la clase de Fréchet)
 - Para procesos con un índice extremo, los máximos normalizados convergen en distribución siempre que el máximo del proceso iid

asociado converja en distribución (que la función de distribución subyacente $F \in MDA(H_\xi)$ para alguna $\xi > 0$)

- Además, como $H_\xi^\theta(x)$ es una distribución del mismo tipo que $H_\xi(x)$, la distribución limitante de los máximos normalizados de las series dependientes es una distribución GEV con el mismo parámetro ξ como el límite para los datos iid asociados (solo la localización y la escalación de la distribución pueden cambiar)
- Escribiendo $u = c_n x + d_n$, se observa que para una n lo suficientemente grande, entonces obtener los siguientes resultados:

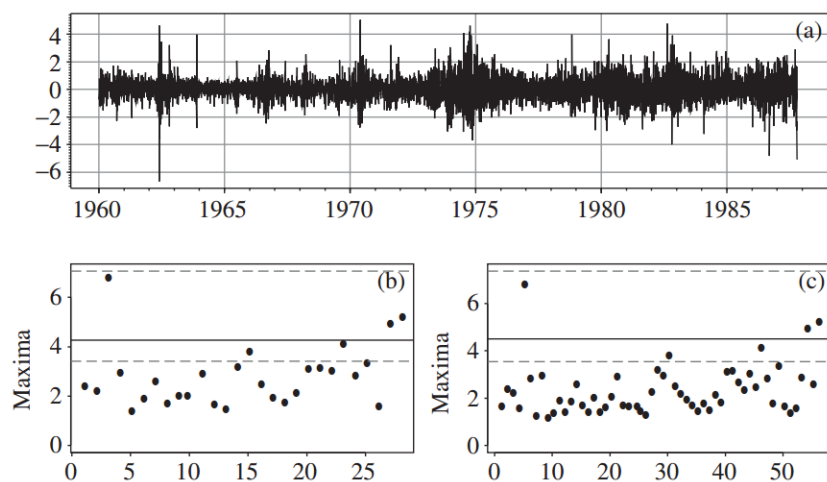
$$P(M_n \leq u) \approx P^\theta(\tilde{M} \leq u) = F^{n\theta}(u)$$

- Para una u grande, la distribución de probabilidad del máximo de n observaciones de la serie temporal con el mismo extremo θ se puede aproximar por la distribución del máximo de $n\theta < n$ observaciones para las series iid asociadas. La cantidad $n\theta$ se puede entender como una que cuenta el número de clústeres independientes en n observaciones, y θ normalmente se interpreta como el recíproco del tamaño medio de un clúster
- No todos los procesos estrictamente estacionarios tienen un índice extremo, pero para los tipos de series temporales usadas en el modelaje financiero, generalmente sí existe uno
 - Esencialmente, solo se tiene que distinguir entre casos cuando $\theta = 1$ y los casos cuando $\theta < 1$. Para el primer caso, no hay tendencia para clusterizar en niveles altos, y los máximos de muestras grandes de la serie temporal se comportan exactamente como los máximos de muestras iid de tamaños similares; para el segundo caso, uno debe tener en cuenta de la tendencia de los valores extremos para clusterizarse
 - Los procesos de ruido blanco estrictos (variables aleatorias iid) tienen índices extremos $\theta = 1$; los procesos ARMA con ruido blanco gaussiano estrictos tienen $\theta = 1$, pero si la distribución de innovaciones está en $MDA(H_\xi)$ para $\xi < 0$, entonces $\theta < 1$; y los modelos ARCH y GARCH tienen $\theta < 1$

Table 5.1. Approximate values of the extremal index as a function of the parameter α_1 for the ARCH(1) process in (4.22).

α_1	0.1	0.3	0.5	0.7	0.9
θ	0.999	0.939	0.835	0.721	0.612

- Habiendo desarrollado la teoría sobre el modelaje de máximos de bloque, se pueden tratar aspectos metodológicos más concretos
 - Se supone que se tienen datos de una distribución desconocida F que se asume que está en el dominio de atracción de una distribución de valor extremo H_ξ para alguna ξ
 - Si los datos son realizaciones de las variables iid, o de variables de un proceso con índice extremo tal como un GARCH, la implicación de la teoría es que la verdadera del máximo M_n del bloque de n observaciones se puede aproximar por una distribución GEV de tres parámetros $H_{\xi, \mu, \sigma}$ con una n lo suficientemente grande. Se usa esta idea ajustando el modelo a los datos
 - Obviamente se necesitan observaciones repetidas de un máximo de un bloque de n observaciones, y se asume de que los datos se pueden dividir en m bloques de tamaño n . Esto tiene más sentido cuando hay maneras naturales de hacer bloques con los datos
 - Se consideran aplicaciones financieras en donde datos de rendimientos diarios se dividen en bloques anuales (o semestrales o trimestrales) y se analiza el máximo diario cae dentro de estos bloques



- Se denota el máximo del bloque j como M_{nj} , de modo que los datos son $M_{n1}, M_{n2}, \dots, M_{nm}$. La distribución GEV puede ajustarse usando varios métodos, incluyendo el de máxima verosimilitud
 - Una alternativa es el método de momentos ponderados por probabilidad

- Al implementar la máxima verosimilitud, normalmente se asume que el bloque de tamaño n es lo suficientemente grande, de modo que, independientemente de si los datos subyacentes son dependientes o no, las observaciones máximas del bloque se pueden tomar como independientes. En este caso, escribiendo $h_{\xi,\mu,\sigma}$ como la densidad de la distribución GEV, la verosimilitud logarítmica sería la siguiente:

$$\begin{aligned}
 l(\xi, \mu, \sigma; M_{n1}, \dots, M_{nm}) &= \sum_{i=1}^m \ln h_{\xi,\mu,\sigma}(M_{ni}) = \\
 &= -m \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \ln \left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right) - \sum_{i=1}^m \ln \left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right)^{-\frac{1}{\xi}} \\
 &\text{such that } \sigma > 0 \quad \& \quad 1 + \xi \frac{M_{ni} - \mu}{\sigma} > 0 \text{ for } \forall i
 \end{aligned}$$

- Mientras que esta representa un problema de verosimilitud irregular, debido a la dependencia del espacio paramétrico con los valores de los datos, la consistencia y la eficiencia asintótica de los estimadores resultantes se puede establecer usando el caso en el que $\xi > -1/2$
- Al determinar el número y el tamaño de los bloques (m y n respectivamente), hay un *trade off*: un valor grande de n conlleva una aproximación más precisa de la distribución de los máximos de bloque por una GEV y un sesgo más pequeño en las estimaciones de los parámetros. En el caso de datos dependientes, se recomienda el uso de algunos bloques más grandes que en el caso de independencia, dado que la dependencia hace que la convergencia de la GEV sea más lenta (el tamaño muestral θn es menor a n)
- La GEV ajustada se puede utilizar para estimar dos cantidades relacionadas que describen la ocurrencia de eventos de estrés o *stress events*
 - Por un lado, se puede estimar el tamaño del evento que ocurre en una frecuencia predeterminada, llamado el problema del rendimiento y el nivel, o *return-level problem*
 - Por otro lado, se puede estimar la frecuencia de un evento de estrés que tiene un tamaño predeterminado, llamado el problema de rendimiento y periodo, o *return-period problema*

- Siendo H la función de distribución real del máximo del bloque de n observaciones, el nivel de rendimiento k del bloque de n observaciones es $r_{n,k} = q_{1-1/k}(H)$ (el cuartil $1 - 1/k$ de H)

- Este se puede interpretar como el nivel que se excede en uno de cada k bloques de n observaciones de media
- Usando el modelo ajustado, el estimador del nivel de rendimiento será el siguiente:

$$\hat{r}_{n,k} = H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}^{-1} \left(1 - \frac{1}{k} \right) = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(-\ln \left(1 - \frac{1}{k} \right) \right)^{-\hat{\xi}} - 1 \right]$$

- Siendo H la función de distribución real del máximo del bloque de n observaciones, el periodo de rendimiento de un evento $\{M_n > u\}$ se define como $k_{n,u} = 1/\bar{H}(u)$

- El periodo de rendimiento $k_{n,u}$ se define de tal manera que el nivel de rendimiento $k_{n,u}$ del bloque de n observaciones es u . De este modo, se espera que en $k_{n,u}$ bloques se vea un solo bloque en el que el nivel u se ha excedido
- Si ha habido una gran tendencia a que los valores extremos se clustericen, uno puede esperar ver muchos excesos del nivel dentro del bloque
- Asumiendo que H es la función de distribución de una distribución GEV y usando el modelo ajustado, se estima el periodo de rendimiento con $\hat{k}_{n,u} = 1/\bar{H}_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}(u)$

- Tanto $\hat{r}_{n,k}$ como $\hat{k}_{n,u}$ son formas simples funcionales de los parámetros estimados de la distribución GEV. Además de calcular estimadores puntuales para estas cantidades, se deberían dar intervalos de confianza que reflejen el error en la estimación de los parámetros

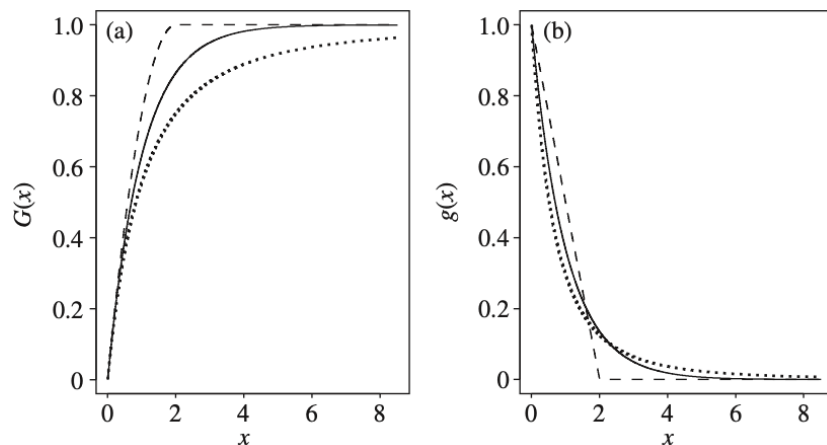
- Un buen método es basar estos intervalos de confianza en el estadístico de razón de verosimilitud, reparametrizando la distribución GEV en términos de las cantidades de interés
- En el caso del nivel de rendimiento, si $\phi = H_{\xi, \mu, \sigma}^{-1}(1 - 1/k)$ y se parametriza la distribución GEV como $\theta = (\phi, \xi, \sigma)'$ en vez de $\theta = (\xi, \mu, \sigma)'$. La estimación de máxima verosimilitud de ϕ es el estimador visto anteriormente, y el intervalo de confianza se puede construir a partir de este

La teoría del valor extremo: excesos de umbrales

- El método de los máximos de bloque malgasta muchos datos, dado que solo se realiza el análisis en las pérdidas máximas en grandes bloques, por lo que se suelen utilizar los métodos basados en excesos de umbrales, donde se usan todos los datos que exceden un nivel alto particular
 - El modelo distribucional principal para los excesos es la distribución generalizada de Pareto (GPD). La función de distribución de la GPD se da por la siguiente expresión:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}} & \text{for } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{for } \xi = 0 \end{cases}$$

- En este caso, el parámetro β es el parámetro de forma y $\beta > 0$, mientras que ξ es el parámetro de escala, y $x \geq 0$ cuando $\xi \geq 0$ y $0 \leq x \leq -\beta/\xi$ cuando $\xi < 0$
- Cuando $\xi > 0$, la distribución es una distribución de Pareto (con una cola gruesa); si $\xi = 0$, la distribución es una exponencial (con una cola exponencial), y si $\xi < 0$, se tiene una distribución de Pareto tipo II (con colas cortas)



- Para una x fija, la forma paramétrica es continua en ξ , de modo que $G_{\xi,\beta}(x) \rightarrow G_{0,\beta}(x)$ cuando $\xi \rightarrow 0$

- En términos del dominio de atracción se tiene que $G_{\xi,\beta}(x) \in MDA(H_\xi)$ para toda $\xi \in \mathbb{R}$. Cuando $\xi \neq 0$, es fácil ver como esta proposición se cumple
- La media de la GPD se define suponiendo que $\xi < 1$, de modo que resulta en la siguiente expresión:

$$E(X) = \frac{\beta}{1 - \xi}$$

- El rol de la GPD en la teoría del valor extremo es de ser un modelo natural para la distribución de exceso sobre un umbral alto, y se puede definir simultáneamente el concepto de función de exceso medio

- Siendo X una variable aleatoria con función de distribución F , la distribución de exceso sobre el umbral $u > 0$ tiene la siguiente función de distribución, donde $x_F \leq \infty$ es el punto asintótico derecho de F :

$$F_u(X) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}$$

$$\text{for } 0 \leq x < x_F - u$$

$$f_u(X) = \frac{f(x + u)}{1 - F(u)}$$

- Siendo X una variable aleatoria con función de distribución F y media finita, la función de exceso medio viene dada por la siguiente ecuación:

$$me(u) = E(X - u \leq x | X > u)$$

- La función de distribución de exceso F_u describe la distribución de las pérdidas excesivas sobre el nivel u , dado que u sea excedido, mientras que la función de exceso medio expresa la media de F_u como una función de u
- En el caso de que se tenga una distribución de Pareto generalizada, se pueden calcular las formas funcionales de cada uno:

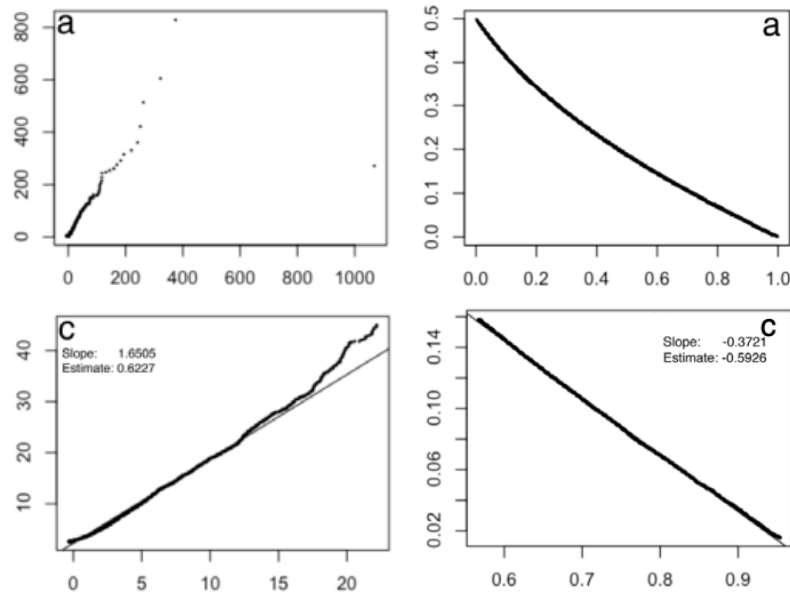
$$F_u(x) = G_{\xi,\beta}(x) \quad \text{and} \quad \beta(u) = \beta + \xi u$$

$$\text{where } 0 \leq x < \infty \text{ if } \xi \geq 0 \quad \& \quad 0 \leq x < -\frac{\beta}{\xi} \text{ if } \xi < 0$$

$$me(u) = \frac{\beta(u)}{1 - \xi} = \frac{\beta + \xi u}{1 - \xi}$$

where $0 \leq u < \infty$ if $0 \leq \xi < 1$ & $0 \leq u < -\frac{\beta}{\xi}$ if $\xi < 0$

- Como se puede ver, la F_u será una GPD, pero con parámetro de escala linealmente creciente en u y la $me(u)$ será una función lineal de u , de modo que la GPD cumple con un tipo de propiedad de estabilidad. La pendiente, tal y como se ha visto anteriormente, puede tener pendiente positiva o negativa



- A través de la función $me(u)$, es posible obtener un estimador para poder obtener un umbral \hat{u} a partir de los datos

- Es posible estimar la función a través del siguiente estimador:

$$\widehat{me}(u) = \frac{\sum_{i=1}^n (X_i - u) I_{[X_i > u]}}{\sum_{i=1}^n I_{[X_i > u]}} \quad \text{for } u \geq 0$$

- Una vez con la función empírica, se puede crear un gráfico para la función de exceso medio muestral y analizar su comportamiento. Bajo una distribución GPD, la función tendría que ser lineal (por la expresión derivada anteriormente), de modo que se puede escoger un nivel u en el punto en el que el gráfico comienza a aproximarse a una función lineal

- El teorema de Pickands-Balkema-de Haan expresa que se puede encontrar una función $\beta(u)$ que cumpla la siguiente convergencia, si, y solo si, $F \in MDA(H_\xi)$ con $\xi \in \mathbb{R}$:

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0$$

- Por lo tanto, las distribuciones con máximos normalizados que convergen a una distribución GEV constituyen un conjunto de distribuciones cuyas distribuciones de exceso que convergen a una GPD cuanto más incrementa el nivel ($u \rightarrow x_F$)
- Como muchas distribuciones utilizadas en estadística pertenecen a la $MDA(H_\xi)$ para alguna ξ , este teorema muestra que la distribución GPD es una distribución canónica para modelar las pérdidas excesivas sobre niveles altos
- Se puede explotar el teorema anterior asumiendo que se está lidiando con la distribución de pérdida $F \in MDA(H_\xi)$ para la cual, dado un umbral alto u , uno puede modelar F_u con una GPD
 - Se considera que F es una distribución de pérdida con punto extremo derecho x_F y se asume que para un umbral alto u se tiene que $F_u(x) = G_{\xi, \beta}(x)$ para $0 \leq x < x_F - u$ y alguna $\xi \in \mathbb{R}$ y $\beta > 0$
 - Esta es claramente una idealización, dado que, en la práctica, la distribución no será exactamente una GPD, pero se usa la suposición para hacer unos cálculos
 - Dados los datos de pérdidas X_1, \dots, X_n de F , un número aleatorio N_u excederá el umbral u , etiquetando así estos datos como $\tilde{X}_1, \dots, \tilde{X}_{N_u}$
 - Para cada uno de estos excesos, se calcula la cantidad $Y_j = \tilde{X}_j - u$ de pérdida excesiva. Se desea estimar los parámetros de un modelo GPD al ajustar la distribución a las N_u pérdidas excesivas
 - Hay varias maneras de ajustar la GPD, incluyendo estimadores de máxima verosimilitud y momentos ponderados por probabilidad. El método de máxima verosimilitud se utiliza más comúnmente y es fácil de implementar si los datos de exceso provienen de realizaciones de variables aleatorias independientes, porque de este modo la densidad conjunta será el producto de densidades conjuntas
 - Escribiendo $g_{\xi, \beta}$ para la densidad de la GPD, la verosimilitud logarítmica es la siguiente, la cual se tiene que maximizar sujeta a que $\beta > 0$ y $1 + \xi Y_j / \beta > 0$ para toda j :

$$\begin{aligned}\ln L(\xi, \beta; Y_1, \dots, Y_{N_u}) &= \sum_{j=1}^{N_u} \ln g_{\xi, \beta}(Y_j) = \\ &= -N_u \ln \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \ln \left(1 + \xi \frac{Y_j}{\beta}\right)\end{aligned}$$

- Para datos sobre riesgo operacional o asegurador, la suposición de datos independientes e idénticamente distribuidos no suele ser un problema, pero sí lo es para series temporales
 - Si los datos son dependientes serialmente pero no enseñan ninguna tendencia a formar clústeres de valores extremos, entonces esto puede sugerir que el proceso tiene un índice extremo $\theta = 1$. En este caso, la teoría asintótica sugiere que el modelo limitante para los excesos en un umbral alto es un proceso de Poisson y que las cantidades de pérdidas excesivas siguen una distribución de Pareto generalizada y son iid
 - Si hay clusterización de los valores extremos, se puede sugerir un índice extremo $\theta < 1$ (como un proceso GARCH), la suposición es menos satisfactoria. El enfoque más fácil es ignorar el problema y considerar el método de máxima verosimilitud como uno de cuasi-máxima verosimilitud, donde la verosimilitud no se especifica correctamente con respecto a la estructura de dependencia serial de los datos
 - Este último enfoque dará estimadores puntuales razonables, aunque los errores estándar pueden ser demasiado pequeños
- A partir del modelo que se ha ajustado para la distribución de excesos sobre u , uno puede fácilmente inferenciar un modelo para la distribución de exceso sobre cualquier umbral alto
 - Bajo la suposición hecha, $F_{v(x)} = G_{\xi, \beta + \xi(v-u)}(x)$ para cualquier umbral alto $v \geq u$. Considerando $\bar{F}(x) = 1 - F(x)$, se pueden obtener los siguientes resultados:

$$\begin{aligned}\bar{F}_v(x) &= \frac{\bar{F}(v+x)}{\bar{F}(v)} = \frac{\bar{F}(u + (x + v - u))}{\bar{F}(u)} \frac{\bar{F}(u)}{\bar{F}(u + (v - u))} = \\ &= \frac{\bar{F}_u(x + v - u)}{\bar{F}_u(v - u)} = \frac{\bar{G}_{\xi, \beta}(x + v - u)}{\bar{G}_{\xi, \beta}(v - u)} = \bar{G}_{\xi, \beta + \xi(v-u)}(x)\end{aligned}$$

- Por lo tanto, la distribución de excesos sobre umbrales altos sigue siendo una GPD con el mismo parámetro ξ pero una escala que crece linealmente con el umbral v . Si $\xi < 1$, la función de exceso medio se da por la siguiente expresión:

$$me(v) = \frac{\beta + \xi(v - u)}{1 - \xi} = \frac{\xi v}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi}$$

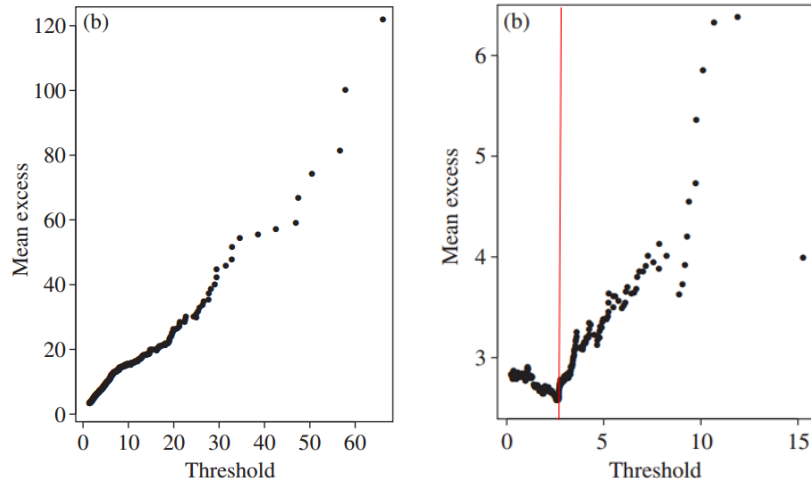
$$for \begin{cases} u \leq v < \infty & \text{if } \xi \in [0,1) \\ u \leq v \leq u - \frac{\beta}{\xi} & \text{if } \xi < 0 \end{cases}$$

- La linealidad de la función en v normalmente se usa como un diagnóstico para los datos que admiten un modelo GPF para la distribución de excesos. Forma una base para el método gráfico para escoger un umbral apropiado
- Para datos de pérdida positivos X_1, \dots, X_n , se define la función de exceso medio muestral como un estimador empírico de la función de exceso medio, dado por la siguiente expresión:

$$me_n(v) = \frac{\sum_{i=1}^n (X_i - v) I_{\{X_i > v\}}}{\sum_{i=1}^n I_{\{X_i > v\}}}$$

- Para estudiar la función, generalmente se construye un gráfico de excesos medios $\{(X_{i,n}, e_n(X_{i,n})): 2 \leq i \leq n\}$, donde $X_{i,n}$ denota el estadístico de orden (en orden ascendente). Si los datos apoyan un modelo GPD sobre un umbral muy alto, entonces la función de exceso medio anterior sugiere que el diagrama se debería volver cada vez más lineal para valores altos de v
- Una línea con pendiente positiva indica que el modelo GPD tiene un parámetro de forma aproximadamente nulo, o, en otras palabras, una distribución exponencial de los excesos. Una línea con pendiente negativa, indica que el parámetro de forma será negativo
- Estas son las situaciones ideales, pero en la práctica se requiere experiencia para leer estos diagramas de exceso medio
- Aún para datos que son genuinamente de una GPD, el diagrama pocas veces es perfectamente lineal, particularmente si se mira el extremo derecho, en donde se promedia un pequeño número de excesos grandes

- En verdad, uno normalmente omite los puntos al final por consideración, ya que pueden distorsionar mucho las conclusiones. Si se observa evidencia visual de que el diagrama se vuelve lineal, entonces se puede seleccionar como umbral u aquel valor en donde la función deja de ser lineal o hay un *kink* o punta



- El modelo GPD para las pérdidas excesivas permite estimar la cola de la distribución de pérdidas subyacente F y sus medidas de riesgo asociadas
 - Bajo la suposición de que $F_u(x) = G_{\xi, \beta}(x)$ para $0 \leq x \leq x_F - u$, que $\xi > 0$ y que $\beta > 0$, se puede obtener la siguiente igualdad:

$$\begin{aligned}
 [1 - F(x)] &= P(X > u)P(X > x|X > u) = \\
 &= [1 - F(u)]P(X - u > x - u|X > u) = \\
 &= [1 - F(u)][1 - F(x - u)] = [1 - F(u)] \left(1 + \xi \frac{x - u}{\beta}\right)^{-1/\xi}
 \end{aligned}$$

- Si se sabe $F(u)$, es posible calcular la probabilidad de las colas para un valor $x \geq u$
- Esta fórmula se puede invertir para poder obtener un cuartil alto de la distribución subyacente, que se puede interpretar como el VaR_α y, a la vez, obtener el ES_α
 - Para $\alpha \geq F(u)$ se tiene que el VaR se puede obtener con la siguiente fórmula:

$$VaR_\alpha = q_\alpha(F) = u + \frac{\beta}{\xi} \left(\left(\frac{1 - \alpha}{1 - F(u)} \right)^{-\xi} - 1 \right)$$

- Asumiendo que $\xi < 1$, el ES_α asociado se puede calcular de la siguiente manera:

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 q_x(F) dx = \frac{VaR_\alpha}{1-\xi} + \frac{\beta - \xi u}{1-\xi}$$

- La suposición hecha sobre la distribución de pérdida y el hecho de que $F_v(x) = G_{\xi, \beta + \xi(v-u)}(x)$ implica que las pérdidas excesivas por encima del VaR_α siguen una GPD satisfaciendo $F_{VaR_\alpha} = G_{\xi, \beta + \xi(VaR_\alpha - u)}(x)$

- El estimador del *expected shortfall* se puede obtener sumando la media de esta distribución al VaR_α , de modo que $ES_\alpha = VaR_\alpha + me(VaR_\alpha)$
- Es interesante ver como la *ratio* entre ambas medidas de riesgo se comportan para valores grandes del cuantil de probabilidad α . Se puede obtener, por tanto, el siguiente límite cuando $\alpha \rightarrow 1$:

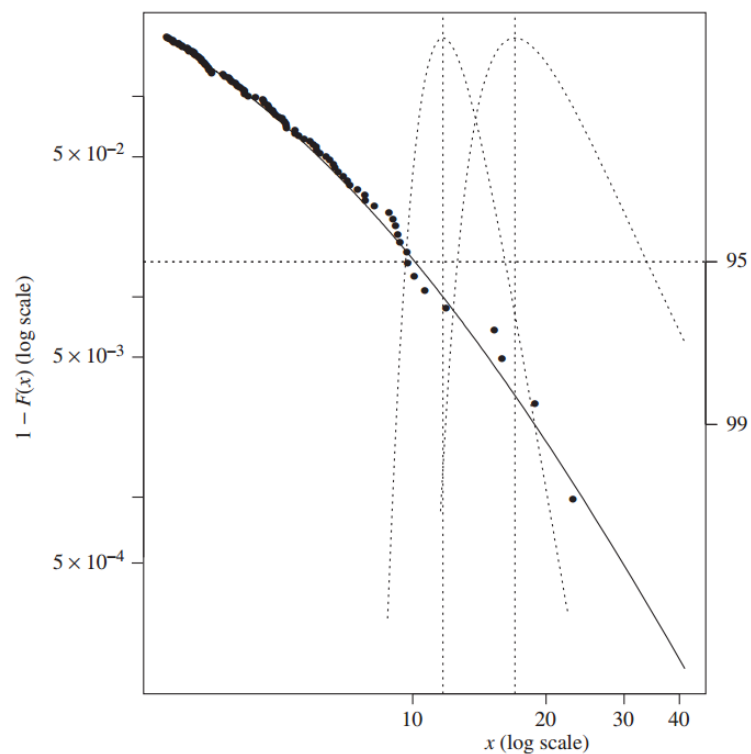
$$\lim_{\alpha \rightarrow 1} \frac{ES_\alpha}{VaR_\alpha} = \begin{cases} (1-\xi)^{-1} & \text{if } 0 \leq \xi < 1 \\ 1 & \text{if } \xi < 0 \end{cases}$$

- Se destaca que, bajo la suposición, las probabilidades de cola, los $VaRs$ y los ES se dan por fórmulas de la forma $g(\xi, \beta, 1 - F(u))$

- Asumiendo que se ha ajustado un modelo GPD para las pérdidas excesivas sobre el umbral u , se estiman estas cantidades reemplazando ξ y β en las fórmulas anteriores. Obviamente, se requiere una estimación de $1 - F(u)$ y se puede tomar un estimador empírico simple como N_u/n
- Al hacer esto, implícitamente se asume de que hay una proporción suficiente de valores muestrales sobre el umbral u para estimar $1 - F(u)$ de manera fiable. No obstante, se intenta recuperar el método empírico usando un tipo de extrapolación basada en la GPD para más probabilidades de cola extremas y medidas de riesgo
- Para las probabilidades de cola se obtiene un estimador, el cual solo es válido para $x \geq u$. Para $\alpha \geq 1 - N_u/n$, se obtiene estimadores puntuales análogos de VaR_α y ES_α , el cual tiene la siguiente forma:

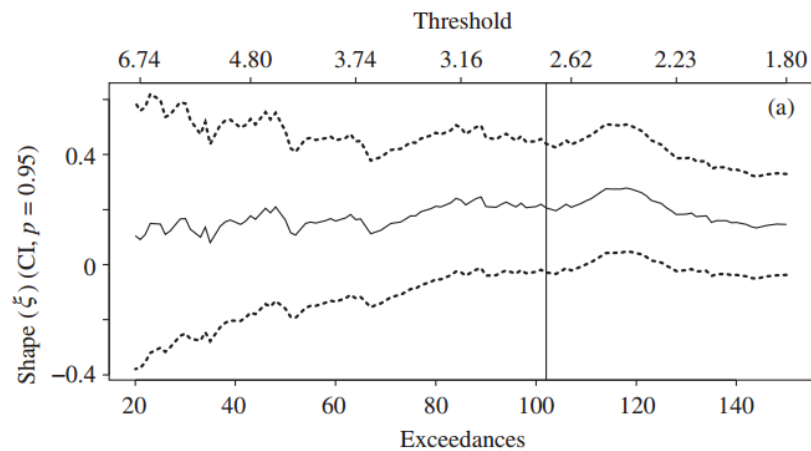
$$\hat{F} = \frac{N_u}{n} \left(1 + \xi \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}}$$

- Si se ha tomado el enfoque de máxima verosimilitud para estimar ξ y β , entonces se pueden dar intervalos de confianza sencillos para $g(\hat{\xi}, \hat{\beta}, N_u/n)$ para tener en cuenta la incertidumbre de $\hat{\xi}$ y $\hat{\beta}$ (pero ignorar la de N_u/n como estimador de $1 - F(u)$). Se usa el enfoque descrito anteriormente para los niveles de rendimiento, en donde se reparametriza en términos de $\phi = g(\xi, \beta, N_u/n)$ y el intervalo de confianza para $\hat{\phi}$ se construye basado en el contraste de razón de verosimilitud
- Se puede hacer un diagrama con las probabilidades de cola estimadas (con ejes logarítmicos), en donde las líneas de puntos son estimaciones de medidas de riesgo e intervalos de confianza. En este caso, las curvas de puntos son curvas de perfiles de la verosimilitud para el VaR como un parámetro de la GPD (se construye usando el contraste de razón de verosimilitud); la recta horizontal indica el nivel de confianza; y la recta vertical indica el estimador del cuartil

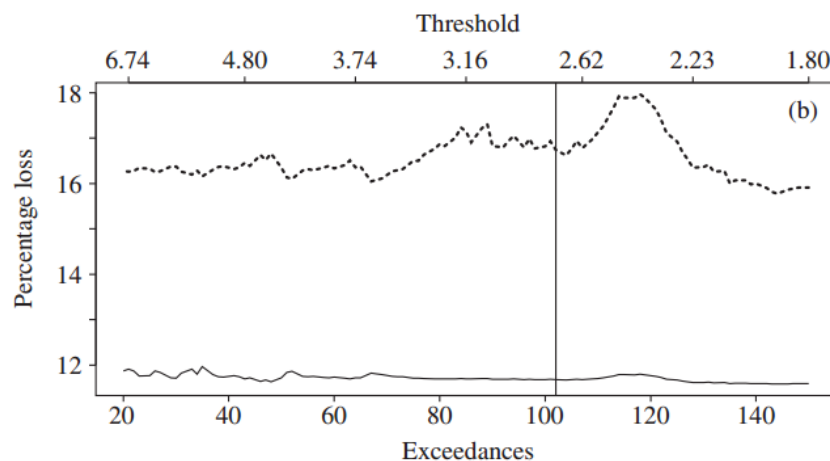


- Antes de acabar con el modelaje de la cola con la GPD, es importante ver que tan sensible son las estimaciones de las medidas de riesgo a la elección del umbral
 - Debido a la importancia del parámetro ξ en determinar el peso de la cola y la relación entre cuantiles y *expected shortfalls*, se puede estimar ξ para cada diferentes umbrales que contienen un número de excesos dentro de un intervalo concreto. El punto

en el que las estimaciones comienzan a diferir se puede encontrar a través de un diagrama de las estimaciones y las bandas de confianza para las diferentes estimaciones



- También se puede hacer el mismo análisis para otras medidas como el VaR a un nivel α concreto



- El método de la GPD no es el único método para poder estimar la distribución de la cola, sino que también existen métodos alternativos como el de Hill, que permiten modelar colas de distribuciones con colas gruesas
 - Se asume que la función de distribución subyacente está en el dominio de atracción máximo de la distribución de Fréchet, de modo que la cola tiene la siguiente forma, para una función que L que varíe lentamente y un parámetro positivo α :

$$1 - F(x) = x^{-\alpha} L(x)$$

- Tradicionalmente, la atención se centra en el índice de cola α , el cual es el recíproco ξ^{-1} , de modo que se intenta encontrar un estimador para α basado en datos idénticamente distribuidos X_1, X_2, \dots, X_n

- Un resultado importante a tener en cuenta para la derivación de propiedades útiles del estimador de Hill es el teorema de Karamata. Siendo L una función que varía lentamente y que está acotada en el intervalo $[x_0, \infty)$ para alguna $x_0 \geq 0$, entonces se cumplen las siguientes proposiciones:

$$(a) \text{ for } \kappa > -1, \int_{x_0}^x t^\kappa L(t) dt \rightarrow \frac{1}{\kappa + 1} x^{\kappa+1} L(x) \text{ as } x \rightarrow \infty$$

$$(b) \text{ for } \kappa < -1, \int_x^\infty t^\kappa L(t) dt \rightarrow -\frac{1}{\kappa + 1} x^{\kappa+1} L(x) \text{ as } x \rightarrow \infty$$

- Esencialmente, el teorema expresa que la función que varía lentamente se puede llevar afuera de la integral como si fuera una constante
- El estimador de Hill se puede derivar a través de considerar la función de exceso medio de una pérdida logarítmica $\ln X$, donde X es una variable aleatoria con función de distribución $x^{-\alpha} L(x)$

$$\begin{aligned} me^*(\ln u) &= E(\ln X - \ln U \mid \ln X > \ln U) = \\ &= \frac{1}{1 - F(u)} \int_u^\infty (\ln x - \ln u) dF(x) = \frac{1}{1 - F(u)} \int_u^\infty \frac{1 - F(x)}{x} dx = \\ &= \frac{1}{1 - F(u)} \int_u^\infty x^{-(\alpha+1)} L(x) dx \end{aligned}$$

- Para una u suficientemente grande, la función $L(x)$ para $x \geq u$ puede ser tratada esencialmente como una constante sacarse de la integral. A través del teorema de Karamata, se obtiene que, para $u \rightarrow \infty$, $me^*(\ln u)$ sigue la siguiente distribución:

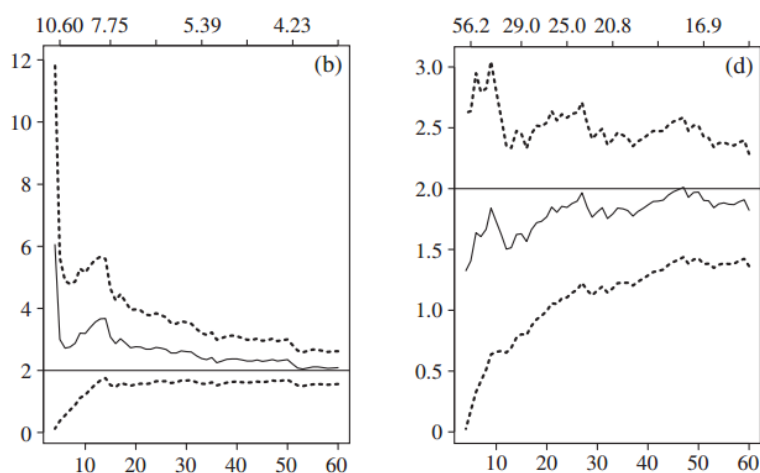
$$me^*(\ln u) \rightarrow \frac{L(u)u^{-\alpha}\alpha^{-1}}{1 - F(u)} = \alpha^{-1} = \xi \text{ as } n \rightarrow \infty$$

- De este modo, $me^*(\ln u)\alpha \rightarrow 1$ cuando $u \rightarrow \infty$. Se espera que se observe un comportamiento similar en la función exceso medio muestral me_n^* , por lo que se espera que $me^*(\ln X_{k,n}) \approx \alpha^{-1}$ para una n grande y una k lo suficientemente pequeña (donde $X_{n,n} \leq \dots \leq X_{1,n}$ son los estadísticos de orden)

- Evaluando $me^*(\ln X_{k,n})$ se puede obtener un estimador $\hat{\alpha}_{k,n}^{(H)} = [(k-1)^{-1} \sum_{j=1}^k \ln X_{(j)} - \ln X_{(k)}]$, y a partir de este se puede construir el estimador de Hill a través de una pequeña modificación

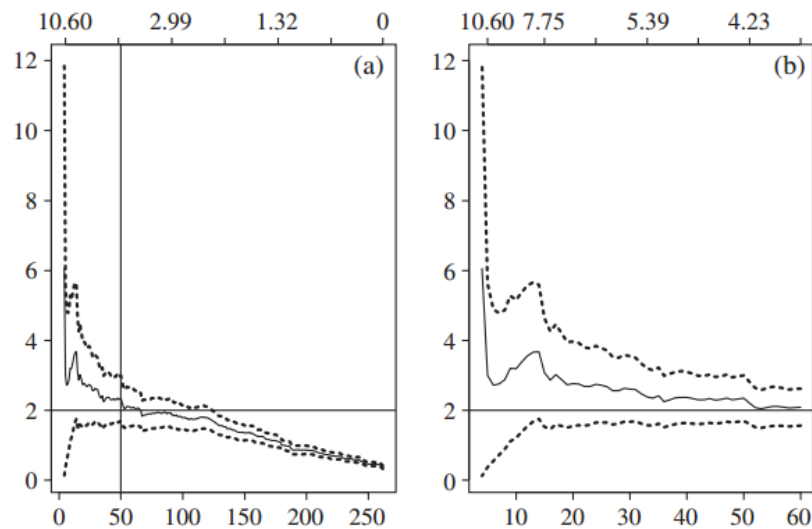
$$\hat{\alpha}_{k,n}^{(H)} = \left[\frac{1}{k} \sum_{j=1}^k \ln X_{(j)} - \ln X_{(k)} \right]^{-1} = \hat{\xi}^{-1} \text{ for } 2 \leq k \leq n$$

- Es posible construir gráficos con las estimaciones para $\hat{\alpha}$, en donde el eje vertical son los valores de $\hat{\alpha}$ (de los cuales se puede deducir $\hat{\xi}$) y el eje horizontal es la posición de los estadísticos de orden utilizados para el estimador (para un bloque de n observaciones)



- Este estimador es uno de los estimadores más estudiados en la teoría del valor extremo. Las propiedades asintóticas (consistencia y normalidad asintóticas) de este estimador se han investigado mucho para varios modelos de los datos, incluyendo los modelos ARCH y GARCH
- Cuando los datos provienen de una distribución con una distribución de cola que es casi una función potencial perfecta, el estimador de Hill suele ser un buen estimador de α o del recíproco de ξ
 - En la práctica, la estrategia general es usar los gráficos de Hill para poder encontrar una región estable (un intervalo del eje vertical) en donde las estimaciones construidas a partir de diferentes números de estadísticos de orden sean muy similares. En cuanto se localice la región estable, se tiene que mirar arriba el umbral en el que comienza esta región o un umbral cercano (parte de arriba de los gráficos, normalmente)

- La experiencia muestra que se tienen que coger 10-50 estadísticos de orden en una muestra de tamaño 1000, dado que considerar todos los valores de k posibles es engañoso



- La interpretación de los gráficos de Hill puede ser difícil: en la práctica, muchas desviaciones de la situación ideal pueden ocurrir. Si los datos no provienen de una distribución con una cola que varíe regularmente, el método de Hill no es realmente apropiado y los gráficos no son útiles
- Además, la dependencia serial de los datos puede reducir el rendimiento del estimador, aunque esto también es verdad para el estimador GPD
- Para las aplicaciones a la gestión del riesgo, lo más importante no es el índice de cola sino las estimaciones de la cola y de las medidas de riesgo, por lo que se puede hacer una justificación heurística para usar el estimador de cola estándar basándose en el enfoque de Hill
 - Se asume que un modelo de cola de la forma $1 - F(x) = Cx^{-\alpha}$ para $x \geq u > 0$ para un umbral alto u (se sustituye una función que varía lentamente por una constante para una x lo suficientemente grande)
 - Para un valor apropiado de k , el índice de cola α se estima con $\hat{\alpha}_{k,n}^{(H)}$ y el umbral u se reemplaza por $X_{k,n}$. Como C se puede escribir como $C = u^\alpha [1 - F(u)]$, esto es equivalente que estimar $1 - F(u)$, de modo que el estimador empírico obvio es k/n

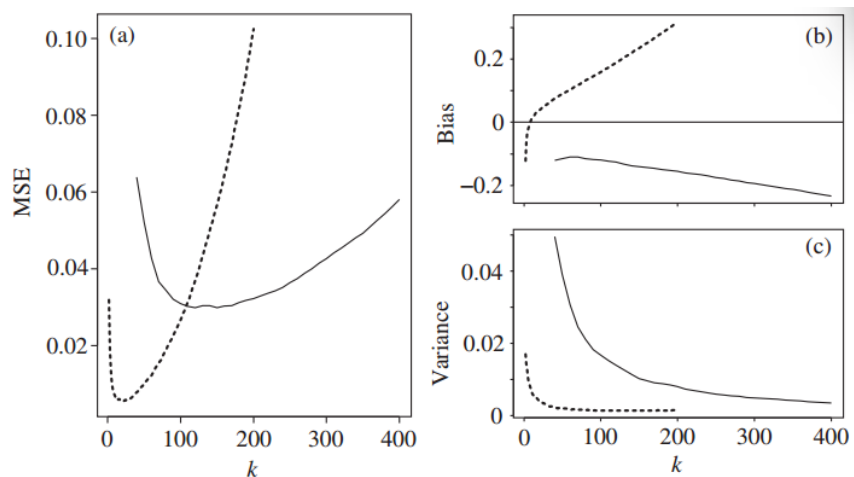
- Poniendo estas ideas juntas se puede obtener la forma estándar para el estimador de cola de Hill (definiendo $\bar{F}(x)$ como $1 - F(x)$)

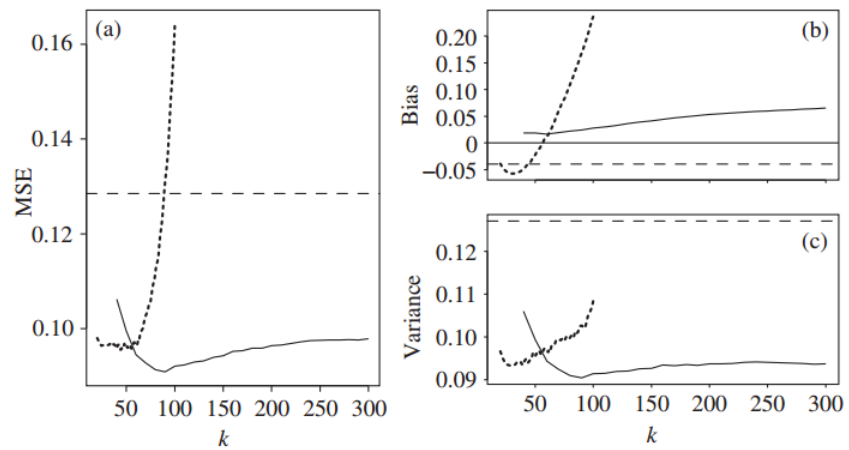
$$\hat{\bar{F}}(x) = \frac{k}{n} \left(\frac{x}{X_{k,n}} \right)^{-\hat{\alpha}_{k,n}^{(H)}} \quad \text{for } x \geq X_{k,n}$$

- Para cada par k y n , tanto el estimador de Hill como el estimador de cola asociado se tratan como función de los k estadísticos de orden de una muestra de tamaño n . Obviamente, es posible invertir este estimador para obtener un estimador cuartílico y también es posible crear un estimador para el *expected shortfall* usando argumentos sobre colas que varían regularmente
- El estimador basado en la GPD normalmente se trata como una función de un número aleatorio N_u de estadísticos de orden para un umbral fijo u . Suponiendo que se reescribe el estimador visto ahora con la notación del original para la GPD, se puede obtener el siguiente estimador (en donde falta el parámetro de escala β)

$$\hat{\bar{F}}(x) = \frac{k}{n} \left(1 + \hat{\xi}^{(H)} \frac{x - u}{\hat{\xi}^{(H)} u} \right)^{-1/\hat{\xi}^{(H)}} \quad \text{for } x \geq u$$

- Se pueden comparar ambos tipos de estimadores a través de la simulación del MSE, el sesgo y la varianza de parámetros como ξ y medidas del riesgo como el VaR
- Como la evaluación analítica del sesgo y de la varianza no es posible, se usan estimaciones de Monte Carlo, en donde la línea lisa es el estimador de la GPD y la de puntos es el de Hill





- Se puede observar como la robustez del estimador de cola basado en la GPD a cambios en k hace que escoger el umbral para este estimador no sea tan importante como para el método de Hill
- El método GPD permite obtener estimaciones de medidas de riesgo para distribuciones estacionarias (o incondicionales) cuando se aplica a excesos de umbrales en series temporales de rendimientos. Se puede considerar una adaptación simple de este método para poder obtener estimaciones para la distribución condicional (o predicción a un paso adelante) de las series
 - Se considera que X_{t-n+1}, \dots, X_t es una serie de rendimientos logarítmicos negativos (con el signo opuesto, de modo que son pérdidas) y se asume que estos provienen de una serie temporal estrictamente estacionaria $\{X_t\}$
 - Además, se asume que el proceso satisface las ecuaciones de la forma $X_t = \mu_t + \sigma_t Z_t$, donde μ_t y σ_t son \mathcal{F}_{t-1} -medibles y $\{Z_t\}$ son innovaciones independientes e idénticamente distribuidas con una función de distribución desconocida F_Z . Un ejemplo sería un modelo ARMA con errores GARCH
 - Se quiere obtener estimaciones del VaR y del *expected shortfall* para la distribución condicional $F_{X_{t+1}|\mathcal{F}_t}$, las cuales se puede obtener con las siguientes ecuaciones:

$$VaR_{\alpha}^t = \mu_{t+1} + \sigma_{t+1} q_{\alpha}(Z) \quad ES_{\alpha}^t = \mu_{t+1} + \sigma_{t+1} ES_{\alpha}(Z)$$

- Estas ecuaciones sugieren un método el cual consta de pocos pasos
 - Primero se ajusta un modelo ARMA-GARCH a través de cuasi-máxima verosimilitud (dado que no se quiere asumir una distribución particular de las innovaciones) y se usa para estimar

μ_{t+1} y σ_{t+1} . Como alternativa, se puede usar la predicción EWMA de la volatilidad

- Para estimar $q_\alpha(Z)$ y $ES_\alpha(Z)$ se aplica el procedimiento de estimación de la cola del GPD para la distribución de innovación F_Z
- Para poder superar el problema de que no se observan los datos de las innovaciones directamente, se trata a los residuos del análisis GARCH como los datos y se aplica el método de estimación de cola de la GPD a estos residuos. En particular, se estima $q_\alpha(Z)$ y $ES_\alpha(Z)$ usando las fórmulas vistas anteriormente para el modelo GPD

Las cópulas

- Toda función de distribución conjunta para un vector aleatorio de factores de riesgo implícitamente contiene una descripción de su comportamiento marginal (para cada uno de los factores de riesgo) y una descripción de su estructura de dependencia, y la cópula permite aislar esa descripción de la estructura de dependencia
 - Las cópulas permiten entender la dependencia a un nivel más profundo y facilitan la construcción de modelos multivariante con un enfoque *bottom-up*
 - Las cópulas permiten ver los peligros potenciales de los enfoques para la dependencia que solo se centran en la correlación y enseña como definir más medidas de dependencia alternativas. Además, expresan la dependencia en una escala cuartílica, que es útil para describir la dependencia de valores extremos
 - Normalmente se sabe más sobre el comportamiento marginal de los factores de riesgo que de su estructura de dependencia, de modo que el enfoque proporcionado por las cópulas permite combinar modelos marginales más avanzados con una variedad de posibles modelos de dependencia e investigar la sensibilidad del riesgo a su especificación de independencia
 - Con esta flexibilidad, las cópulas permiten, en teoría, construir un número ilimitado de modelos con distribuciones marginales dadas, aunque hace falta tener en cuenta el comportamiento de diferentes cópulas y su adecuación al contexto
 - Una cópula d -dimensional $C(\mathbf{u}) = C(u_1, u_2, \dots, u_d)$ es una función de distribución sobre $[0,1]^d$ con distribuciones marginales uniformes

estándar $U(0,1)$, por lo que es un mapeado $C: [0,1]^d \rightarrow [0,1]$ (de un hipercubo a un intervalo unitario). Esta debe cumplir tres propiedades:

- Se cumple que $C(u_1, u_2, \dots, u_d) = 0$ si $u_i = 0$ para al menos una i y que $C(1, 1, \dots, u_i, \dots, 1, 1) = u_i$ para toda $i \in \{1, 2, \dots, d\}$ y $u_i \in [0, 1]$. Esta segunda propiedad corresponde al requerimiento de que las distribuciones marginales sean uniformes
- Se cumple que para toda $(a_1, a_2, \dots, a_d) \in [0, 1]^d$ y $(b_1, b_2, \dots, b_d) \in [0, 1]^d$ con $a_i \leq b_i$, se obtiene la siguiente desigualdad:

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{\sum_{r=1}^d i_r} C(u_{1i_1}, u_{1i_2}, \dots, u_{1i_d}) \geq 0$$

where $u_{j1} = a_j$ & $u_{j2} = b_j$ for $\forall i \in \{1, 2, \dots, d\}$

- La desigualdad rectangular es la expresión anterior, y asegura de que si el vector aleatorio $(U_1, U_2, \dots, U_d)'$ tiene una función de distribución C , entonces $P(a_1 \leq U_1 \leq b_1, \dots, a_d \leq U_d \leq b_d)$ no es negativa
 - Para $2 \leq k < d$, las distribuciones marginales k -dimensionales de una cópula d -dimensional también son cópulas
- Al trabajar con cópulas, uno tiene que estar familiarizado con las operaciones de transformaciones de probabilidad y de cuartil y a las propiedades de inversas generalizadas
- Siendo F una función de distribución y F^\leftarrow su inversa generalizada (la función $F^\leftarrow(u) \equiv \inf\{x : F(x) \geq u\}$), si $U \sim U(0,1)$, entonces $P(F^\leftarrow(U) \geq x) = F(x)$. Esto se puede demostrar fácilmente de la siguiente manera:

$$F(x) \geq u \Leftrightarrow F^\leftarrow(u) \leq x$$

$$\Rightarrow P(F^\leftarrow(U) \geq x) = P(F(x) \geq U) = F(x)$$

- Siendo F una función de distribución y F^\leftarrow su inversa generalizada, si X tiene una función de distribución F continua univariante, entonces $F(X) \sim U(0,1)$. Esto se puede demostrar fácilmente de la siguiente manera:

$$\begin{aligned} P(F(X) \leq u) &= P(F^\leftarrow(F(X)) \leq F^\leftarrow(u)) = \\ &= F(u) \circ F^\leftarrow(u) = u \end{aligned}$$

- La primera proposición es muy importante para la simulación estocástica, dado que si se puede generar una uniforme estándar U y la inversa de una función de distribución $F(x)$, entonces se puede obtener una muestra de esta función. Tomando ambas proposiciones, se puede ver que se pueden transformar los riesgos con cualquier función de distribución continua particular para obtener cualquier otra función de distribución
- La importancia de las cópulas en el estudio de las funciones de distribución multivariante se resume en el teorema de Sklar, el cual enseña que todas las funciones de distribución multivariante contienen cópulas, y que las cópulas se pueden usar conjuntamente con funciones de distribución univariantes para construir funciones de distribución multivariante

- Siendo F una distribución conjunta con funciones de distribución marginales univariantes F_1, F_2, \dots, F_d , existe una cópula $C: [0,1]^d \rightarrow [0,1]$ tal que, para toda x_1, x_2, \dots, x_d en $\mathbb{R} = [-\infty, \infty]$, se cumple la siguiente igualdad:

$$F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d))$$

- Si los márgenes son continuos, entonces la cópula C es única, y de otro modo, la cópula es determinada únicamente por el producto cartesiano del rango de cada una de las distribuciones $F_i = F(\mathbb{R})$ para toda $i \in \{1, 2, \dots, d\}$

$$\text{ran}(F_1) \times \text{ran}(F_2) \times \dots \times \text{ran}(F_d)$$

- Conversamente, si C es una cópula y F_1, F_2, \dots, F_d son funciones de distribución marginales univariantes, entonces la función $F(x_1, x_2, \dots, x_d)$ definida anteriormente es una función de distribución conjunta con distribuciones marginales univariantes F_1, F_2, \dots, F_d

$$C(u_1, u_2, \dots, u_d) = F[F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2), \dots, F_d^{\leftarrow}(u_d)]$$

- Las dos ecuaciones mostradas en este teorema son fundamentales para lidiar con cópulas. La primera muestra como las distribuciones conjuntas F se forman acoplando distribuciones marginales con cópulas C , mientras que la otra muestra como las cópulas expresan dependencia en una escala de cuartiles (ya que $C(u_1, u_2, \dots, u_d)$ es la probabilidad conjunta de que X_i sea menor que su cuantil u_i para $i = 1, 2, \dots, d$)

- La forma general del teorema de Sklar se puede demostrar usando la transformación distribucional en vez de la transformación de probabilidad
- El teorema de Sklar también sugiere que, en caso de márgenes continuos, es natural definir la noción de cópula de una distribución
 - Si el vector aleatorio \mathbf{X} tiene una función de distribución de probabilidad F con funciones de distribución marginales F_1, F_2, \dots, F_d , entonces la cópula de F (o de \mathbf{X}) es la función de distribución C de $(F_1(X_1), F_2(X_2), \dots, F_d(X_d))$
 - El concepto de cópula es un poco menos natural para distribuciones discretas multivariantes, lo cual es debido a que hay más de una cópula que se pueden usar para juntar los márgenes y así formar una función de distribución conjunta
- Una de las propiedades más importantes de la cópula de una distribución es su invarianza bajo transformaciones estrictamente crecientes de las funciones de distribución marginales
 - Por lo tanto, se puede interpretar la cópula de una distribución como una manera natural de representar la estructura de dependencia de esta distribución (sobre todo en el caso de márgenes continuos)
 - Siendo (X_1, X_2, \dots, X_d) un vector aleatorio con funciones de distribución marginales continuas y cópula C y T_1, T_2, \dots, T_d funciones estrictamente crecientes, entonces el vector aleatorio $(T_1(X_1), T_2(X_2), \dots, T_d(X_d))$ también tiene cópula C
- Las distribuciones multivariantes y las cópulas tienen una cota superior e inferior, llamadas cotas de Fréchet. Para una función de distribución multivariante F con funciones de distribución marginales F_1, F_2, \dots, F_d , se tienen las siguientes cotas:

$$\max\left(1 - d + \sum_{i=1}^d F_i(x_i), 0\right) \leq F(\mathbf{x}) \leq \min(F_1(x_1), F_2(x_2), \dots, F_d(x_d))$$

- Por lo tanto, para cualquier cópula $C(u_1, u_2, \dots, u_d)$, se tienen las siguientes cotas:

$$\max\left(1 - d + \sum_{i=1}^d u_i, 0\right) \leq C(\mathbf{u}) \leq \min(u_1, u_2, \dots, u_d)$$

- La segunda desigualdad proviene del hecho de que, para cualquier i , se da la siguiente condición:

$$\bigcap_{1 \leq j \leq d} \{U_j \leq u_j\} \subset \{U_i \leq u_i\} \quad \text{for } \forall i = 1, 2, \dots, d$$

$$\Rightarrow C(\mathbf{u}) = P\left(\bigcap_{1 \leq j \leq d} \{U_j \leq u_j\}\right) \leq \min(u_1, u_2, \dots, u_d)$$

- Para la primera desigualdad, se puede obtener la siguiente desigualdad:

$$\begin{aligned} C(\mathbf{u}) &= P\left(\bigcap_{1 \leq i \leq d} \{U_i \leq u_i\}\right) = 1 - P\left(\bigcup_{1 \leq i \leq d} \{U_i > u_i\}\right) \geq \\ &1 - \sum_{i=1}^d P(U_i > u_i) = 1 - \left(d - \sum_{i=1}^d u_i\right) = 1 - d + \sum_{i=1}^d u_i \end{aligned}$$

- Existen varios ejemplos de cópulas, las cuales se pueden dividir en tres tipos: las cópulas fundamentales, las cópulas implícitas y las cópulas explícitas
 - Las características de los tres tipos de cópula mencionados son las siguientes:
 - Las cópulas fundamentales son estructuras de dependencia especiales importantes
 - Las cópulas implícitas son extraídas de distribuciones multivariante conocidas a través del teorema de Sklar, pero que no necesariamente tienen una expresión cerrada simple
 - Las cópulas explícitas son cópulas que posiblemente tienen una expresión cerrada simple y siguen construcciones matemáticas conocidas que permiten obtener cópulas
 - Las cópulas implícitas y explícitas no son categorías mutuamente exclusivas, y hay cópulas que pueden tener una representación implícita y explícita
 - Las cópula de independencia es una cópula fundamental que tiene la siguiente forma:

$$\Pi(u_1, u_2, \dots, u_d) = \prod_{i=1}^d u_i$$

- A partir del teorema de Sklar, se puede ver que las variables aleatorias con distribuciones continuas son independientes si, y solo si, su estructura de dependencia se da por $\Pi(u_1, u_2, \dots, u_d)$
- Las cópula de conmonotonicidad es una cópula fundamental que tiene la forma de la cota superior de Fréchet:

$$M(u_1, u_2, \dots, u_d) = \min(u_1, u_2, \dots, u_d)$$

- Esta cópula especial es la función de distribución conjunta del vector aleatorio (U, U, \dots, U) , donde $U \sim U(0,1)$. Esta representa una relación positiva perfecta
- Suponiendo que las variables aleatorias X_1, X_2, \dots, X_d tienen una función de distribución continua y son positiva y perfectamente dependientes, en el sentido de que son funciones estrictamente crecientes casi seguro unas de las otras, de modo que $X_i = T(X_1)$ casi seguro para $i = 2, 3, \dots, d$
- Por lo tanto, la cópula (X_1, X_2, \dots, X_d) y la cópula (X_1, X_1, \dots, X_1) son equivalentes. No obstante, la cópula (X_1, X_1, \dots, X_1) es la función de distribución de (U, U, \dots, U) , donde $U = F_1(X_1)$, que es la cópula $\min(u_1, u_2, \dots, u_d)$
- Las cópula de contramonotonicidad es una cópula fundamental que tiene la forma de la cota inferior bidimensional de Fréchet:

$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$$

- La cópula es la función de distribución conjunta de del vector aleatorio $(U, 1 - U)$, donde $U \sim U(0,1)$. Esta representa una relación negativa perfecta
- Si X_1 y X_2 tienen funciones de distribución continuas y son negativa y perfectamente dependientes, en el sentido de que son funciones estrictamente decrecientes casi seguro unas de las otras, de modo que $X_i = T(X_1)$ casi seguro para $i = 2, 3, \dots, d$
- La extensión a dimensiones mayor a dos para esta cópula no es posible
- La cópula gaussiana es una cópula implícita para el cual el vector aleatorio sigue una distribución normal multivariante
- Si $\mathbf{Y} \sim N_d(\boldsymbol{\mu}, \Sigma)$, la operación de estandarizar los márgenes se basa en aplicar una serie de transformaciones estrictamente

crecientes, y la proposición vista anteriormente sobre las cópula de un vector aleatorio transformado implica que la cópula de \mathbf{Y} es exactamente la misma que la cópula de $\mathbf{X} \sim N_d(\mathbf{0}, P)$, donde $P = \rho(\Sigma)$ es la matriz de correlación de \mathbf{Y}

- Por definición, la cópula viene dada por la siguiente expresión, donde Φ es la función de distribución normal estándar y Φ_P denota la función de distribución conjunta de \mathbf{X} :

$$\begin{aligned} C_P^{Ga}(\mathbf{u}) &= P(\Phi(X_1) \leq u_1, \dots, \Phi(X_d) \leq u_d) = \\ &= \Phi_P(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)) \end{aligned}$$

- La notación C_P^{Ga} enfatiza que la cópula está parametrizada por los parámetros de la matriz de correlación $d(d-1)/2$. En dos dimensiones se escribe C_ρ^{Ga} , donde $\rho = \rho(X_1, X_2)$
- La cópula de Gauss no tiene una forma cerrada simple, pero se puede expresar como la integral sobre la densidad de \mathbf{X} . En dos dimensiones para $|\rho| < 1$, la expresión es la siguiente:

$$\begin{aligned} C_\rho^{Ga}(u_1, u_2) &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{\frac{2\rho s_1 s_2 - s_1^2 - s_2^2}{2(1-\rho^2)}} ds_1 ds_2 \\ c_\rho^{Ga}(u_1, u_2) &= \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{\frac{2\rho s_1 s_2 - s_1^2 - s_2^2}{2(1-\rho^2)}} \end{aligned}$$

- Las cópulas de independencia y de conmonotonidad son casos especiales de la cópula de Gauss. Si $P = I_d$, se obtiene la cópula de independencia, si $P = J_d$ (una matriz $d \times d$ consistiendo de unos) entonces se obtiene la cópula de conmonotonidad. Además, para $d = 2$ y $\rho = -1$ la cópula de Gauss es igual es la contramonotonidad
- Por lo tanto, en dos dimensiones, la cópula de Gauss se puede entender como la estructura de dependencia que interpola entre correlación positiva perfecta y negativa perfecta, donde el parámetro ρ representa la fuerza de la dependencia
- La t -cópula d -dimensional es una cópula implícita para el cual el vector aleatorio sigue una distribución t -Student multivariante estándar con ν grados de libertad

$$C_{\nu, P}^t(\mathbf{u}) = \mathbf{t}_{\nu, P}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \dots, t_\nu^{-1}(u_d))$$

- En dos dimensiones, la expresión de la función de densidad de esta cópula es la siguiente:

$$c_{\nu, P}^t(u_1, u_2) = \frac{\Gamma\left(\frac{\nu+2}{\nu}\right)}{\Gamma\left(\frac{\nu+2}{\nu}\right) 2\pi(1-\rho^2)^{\frac{1}{2}}} e^{1 + \frac{2\rho s_1 s_2 - s_1^2 - s_2^2}{\nu(1-\rho^2)}}$$

- En este caso, t_ν es la función de distribución de una distribución t -Student estándar con ν grados de libertad, mientras que $\mathbf{t}_{\nu, P}$ es la función de distribución conjunta del vector $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$ y P es la matriz de correlaciones
 - Igual que en el caso de la cópula Gaussiana, si $P = J_d$ entonces se obtiene la cópula de conmonotonicidad. No obstante, si $P = I_d$, no se obtiene la cópula de independencia (ya que $\nu < \infty$) porque distribuciones t -Student multivariante no correlacionadas no son independientes unas de otras
- La cópula de Gumbel es una cópula explícita la cual tiene la siguiente forma cerrada:

$$C_\theta^{Gu}(\mathbf{u}) = e^{-((- \ln u_1)^\theta + (- \ln u_2)^\theta)^{\frac{1}{\theta}}} \text{ for } 1 \leq \theta < \infty$$

- Si $\theta = 1$, se puede obtener la cópula de independencia como un caso especial, y el límite de $C_\theta^{Gu}(\mathbf{u})$ cuando $\theta \rightarrow \infty$ es la cópula de conmonotonicidad bidimensional. Por lo tanto, la cópula de Gumbel permite interpolar entre la independencia y la dependencia perfecta positiva, siendo θ la representación de la fuerza de la dependencia
- La función de densidad en el caso bidimensional de la cópula de Gumbel es la siguiente:

$$c_\theta^{Gu}(u_1, u_2) =$$

$$C_\theta^{Gu}(u_1, u_2) u_1^{-1} u_2^{-1} \left(L^{-2 + \frac{2}{\theta}} \right) (\ln u_1 \ln u_2)^{\theta-1} \left[1 - (\theta + 1) L^{-\frac{1}{\theta}} \right]$$

$$\text{where } L \equiv (- \ln u_1)^\theta + (- \ln u_2)^\theta$$

- La cópula de Clayton es una cópula explícita la cual tiene la siguiente forma cerrada:

$$C_\theta^{Cl}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}} \text{ where } 0 \leq \theta < \infty$$

- El caso en el que $\theta \rightarrow 0$, la cópula de Clayton tiende a la cópula de independencia, mientras que si $\theta \rightarrow \infty$, la cópula tiende a la cópula de conmonotonidad
- La función de densidad en el caso bidimensional de la cópula de Clayton es la siguiente:

$$c_{\theta}^{Gu} = \left(1 + \frac{1}{\theta}\right) R^{-\theta-2} [(1-u_1)(1-u_2)]^{-1-\frac{1}{\theta}}$$

$$\text{where } R \equiv (1-u_1)^{-\frac{1}{\theta}} + (1-u_2)^{-\frac{1}{\theta}} - 1$$

- La cópula de Frank es una cópula explícita la cual tiene la siguiente forma cerrada:

$$C_{\theta}^{Fr}(u_1, u_2) = -\frac{1}{\theta} \ln \left(1 - \frac{(1-e^{\theta u_1})(1-e^{\theta u_2})}{1-e^{\theta}} \right)$$

$$\text{where } \theta \in (-\infty, 0) \cup (0, \infty)$$

- Igual que en el caso de la cópula gaussiana y de la t -cópula, las cópulas de Frank son simétricas
- La función de densidad de la cópula de Frank se puede expresar de la siguiente manera:

$$c_{\theta}(u_1, u_2) = -\theta e^{\theta} \left(\frac{1 + e^{\theta(u_1+u_2)}}{e^{\theta u_1} e^{\theta u_2} + e^{2\theta}} \right)$$

Las cópulas: conceptos de dependencia y las medidas

- Es posible definir los conceptos de dependencia perfecta positiva y negativa usando las cópulas fundamentales vistas anteriormente, llamadas conmonotonidad y contramonotonidad, respectivamente
 - El concepto de conmonotonidad puede ser definido de varias maneras equivalentes, pero se puede dar una definición basada en cópulas y derivar representaciones alternativas que muestran como variables aleatorias conmonótonas pueden interpretarse como variables no diversificables
 - Las variables aleatorias X_1, X_2, \dots, X_d son conmonótonas si admiten como cópula la cota superior de Fréchet $M(u_1, u_2, \dots, u_d) = \min(u_1, u_2, \dots, u_d)$

- Las variables X_1, X_2, \dots, X_d son conomonótonas si, y solo si, se cumple que para alguna variable aleatoria Z y funciones crecientes v_1, v_2, \dots, v_d se cumple lo siguiente:

$$(X_1, X_2, \dots, X_d) \sim (v_1(Z), v_2(Z), \dots, v_d(Z))$$

- Lo que esta propiedad quiere decir es que la fuente de variación es la misma, de modo que la magnitud no se tiene en cuenta en esta noción de dependencia positiva, solo el “movimiento”
 - Este resultado demuestra que las variables conomonótonas son funciones monótonamente crecientes de una sola variable aleatoria, por lo que solo hay una única fuente de riesgo y las variables conomonótonas se mueven determinísticamente en base a esta fuente
 - Siendo X_1, X_2, \dots, X_d variables aleatorias con función de distribución continua, estas son conomonótonas si, y solo si, para todo par (i, j) se tiene que $X_j = T_{ij}(X_i)$ de manera casi segura para una transformación creciente T_{ij}
 - El resultado se basa en la demostración anterior, notando que que la variable aleatoria U puede tomarse como $F(X_i)$ para cualquier i
- Un resultado importante para este tipo de variables es la aditividad de la función de cuartiles. Además de la medida del riesgo VaR, la propiedad de aditividad conomonótona se aplica a una clase de medidas del riesgo llamadas medidas de riesgo de distorsión, que incluye el *expected shortfall* y otras
 - Dos variables X_1 y X_2 son contramonótonas si tienen como cópula la cota inferior de Fréchet $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$

- Las variables aleatorias X_1 y X_2 son contramonótonas si, y solo si, para una variable Z y una función creciente v_1 y una decreciente v_2 , o viceversa, se cumple lo siguiente:

$$(X_1, X_2) \sim (v_1(Z), v_2(Z))$$

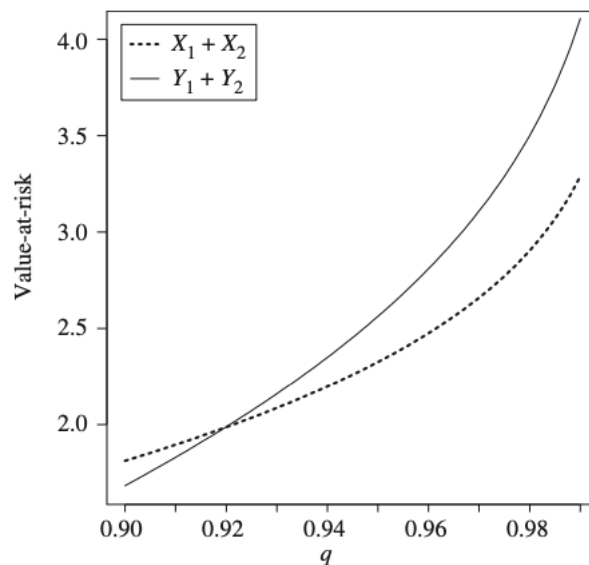
- Lo que esta propiedad quiere decir es que la fuente de variación es la misma, de modo que la magnitud no se tiene en cuenta en esta noción de dependencia negativa, solo el “movimiento”
- En el caso en el que X_1 y X_2 tengan funciones de distribución continuas, hay un resultado más simple para la contramonotonidad, expresando que esta es equivalente a

$X_2 = T(X_1)$ de manera casi segura para una función decreciente T

- El concepto de contramonotonidad no se generaliza a dimensiones más grandes, dado que la cota inferior de Fréchet no es una cópula en sí misma para $d > 2$ al no ser una función de distribución debida y no satisfacer la tercera propiedad de las cópulas
- La correlación juega un rol central en la teoría financiera, pero este concepto solo es realmente natural si se consideran distribuciones normales multivariantes o modelos elípticos. No obstante, no es natural en otras distribuciones y trae desventajas consigo que se pueden ilustrar a través de las cópulas
 - Como se ha visto, las distribuciones elípticas están descritas completamente por un vector de medias, una matriz de varianzas y covarianzas y una función característica
 - Como las medias y las varianzas son características de las distribuciones marginales, las cópulas de las distribuciones elípticas se pueden interpretar como distribuciones dependientes solo en la matriz de correlación y en la función característica. Esto hace que la matriz de correlaciones tenga un rol paramétrico en estos modelos (a diferencia de en otros modelos multivariante)
 - La correlación $\rho(X_1, X_2)$ entre las variables X_1 y X_2 es una medida lineal de dependencia que toma valores entre $[-1, 1]$. A esta medida se le pueden hacer algunas observaciones importantes como las siguientes:
 - Aunque variables independientes impliquen correlación nula, lo converso no es cierto, dado que hay distribuciones como las de mezcla de normal y las esféricas, en donde dos variables dependientes tienen correlación nula
 - Si $|\rho(X_1, X_2)| = 1$, entonces es equivalente a decir que X_1 y X_2 son perfectamente linealmente dependientes, de modo que se cumple que $X_2 = a + bX_1$ casi seguro para $a \in \mathbb{R}$ y $b \neq 0$, con $b > 0$ si la relación lineal es positiva y $b < 0$ si es negativa
 - Cuando $b_1, b_2 > 0$, la correlación es invariante bajo transformaciones lineales estrictamente crecientes. No obstante, esta no es invariante bajo transformaciones no lineales (aunque sean estrictamente crecientes)

$$\rho(a + b_1X_1, a + b_2X_2) = \rho(X_1, X_2)$$

- La correlación solo está definida cuando la varianza de cada variable es finita, pero no es una buena restricción para medidas de dependencia y puede causar problemas al trabajar con distribuciones con colas gruesas
- Existen varias falacias asociadas a la correlación lineal que ilustran los peligros de construir modelos de riesgo multivariantes de distribuciones marginales y estimaciones de las correlaciones entre riesgos. Estas solo son verdad para distribuciones elípticas, pero son falsas en general
 - La primera falacia es que las funciones de distribución marginal y las correlaciones a pares de un vector aleatorio determinan su función de distribución conjunta
 - Es posible construir diferentes cópulas tal que dos vectores aleatorios con función de distribución igual a esas cópulas tengan una misma correlación y el mismo tipo de distribución univariante (márgenes). En el siguiente gráfico es posible observar como la suma de los componentes de esos dos vectores pueden diferir en términos de cuartiles (y, por tanto, el VaR no sería el mismo para ambas sumas)



- La correlación de dos fuentes de riesgo no depende únicamente de su cópula, ya que, de ser así, la correlación sería invariante para transformaciones estrictamente crecientes (cuales sean) debido a que una cópula no se ve afectada por las transformaciones crecientes de sus componentes (lo cual es falso)
- La segunda falacia es que, para funciones de distribución univariantes F_1 y F_2 y cualquier correlación ρ en $[-1,1]$, siempre es posible construir una distribución conjunta F con márgenes F_1 y F_2 y correlación ρ

- Esto muestra que la correlación está relacionada intrínsecamente con las distribuciones marginales de los riesgos, y esto impone ciertas restricciones sobre los valores que puede tomar la correlación (por lo que no se puede)
- Las correlaciones asequibles pueden formar un subconjunto estricto del intervalo $[-1,1]$, lo cual se demuestra en el siguiente teorema
- Siendo (X_1, X_2) un vector aleatorio con funciones de distribución marginales F_1 y F_2 con varianza finita positiva y una distribución conjunta no especificada, se cumplen las siguientes proposiciones:
 - Las correlaciones asequibles forman un intervalo cerrado $[\rho_{min}, \rho_{max}]$ con $\rho_{min} < 0 < \rho_{max}$
 - La correlación mínima ρ_{min} se consigue si, y solo si, X_1 y X_2 son contramonótonas, mientras que la correlación máxima ρ_{max} se consigue si, y solo si, X_1 y X_2 son comonótonas
 - La correlación ρ_{min} es -1 si, y solo si, X_1 y $-X_2$ son del mismo tipo y $\rho_{max} = 1$ si, y solo si, X_1 y X_2 son del mismo tipo
- La tercera falacia es que, para variables aleatorias $X_1 \sim F_1$ y $X_2 \sim F_2$ y para una α dada, el cuartil de la suma $F_{X_1+X_2}^{\leftarrow}(\alpha)$ es maximizado cuando la distribución conjunta F consigue su correlación máxima
 - Cualquier ejemplo de superaditividad en la función cuartílica (o en la medida VaR) es un contraejemplo. Por ejemplo, si X_1 y X_2 son comonótonas, entonces $F_{Y_1}^{\leftarrow}(\alpha) + F_{Y_2}^{\leftarrow}(\alpha) = F_{Y_1+Y_2}^{\leftarrow}(\alpha)$ para $X_1 \sim Y_1$ y $X_2 \sim Y_2$, pero eso implicaría que se consigue correlación máxima, por lo que $F_{X_1+X_2}^{\leftarrow}$ no se maximiza cuando hay correlación perfecta

$$F_{X_1+X_2}^{\leftarrow}(\alpha) > F_{X_1}^{\leftarrow}(\alpha) + F_{X_2}^{\leftarrow}(\alpha) = F_{Y_1+Y_2}^{\leftarrow}(\alpha)$$

- Las correlaciones de rango son medidas escalares simples de dependencia que dependen solo de la cópula de una distribución bivalente y no de sus distribuciones marginales, a diferencia de la correlación lineal que depende de ambas
 - Los estimadores empíricos estándar de la correlación de rango se pueden calcular mirando los rangos de datos únicamente, de modo que solo se necesita el orden de la muestra para cada variables de interés y no los valores numéricos reales

- La razón principal para el uso de estas correlaciones es que se pueden usar para calibrar las cópulas a datos empíricos
- Hay dos variedades principales de correlación de rango: la de Kendall y la de Spearman, y las dos se pueden entender como una medida de concordancia entre vectores aleatorios bivariantes. Se dice que dos puntos en \mathbb{R}^2 son concordantes si se cumple que $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0$, y discordantes si $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0$
- Considerando un vector aleatorio (X_1, X_2) y una copia independiente $(\tilde{X}_1, \tilde{X}_2)$ (con la misma distribución, pero independiente), si X_2 tiende a crecer con X_1 , se espera que la probabilidad de concordancia sea relativamente más alta que la de discordancia, y lo opuesto en caso de que X_2 tienda a decrecer

- Esto motiva la correlación de rango de Kendall, que es la probabilidad de concordancia menos la probabilidad de discordancia de los pares

$$\rho_\tau(X_1, X_2) = P[(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0] - P[(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0]$$

- Para dos variables aleatorias X_1 y X_2 y la copia independiente $(\tilde{X}_1, \tilde{X}_2)$ del vector aleatorio (X_1, X_2) , la tau de Kendall se da por la siguiente expresión:

$$\rho_\tau(X_1, X_2) = E \left[\text{sign} \left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) \right) \right] = E[I_{>0} - I_{<0}]$$

- Para dimensiones más altas, la matriz de la tau de Kendall de un vector aleatorio \mathbf{X} se puede denotar por $\rho_\tau(\mathbf{X}) = \text{Cov}(\mathbf{Y})$, donde $\mathbf{Y} \equiv \text{sign}(\mathbf{X} - \tilde{\mathbf{X}})$ y esta se obtiene de aplicar la función signo en cada uno de los componentes de $\mathbf{X} - \tilde{\mathbf{X}}$
- Suponiendo que X_1 y X_2 tiene funciones de distribución marginales continuas y una única cópula C , entonces la tau de Kendall se puede obtener de la siguiente manera:

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1$$

- Para funciones de distribución continuas, la tau de Kendall solo depende de una única cópula C de (X_1, X_2)
- Esta expresión para la tau de Kendall muestra como esta correlación no depende de la magnitud (del valor de los cuartiles en la distribución) sino que solo depende del orden de estos, dado que las variables de las que dependen siguen una

distribución uniforme estándar (solo se miden de 0 a 1, lo que representa los cuartiles)

- La rho de Spearman también se puede definir en términos de la concordancia y discordancia de dos vectores aleatorios bivariantes, pero ahora se consideran pares de vectores independientes (X_1, X_2) y $(\tilde{X}_1, \tilde{X}_2)$ y se asume que tienen distribuciones marginales idénticas pero que el segundo par es tiene componentes independientes (de modo que \tilde{X}_1 y \tilde{X}_2 son variables independientes)

- Para variables aleatorias X_1 y X_2 y donde \tilde{X}_1 y \tilde{X}_2 son variables aleatorias satisfaciendo $\tilde{X}_1 \sim X_1$ y $\tilde{X}_2 \sim X_2$, ambas siendo independientes de (X_1, X_2) y entre ellas, la rho de Spearman se define de la siguiente manera:

$$\rho_S(X_1, X_2) = 3[P[(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0] - P[(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0]]$$

- Aunque parezca que la definición no permite obtener una medida en el intervalo $[-1, 1]$, una representación alternativa permite ver que sí la da. Si X_1 y X_2 son variables con distribuciones marginales continuas F_1 y F_2 , entonces se da la siguiente igualdad:

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2))$$

- La matriz de la rho de Spearman para un vector aleatorio general multivariante \mathbf{X} con márgenes continuos se da por $\rho_S(\mathbf{X}) = \rho(F_1(X_1), F_2(X_2), \dots, F_d(X_d))$ y tiene que ser semidefinida positiva
- La fórmula de la rho en términos de la cópula se puede obtener a través de la fórmula de Höffding
- Suponiendo que X_1 y X_2 tienen distribuciones marginales continuas y una única cópula C , entonces la rho de Spearman se puede obtener de la siguiente manera:

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2] du_1 du_2$$

- Igual que con la tau de Kendall, esta expresión muestra como esta correlación no depende de la magnitud (del valor de los cuartiles en la distribución) sino que solo depende del orden de estos, dado que las variables de las que dependen siguen una distribución uniforme estándar (solo se miden de 0 a 1, lo que representa los cuartiles)

- La tau de Kendall y la rho de Spearman tienen propiedades comunes, algunas de las cuales son las siguientes:
 - Ambas medidas son medidas de dependencia simétrica, tomando valores en el intervalo $[-1,1]$
 - Estas dan un valor nulo para variables independientes X_1 y X_2 , aunque no necesariamente una correlación nula implica independencia (igual que en el caso lineal)
 - Se puede demostrar que toman un valor de 1 solo si, y solo si, las variables son comonótonas y un valor de -1 si, y solo si, son contramonótonas
 - Estas son invariantes bajo transformaciones estrictamente crecientes (lineales o no)
 - La primera falacia sobre la determinación de la distribución conjunta por los márgenes y la correlación y la tercera sobre el valor máximo de la correlación siguen aplicando
 - La segunda falacia ya no aplica porque se puede formar una distribución bivalente tal que la correlación de rango esté en el intervalo deseado. Por ejemplo, haciendo una combinación convexa de la siguiente forma:

$$F(x_1, x_2) = \gamma W(F_1(x_1), F_2(x_2)) + (1 - \gamma)M(F_1(x_1), F_2(x_2))$$

$$\text{where } \gamma \in [0,1]$$

- Igual que las correlaciones de rango, los coeficientes de dependencia en la cola son medidas de dependencia que dependen solo de la cópula de una par de variables aleatorias X_1 y X_2 con distribuciones marginales continuas
 - La motivación de mirar estos coeficientes es que proporcionan medidas de dependencia extrema o medidas de la fuerza de la dependencia en las colas de la distribución bivalente
 - Los coeficientes que se describen se definen en términos de las probabilidades condicionales asintóticas de los excesos de cuartiles
 - Siendo X_1 y X_2 dos variables aleatorias con funciones de distribución F_1 y F_2 , el coeficiente de dependencia de la cola superior de X_1 y X_2 si el límite de $\lambda_u \in [0,1]$ existe y se define de la siguiente manera:

$$\lambda_u \equiv \lambda_u(X_1, X_2) = \lim_{q \rightarrow 1^-} P(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q))$$

- Si $\lambda_u \in (0,1]$, entonces X_1 y X_2 muestran dependencia de la cola superior o dependencia extrema en la cola superior, mientras que si $\lambda_u = 0$, las variables son asintóticamente independientes en la cola superior
- Si F_1 y F_2 son funciones de distribución continuas, entonces se puede obtener la siguiente expresión en términos de la cópula única C de la distribución bivalente:

$$\begin{aligned}\lambda_u &= \lim_{q \rightarrow 1^-} \frac{P(X_2 > F_2^{\leftarrow}(q), X_1 > F_1^{\leftarrow}(q))}{P(X_1 > F_1^{\leftarrow}(q))} = \lim_{q \rightarrow 1^-} \frac{\hat{C}(1-q, 1-q)}{1-q} \\ &= \lim_{q \rightarrow 0^+} \frac{\hat{C}(q, q)}{q} \quad \text{where } \hat{C} = \text{surv. copula of } C\end{aligned}$$

- Siendo X_1 y X_2 dos variables aleatorias con funciones de distribución F_1 y F_2 , el coeficiente de dependencia de la cola inferior de X_1 y X_2 si el límite de $\lambda_l \in [0,1]$ existe y se define de la siguiente manera:

$$\lambda_l \equiv \lambda_l(X_1, X_2) = \lim_{q \rightarrow 0^+} P(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q))$$

- Si $\lambda_l \in (0,1]$, entonces X_1 y X_2 muestran dependencia de la cola inferior o dependencia extrema en la cola inferior, mientras que si $\lambda_l = 0$, las variables son asintóticamente independientes en la cola inferior
- Si F_1 y F_2 son funciones de distribución continuas, entonces se puede obtener la siguiente expresión en términos de la cópula única C de la distribución bivalente:

$$\lambda_u = \lim_{q \rightarrow 0^+} \frac{P(X_2 \leq F_2^{\leftarrow}(q), X_1 \leq F_1^{\leftarrow}(q))}{P(X_1 \leq F_1^{\leftarrow}(q))} = \lim_{q \rightarrow 0^+} \frac{C(q, q)}{q}$$

- El cálculo de estos coeficientes es fácil para cópulas explícitas, tales como la cópula de Gumbel o la de Clayton, pero es un poco más difícil para otros tipos de cópulas

- La cópula gaussiana no tiene dependencia en las colas por construcción

$$\lambda_u = \lambda_l = 0$$

- La t -cópula puede tener dependencia en las colas, y al ser simétrica, se puede ver que se cumple la siguiente desigualdad:

$$\lambda_u = \lambda_l \geq 0$$

- La cópula de Gumbel muestra dependencia en la cola superior siempre que $\theta > 1$, y la fuerza de esta dependencia tiende a 1 cuando $\theta \rightarrow \infty$:

$$\lambda_u = 2 - 2^{\frac{1}{\theta}} \quad \lambda_l = 0$$

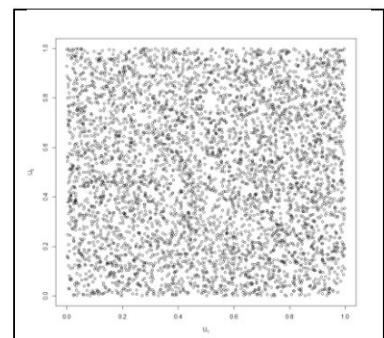
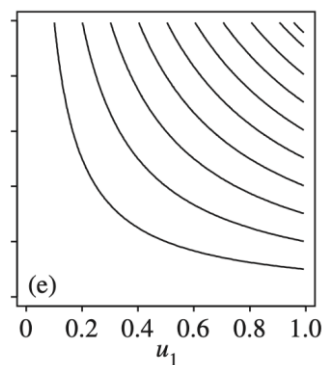
- La cópula de Clayton muestra dependencia en la cola inferior siempre que $\theta > 0$, y la fuerza de esta dependencia tiende a 1 cuando $\theta \rightarrow \infty$:

$$\lambda_u = 0 \quad \lambda_l = 2^{-\frac{1}{\theta}}$$

- La cópula de Frank se tiene un comportamiento de la dependencia en sus colas similar al de la gaussiana, pero tiene más dependencia en el centro
- A partir de los conceptos de dependencia y de las propiedades de las cópulas, se pueden interpretar visualmente las características de las cópulas vistas anteriormente a través de diversos gráficos

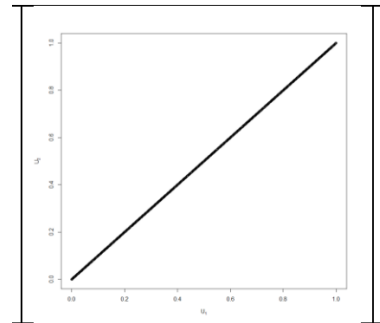
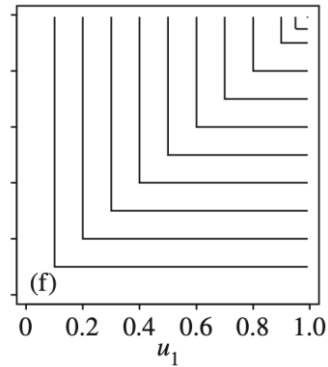
- Para las cópulas fundamentales se pueden obtener los siguientes gráficos ilustrativos:

- Para la cópula de independencia, se puede ver que, como ambas distribuciones son independientes, las esquinas de la función de densidad son muy curvas, indicando que mientras más curvas sean las esquinas, menos dependencia hay en la cola. Además, en el gráfico de puntos se puede ver como estos puntos se reparten por todo el gráfico, ya que no hay relación, lo cual indica que cuanto mayor dispersión de puntos en una zona, menos dependencia hay

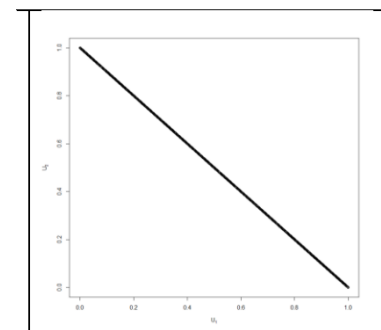
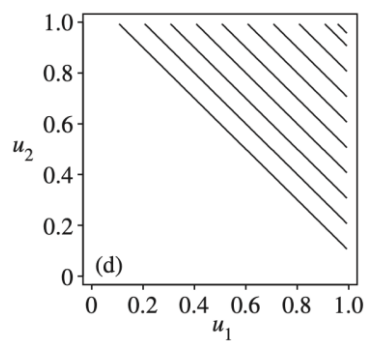


- Para la cópula de conmonotonidad, se puede ver como las esquinas de la función de densidad tienen ángulo recto (debido

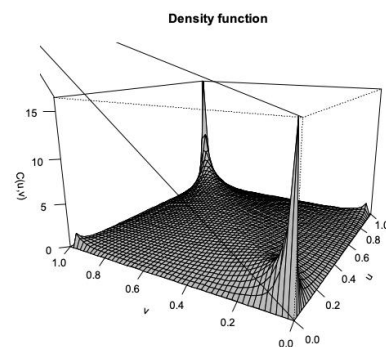
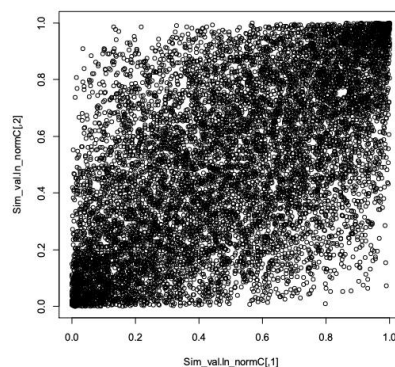
al final de la línea recta), indicando dependencia perfecta positiva en las colas. Además, en el gráfico de puntos se puede ver como estos puntos se reparten a lo largo de una línea recta con pendiente positiva

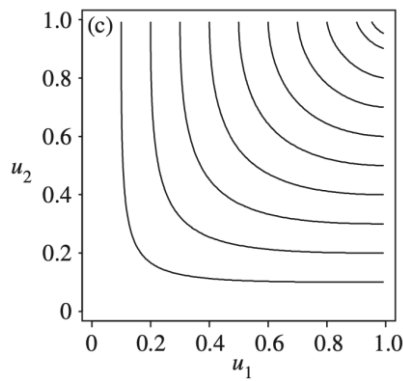


- Para la cópula de contramonotonidad, se puede ver como las esquinas de la función de densidad son rectas (debido a que son la misma línea), indicando dependencia perfecta negativa en las colas (en este caso, las colas son las otras esquinas). Además, en el gráfico de puntos se puede ver como estos puntos se reparten a lo largo de una línea recta con pendiente negativa

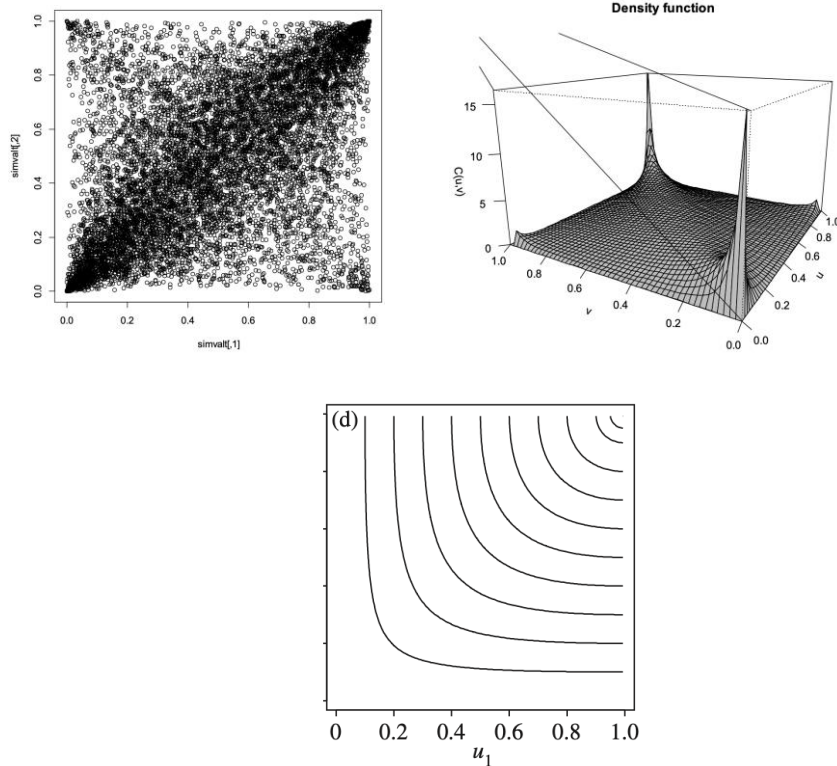


- Para la cópula gaussiana se pueden obtener los siguientes gráficos ilustrativos:





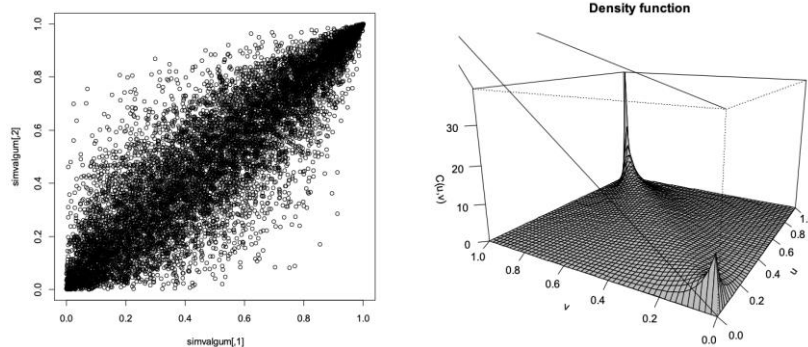
- En el gráfico de puntos, se puede ver como la cópula gaussiana tiene una densidad de puntos relativamente fuerte en las colas (porque su densidad tiende a infinito en esos puntos) pero también en el centro de la cópula
 - El gráfico de las esquinas de las colas muestra, sin embargo, que no hay dependencia en las colas de la cópula, dado que son curvas (como en la cópula de independencia)
 - Se puede apreciar que la cópula es simétrica
- Para la t -cópula se pueden obtener los siguientes gráficos ilustrativos:



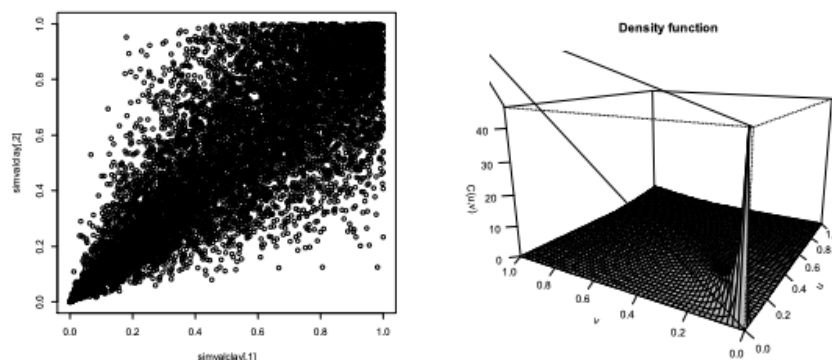
- En el gráfico de puntos, se puede ver como la t -cópula tiene una densidad de puntos relativamente fuerte en las colas (porque su densidad tiende a infinito en esos puntos) pero también en el

centro de la cópula y en las esquinas contrarias (esta última más pronunciada que en la normal)

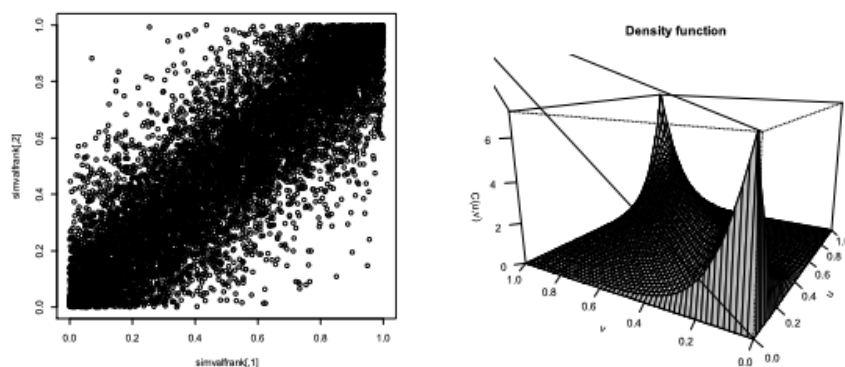
- El gráfico de las esquinas de las colas muestra un comportamiento menos curvo, de modo que hay dependencia en las colas de la cópula (se puede ver que los puntos se concentran mucho más, formando una punta)
 - Se puede apreciar que la cópula es simétrica
- Para la cópula de Gumbel se pueden obtener los siguientes gráficos ilustrativos:



- En el gráfico de puntos, se puede ver como la cópula de Gumbel tiene una densidad de puntos relativamente fuerte en las colas (porque su densidad tiende a infinito en esos puntos) y tiene un poco en el centro de la cópula también. No obstante, no tiene casi nada de densidad en las esquinas contrarias
 - En este caso se puede observar como la cópula de Gumbel exhibe dependencia en la cola derecha o superior, dado que los puntos forman una especie de punta allí, mientras que forman una curva en la otra cola (indicando poca dependencia)
- Para la cópula de Clayton se pueden obtener los siguientes gráficos ilustrativos:



- En el gráfico de puntos, se puede ver como la cópula de Clayton tiene una densidad de puntos relativamente fuerte en la cola izquierda o inferior, pero tiene muy poca en el centro y en la otra cola
- En este caso se puede observar como la cópula de Clayton exhibe dependencia en la cola izquierda o inferior, dado que los puntos forman una especie de punta allí, mientras que forman una curva en la otra cola (indicando poca dependencia)
- Para la cópula de Frank se pueden obtener los siguientes gráficos ilustrativos:



- En el gráfico de puntos, se puede ver como la cópula de Frank tiene una densidad de puntos relativamente fuerte en ambas colas, pero a diferencia de otras, exhibe también más densidad en el centro (más que la gaussiana incluso)
- En este caso se puede observar como la cópula de Frank exhibe menos dependencia en las colas como la gaussiana, dado que los puntos forman curvas pronunciadas allí, pero tienen más dependencia en valores centrales
- Se puede apreciar que la cópula es simétrica

Las cópulas: ajuste a los datos

- Suele ser difícil encontrar un modelo multivariante que describa efectivamente el comportamiento marginal y de estructura de dependencia adecuado, pero a través del ajuste de cópula es posible considerar dependencia en las colas y otros aspectos para construir modelos multivariante
 - Asumiendo que se tienen vectores de datos X_1, X_2, \dots, X_d describiendo pérdidas financieras o rendimientos de factores de riesgo financieros con función de distribución F idéntica, se escribe $X_t =$

$(X_{t,1}, X_{t,2}, \dots, X_{t,d})'$ para un vector de datos individuales y $X = (X_1, X_2, \dots, X_d)'$ para un vector aleatorio genérico con función de distribución F

- Además, se asume que F tiene márgenes continuos, de modo que por el teorema de Sklar, hay una única representación $F(x) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d))$
- Los métodos más importantes para poder ajustar cópulas a los datos son el método de momentos y el método de pseudo-máxima verosimilitud
- Siguiendo métodos estadísticos estándar, normalmente se ajustan modelos de cópula a los datos y se compara la calidad de los modelos ajustados utilizando diversos criterios
 - Los criterios más usados son el criterio de información bayesiana (BIC), el criterio de información de Akaike (AIC) y el estadístico de razón de verosimilitud (LRT)
- Dependiendo de la cópula particular que se quiere utilizar, es más fácil utilizar estimadores empíricos para la correlación de rango de Kendall o Spearman para así inferir una estimación del parámetro de la cópula, los cuales se pueden obtener a través del método de momentos
 - Para la rho de Spearman, los resultados anteriores sugieren que se puede estimar a través de calcular el coeficiente de correlación común para las pseudo-observaciones $\{(F_{i,n}(X_{t,i}), F_{j,n}(X_{t,j})) : t = 1, 2, \dots, n\}$, donde $F_{i,n}$ denota la función de distribución empírica para el margen i
 - En verdad se puede estimar $\rho_S(X_i, X_j)$ calculando el coeficiente de correlación de los rangos de los datos, lo cual coincide con el coeficiente de correlación cuando las observaciones no hay observaciones vinculadas (donde $X_{i,t} = X_{i,s}$ o $X_{j,t} = X_{j,s}$ para $t \neq s$)
 - El rango de $X_{i,t}$, denotado por $rank(X_{i,t})$ es simplemente la posición de $X_{i,t}$ en la muestra $X_{1,i}, X_{2,i}, \dots, X_{n,i}$ cuando las observaciones están ordenadas de menor a mayor. Si hay observaciones vinculadas, se les asigna un rango igual a la media de rangos que tendrían si se dividieran aleatoriamente
 - En el caso de no haber vinculaciones, la siguiente fórmula permite estimar la correlación de rango de Spearman:

$$r_{ij}^S = \frac{12}{n(n^2 - 1)} \sum_{t=1}^n \left(\text{rank}(X_{t,i}) - \frac{1}{2}(n+1) \right) \left(\left(\text{rank}(X_{t,j}) - \frac{1}{2}(n+1) \right) \right)$$

- Se denota $R^S = (r_{ij}^S)$ como la matriz de coeficientes de correlación de rango de Spearman a pares, la cual será semidefinida positiva
- El estimador estándar de la tau de Kendall es el coeficiente de correlación de rango de Kendall, definido por la siguiente fórmula:

$$r_{ij}^\tau = \binom{n}{2}^{-1} \sum_{i \leq t \leq s \leq n} \text{sign}((X_{t,i} - X_{s,i})(X_{t,j} - X_{s,j}))$$

- Se denota $R^\tau = (r_{ij}^\tau)$ como la matriz de coeficientes de correlación de rango de Kendall a pares, la cual será semidefinida positiva y se define de la siguiente manera:

$$R_{ij}^\tau = \binom{n}{2}^{-1} \sum_{i \leq t \leq s \leq n} \text{sign}((\mathbf{X}_t - \mathbf{X}_s)(\mathbf{X}_t - \mathbf{X}_s)')$$

- Estos coeficientes de correlación de rango muestrales se pueden usar para calibrar varias cópulas
 - Para cópulas en donde el coeficiente de correlación de Kendall o de Spearman tienen una forma cerrada, se puede aislar el parámetro con tal de obtenerlo en función del coeficiente, y así obtener una estimación paramétrica sustituyendo el coeficiente por su estimador muestral
- Debido a que las cópulas tienen una función de densidad asociada, es posible construir una función de verosimilitud y poder obtener estimadores de máxima verosimilitud utilizando pseudoóbservaciones, por lo que el método se suele llamar método de estimación por pseudo-máxima verosimilitud
 - Una de las ventajas de las cópulas es que permiten estimar independientemente el parámetro de la cópula y aquellos de los márgenes
 - Este método consiste en generar pseudo-observaciones $(\hat{U}_1, \dots, \hat{U}_n)$ que son obtenidas de las distribuciones marginales empíricas corregidas
 - Siendo C_θ una cópula paramétrica donde θ es el vector de parámetros a ser estimado, el estimador de máxima verosimilitud se obtiene maximizando la siguiente función con respecto al vector:

$$\ln L(\theta; \hat{U}_1, \dots, \hat{U}_n) = \sum_{t=1}^n \ln c_{\theta}(\hat{U}_t)$$

- La calidad estadística de los estimadores dependerá de la calidad de los estimadores de las distribuciones marginales usados en la formación de la pseudo-muestra de la cópula
- En la práctica, para implementar el método de máxima verosimilitud, es necesario derivar la densidad de la cópula
- Normalmente el estimador de máxima verosimilitud se encuentra a través de maximización numérica de la verosimilitud logarítmica resultante

El riesgo de mercado: medición del riesgo de mercado

- El método de varianzas y covarianzas utilizar las propiedades de la distribución normal de las pérdidas con tal de obtener las medidas de riesgo pertinentes
 - Asumiendo que los cambios en los factores de riesgo \mathbf{X}_{t+1} siguen una distribución multivariante normal $N_d(\boldsymbol{\mu}, \Sigma)$ y que la pérdida en $t + 1$ se puede aproximar linealmente con la función L_{t+1}^{Δ} , entonces L_{t+1}^{Δ} se distribuye normal

$$\mathbf{X}_{t+1} \sim N_d(\boldsymbol{\mu}, \Sigma) \quad L_{t+1}^{\Delta} = l_{[t]}^{\Delta}(\mathbf{X}_{t+1}) = -(\mathbf{c}_t + \mathbf{b}_t' \mathbf{X}_{t+1})$$

$$\Rightarrow L_{t+1}^{\Delta} \sim N(-(\mathbf{c}_t + \mathbf{b}_t' \boldsymbol{\mu}), \mathbf{b}_t' \Sigma \mathbf{b}_t)$$

- Esto se puede demostrar a través de las propiedades de la esperanza y de la covarianza:

$$E[-(\mathbf{c}_t + \mathbf{b}_t' \mathbf{X}_{t+1})] = -[\mathbf{c}_t + \mathbf{b}_t' E(\mathbf{X}_{t+1})] = -(\mathbf{c}_t + \mathbf{b}_t' \boldsymbol{\mu})$$

$$Cov[-(\mathbf{c}_t + \mathbf{b}_t' \mathbf{X}_{t+1})] = \mathbf{b}_t' Cov(\mathbf{X}_{t+1}) \mathbf{b}_t = \mathbf{b}_t' \Sigma \mathbf{b}_t$$

- Este método se puede utilizar con aproximaciones condicionales o aproximaciones incondicionales:
 - Con aproximaciones condicionales, cada predicción en $t + 1$ depende del comportamiento de las observaciones anteriores hasta t . Con estas aproximaciones lo que se hace es estimar $\boldsymbol{\mu}$ y Σ a partir de los datos históricos y usar las hipótesis de normalidad de L_{t+1}^{Δ} para calcular las medidas de riesgo
 - Con aproximaciones incondicionales, el comportamiento en $t + 1$ no depende de la realización en observaciones previas, sino

que en cada periodo el mismo fenómeno aleatorio sucede y es independiente de la realización del anterior. Con estas aproximaciones lo que se hace es predecir μ_{t+1} y Σ_{t+1} en el momento $t + 1$ a partir de los datos históricos y usar las hipótesis de normalidad de L_{t+1}^A para calcular las medidas de riesgo

- Las ventajas de este método son que es fácilmente aplicable, permite obtener una solución analítica para el VaR_α y el ES_α y se puede generalizar a una distribución t -Student
 - Sin embargo, se basa en dos hipótesis simplificadoras: la pérdida es una función lineal y se cumplen las hipótesis de normalidad. Además, puede subestimar la importancia de la cola (probabilidad de que se den valores extremos)
- En vez de suponer un modelo paramétrico explícito para el vector X_{t+1} , el método de simulación histórica funciona directamente a través de la función de pérdida L_s o de la función de pérdida linealizada L_s^A
 - En este caso, $s = t - T + i$ para $i = 1, 2, \dots, t$, de modo que se calcula una L_s^A para los T periodos previos

$$L_s^A = l_{[t]}^A(X_s) \quad \text{for } s = t - T + 1, 2, \dots, t$$

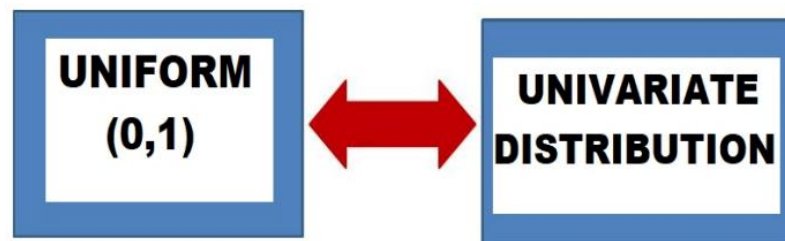
- Como se tienen datos retrospectivos de las pérdidas, estos se tratan como datos de sección transversal en el método incondicional y como una serie temporal en el método condicional
- Con las observaciones de las pérdidas L_s^A , el histograma o la distribución empírica puede ser analizada
- Los valores de L_s^A para $s = t - T + 1, 2, \dots, t$ se ordenan en orden ascendente según su magnitud, y estos valores se pueden denotar como $\tilde{L}_{n,n} \leq \tilde{L}_{n-1,n} \leq \dots \leq \tilde{L}_{2,n} \leq \tilde{L}_{1,n}$
 - El VaR_α para la distribución empírica es $VaR_\alpha = \tilde{L}_{[(1-\alpha)n],n}$, el cual es el cuartil de la distribución que está en la posición $(1 - \alpha)n$ dentro de las observaciones ordenadas (ya que las pérdidas están ordenadas en orden ascendente, y correspondería al cuartil α , dado que su valor sería más grande)
 - Para calcular el ES_α , se aplica la media aritmética de los valores $\tilde{L}_{j,n}$ que exceden el $VaR_\alpha(L)$

- A partir de los valores L_s^A para $s = t - T + 1, 2, \dots, t$, también se puede ajustar una distribución paramétrica y calcular el $VaR_\alpha(L)$ y el $ES_\alpha(L)$ utilizando los parámetros de la distribución ajustada
 - Las ventajas de este método son que es fácilmente aplicable, no se requiere suponer ninguna distribución multivariante para los cambios en los factores de riesgo, hay solución analítica para el $VaR_\alpha(L)$ y el $ES_\alpha(L)$ y se puede generalizar a distribuciones univariantes adecuadas para los extremos
 - No obstante, si hay pocas observaciones, el valor de T será muy limitado, y la diversificación no se tiene en cuenta en los resultados
- El método de simulación de Monte Carlo se basa en simular varios posibles escenarios, posiblemente interrelacionados entre ellos, a través de los datos empíricos y de la distribución ajustada de estos
 - El método de Monte Carlo puede ser condicional o incondicional, y este consiste de los siguientes pasos:
 - Se escoge un modelo de rendimiento histórico para los datos de factores de riesgo (X_{t-n+1}, \dots, X_t)
 - Se generan aleatoriamente m secuencias del comportamiento de los factores de riesgo en el siguiente periodo

$$(X_{t+1}^{(1)}, \dots, X_{t+1}^{(m)})$$
 - Se calculan los m valores de la función de pérdida utilizando las observaciones X_t y las simuladas

$$L_{t+1}^{(j)} = l_{[t]}(X_{t+1}^{(j)}): j = 1, \dots, m$$
 - Se mide el riesgo a través de la distribución empírica o ajustando una distribución paramétrica para medir el riesgo
 - El método de Monte Carlo permite obtener una solución directamente para calcular el VaR_α y ES_α
 - No obstante, el comportamiento estocástico de X_{t+1} es difícil de justificar; si la cartera es grande, entonces puede haber demasiados factores de riesgo a modelar; y no hay criterios definidos para fijar cuantas réplicas m se tienen que simular

- Uno de los tipos de simulación de Monte Carlo más comunes es utilizando una distribución normal multivariante, pero también se pueden utilizar otras distribuciones
 - Debido a que las funciones de distribución toman valores entre cero y uno, es posible simular valores en ese intervalo (simulando valores de una distribución uniforme u por otros métodos) y aplicar la inversa de la función de distribución de una normal univariante para obtener cuartiles de la distribución
 - En verdad, es posible obtener cuartiles (valores) de cualquier tipo de distribución univariante a través de la inversa de la función de distribución y de simulaciones de una distribución uniforme de cero a uno

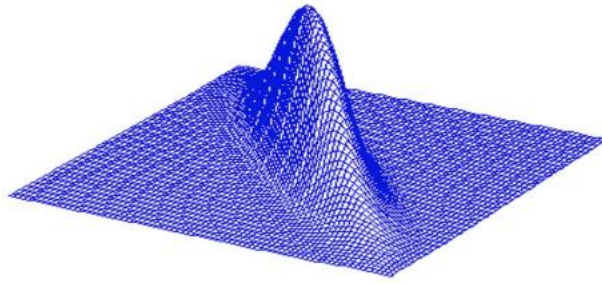


- La simulación de normales no correlacionadas univariantes permiten obtener normales correlacionadas a través de la descomposición de Cholesky
 - Generalmente, una matriz de varianzas y covarianzas se puede descomponer en LL' , en donde L es una matriz triangular

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \times \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}$$

- Encontrando esta matriz triangular, se puede utilizar en la función de densidad de una normal multivariante con tal de generar correlación entre ambas variables normales en su distribución conjunta



- En el caso bivalente, los pasos para obtener una distribución normal bivalente son sencillos:

- Primero se genera un vector aleatorio de variables normales independientes estándar (x_1, x_2)
- Este vector es multiplicado por la matriz triangular de la descomposición de Cholesky para obtener un nuevo vector

$$(y_1, y_2) = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} (x_1, x_2)$$

- Finalmente, este vector se desplaza con el vector de medias deseado y se obtiene así un vector que sigue una distribución normal bivalente estándar

$$(z_1, z_2) = (\mu_1, \mu_2) + (y_1, y_2)$$

- Finalmente, uno de los métodos para poder comprobar que las estimaciones del VaR_α y el ES_α son correctas, se puede utilizar un procedimiento de *backtesting*

- Este procedimiento también se puede utilizar para comprobar las medidas $VaR_\alpha^{(h)}$ y $ES_\alpha^{(h)}$

- En este caso, se puede demostrar que cuando se agregan datos las medidas son las siguientes si la esperanza es nula:

$$VaR_\alpha^{(h)} = \sqrt{h} VaR_\alpha \quad ES_\alpha^{(h)} = \sqrt{h} ES_\alpha$$

- Este procedimiento consiste en los siguientes pasos:

- Primero, se supone que en el momento t se han hecho estimaciones para el VaR_α , el ES_α , el $VaR_\alpha^{(h)}$ y el $ES_\alpha^{(h)}$
- Después de un periodo, es posible comparar los resultados observados con la estimación de un solo periodo

- Sucesivamente para los siguientes h periodos cuando las estimaciones son correctas, las pérdidas solo deberían sobrepasar el VaR_α en una proporción igual a $100(1 - \alpha)\%$. Si el porcentaje es mayor, la estimación no es buena
- Para poder comprobar el comportamiento del ES_α , se puede comprobar si se da la siguiente igualdad, la cual quiere decir que las pérdidas L_{t+1} deberían ser similares al ES_α para los periodos en los que se supera el VaR_α :

$$E(L_{t+1} - ES_\alpha | L_{t+1} > VaR_\alpha) = 0$$

El riesgo crediticio y los instrumentos crediticios

- El riesgo crediticio es el riesgo de pérdida que nace de la incapacidad de la contraparte a cumplir con sus obligaciones contractuales en tiempo y forma
 - Esta definición incluye los cuatro riesgos más importantes: el riesgo de fallida, el riesgo de rebaja, el riesgo de incumplimiento y el riesgo de liquidación
 - El riesgo de fallida es el riesgo de pérdidas debido a la fallida del prestatario o socio comercial. Se produce cuando se toma cualquier posesión como garantía del incumplimiento de la contrapartida, pues la liquidación no es suficiente para recuperar las pérdidas
 - El riesgo de rebaja es el riesgo de pérdidas debido a la deterioración de la calidad crediticia de la contraparte que se traduce en una rebaja en un sistema de calificación
 - El riesgo de incumplimiento se refiere a la falta de pago de intereses y/o principal de un préstamo por parte de la contrapartida al prestatario. El periodo de incumplimiento normalmente está limitado, entre 30-60 días en España
 - El riesgo de liquidación se produce cuando, a la fecha de vencimiento del contrato, una de las contrapartes está en posición de ganancia y la otra en una de pérdidas, haciendo que la última no pueda hacer frente a sus obligaciones de pago
 - El riesgo crediticio es omnipresente en la cartera de una institución financiera común
 - Las carteras de bonos de préstamos y corporativos están afectados por riesgo crediticio

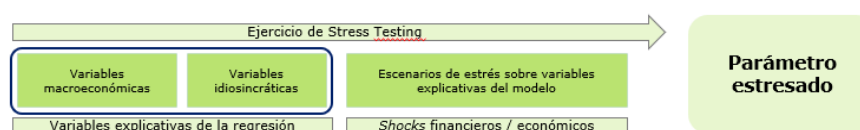
- Cualquier transacción de derivados financieros OTC está acompañada de riesgo crediticio (como en el caso de los *swaps*), dado que la fallida de una de las partes puede afectar significativamente la recompensa de la transacción. No obstante, existe un mercado especializado para derivados crediticios como los *credit default swaps* en los que las instituciones financieras son jugadores activos
 - Esto quiere decir que el riesgo crediticio está relacionado a las actividades básicas de los bancos y de las compañías aseguradoras, las cuales están expuestas sustancialmente al riesgo crediticio en sus carteras y a riesgo de fallida en sus renovaciones de seguros
- La gestión del riesgo crediticio en una institución financiera conlleva muchas actividades
 - Las instituciones necesitan determinar el capital que tiene que mantener para absorber pérdidas por riesgo crediticio (por motivos regulatorios y económicos)
 - También es necesario gestionar el riesgo en el balance de situación, de modo que se tienen que asegurar la diversificación de carteras con instrumentos arriesgados y que están optimizados acorde con el perfil de riesgo-rendimiento. El perfil de riesgo de la cartera se puede mejorar haciendo coberturas con derivados crediticios o transfiriendo el riesgo a los inversores a través de la titularización
 - Además, se tiene que gestionar la cartera de derivados crediticios comerciados, por lo que se tienen que realizar coberturas, valoraciones y gestiones de riesgo de colateral
 - Finalmente, se tiene que gestionar el riesgo crediticio de contraparte en sus negocios y contratos
- En la práctica de la gestión de riesgo de crédito, se dan importancia a aspectos más detallados de lo que anteriormente se ha discutido, ya que las instituciones financieras están sujetas a regulaciones estrictas
 - Tal y como se ha definido anteriormente, el riesgo de crédito es el riesgo ante la posibilidad de que una contrapartida incumpla sus obligaciones crediticias en tiempo y forma
 - En la práctica, el riesgo de crédito puede resultar de varios eventos posibles: incumplimiento de una obligación financiera, un incremento de la probabilidad de incumplimiento de una

obligación, de una pérdida más grave de la esperada debido a un incremento (o descenso) mayor del esperado de la exposición (o recuperación) al momento de incumplimiento, o el incumplimiento en el pago de bienes y servicios ya prestados

- No obstante, la definición de incumplimiento no es trivial e impacta de forma directa en la modelización del riesgo, por los que suelen haber proyectos de nuevas definiciones de fallida o *new definition of default* (NDoD)
 - Existen dos fuentes determinantes del riesgo crediticio: el riesgo sistémico, proveniente de la conyuntura económica, y el riesgo idiosincrático, proveniente de la calidad y el volumen de la contrapartida
- El rol de la gestión del riesgo es evaluar todos los riesgos a los que se enfrenta la institución, comunicar estos riesgos a los responsables de toma de decisiones y dar un seguimiento de estos riesgos
- El objetivo es, por tanto, reconocer que detrás de las posibles grandes pérdidas hay que desarrollar planes de contingencia para reducir su impacto en caso de que ocurran
 - La gestión y medición del riesgo se basa, entre otras cosas, en la definición en una política clara y unas metodologías de clasificación crediticia, recuperaciones y modelos internos de riesgo robustos
 - El proceso de gestión incluye 5 pasos: identificar el riesgo, cuantificar y estimar la exposición al riesgo, determinar los efectos de la exposición al riesgo, desarrollar una estrategia de mitigación al riesgo y evaluar el rendimiento de las medidas adoptadas e implementar las necesarias
- Las entidades de crédito están obligadas a cubrir sus pérdidas por razón de riesgo de crédito, sujetas a normativas diferentes dependiendo de si son esperadas o inesperadas
- Las pérdidas esperadas o *expected losses* (EL) se consideran provisiones, mientras que las pérdidas inesperadas o *unexpected losses* (UL) se consideran capital. Las EL se consideran un coste de negocio, mientras que las UL se consideran una medida de riesgo
 - Existen dos tipos de capitales, el capital regulatorio y el capital económico. El capital regulatorio son los recursos a mantener para cubrir las pérdidas inesperadas según la regulación (en

Basilea, el percentil 99,9% de la distribución de pérdidas), mientras que el capital económico son los recursos que se mantienen para cubrir pérdidas inesperadas según el apetito al riesgo (para percentiles mayores a 99,9%)

- Los modelos de *stress test* tienen múltiples funciones regulatorias y de gestión. Su objeto principal es dar sensibilidad del riesgo ante situaciones o escenarios económicos concretos. Estos miden el impacto de los escenarios en los parámetros, en las exposiciones y colaterales, teniendo en cuenta variables macroeconómicas e idiosincráticas



- La gestión del riesgo se divide en etapa de admisión y seguimiento

Rating/Scoring	Estimación EL (Provisiones)	Cálculo UL (Capital)	<i>Stress Test</i>
Admisión		Seguimiento	

- En la práctica, el modelo se refiere a un método, sistema, o enfoque cuantitativo que aplica teorías, técnicas o asunciones de tipo estadísticas, económicas, financieras o matemáticas para procesar datos y producir estimaciones cuantitativas
 - Las instituciones financieras utilizan modelos para la estimación y posterior gestión del riesgo de crédito de sus carteras de crédito
 - Aunque no hay documentos oficiales de ningún organismo regulador que detalle todas las características que debería cumplir cada modelo de riesgo de crédito, sí que existen algunos estándares que deberían cumplir todos ellos a fin de garantizar unas estimaciones representativas de la realidad y su buena gobernanza
 - Algunos de estos estándares son la integración en la gestión del riesgo (en los distintos procesos), la simplicidad (evitar complejidad sin mejoras evidentes), la robustez (cumplir suposiciones matemáticas para obtener estimaciones consistentes) y su documentación detallada y actualizada

- Una vez construido todo el modelo de estimación de parámetros de riesgo, este debe ser sometido a un proceso de validación que corrobore la robustez y adecuación del modelo a su finalidad
 - El marco de validación de los modelos de cualquier entidad debería estar basado en tres líneas de defensa independientes, donde el Banco de España fija 6 principales ámbitos sobre los que las líneas de defensa independientes deberían emitir su propia opinión: la metodología, la documentación, los datos usados, los aspectos cuantitativos, los aspectos cualitativos y el entorno tecnológico
 - La primera línea de defensa está formada por la propia unidad de desarrollo de modelos. Es primordial que todo analista someta a un ejercicio de validación y bondad de ajuste todos los modelos que diseña
 - La segunda línea de defensa está formada por la unidad de validación de modelos, la cual es responsable de supervisar y garantizar la validez de todos los modelos creados por la unidad de desarrollo de modelos
 - La tercera línea de defensa está formada por la unidad de auditoría interna, la cual es responsable de supervisar y garantizar la validez de los modelos creados por la primera línea y del trabajo de validación realizado por la segunda línea
- Existen varias posibles clasificaciones de los modelos atendiendo a diferentes características:
 - Los modelos se pueden clasificar según su propósito, en donde se pueden distinguir modelos de proyección, modelos de estimación y modelos de cálculo
 - Los modelos se pueden clasificar según el horizonte temporal con el que trabajan, habiendo una distinción entre modelos para un punto en el tiempo (o *point in time*), los de *downturn* (estrés económico), los de largo plazo (o *long run*) y los de *best estimate* (aquella estimación más probable)
 - Los modelos se pueden clasificar según el enfoque que usen, diferenciando entre modelos cuantitativos, cualitativos y de criterio experto
 - Los modelos se pueden clasificar según las técnicas que utilizan, habiendo diferentes clasificaciones como modelos de

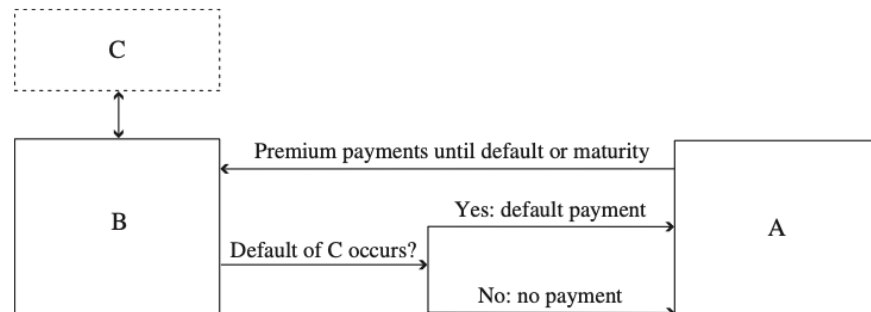
aprendizaje automático, de simulación, de series temporales, de regresiones, etc.

- Los instrumentos más importantes que conllevan riesgo de crédito son los préstamos, los bonos y derivados crediticios tales como *credit default swaps* y otros
 - Los préstamos son los instrumentos crediticios más antiguos y se pueden presentar de varias maneras. Su característica más importante de los préstamos es que una suma de dinero, conocida como principal, se da por adelantado al prestatario para un periodo concreto a cambio de una serie de pagos por intereses definido (que puede depender de los tipos de interés o puede fijarse) y de la devolución del principal al vencimiento
 - Estos se suelen categorizar acorde al tipo de prestatario, y se pueden clasificar en comerciales (particulares y compañías pequeñas o medianas), corporativos (compañías más grandes), interbancarios y soberanos (gobiernos). En cada categoría hay un número diferente de productos financieros
 - Se puede distinguir entre préstamos asegurados y los no asegurados. Los asegurados tienen activos colaterales comprometidos por parte del prestatario, de modo que si no se cumplen las obligaciones (fallida) el prestamista puede tomar posesión de los activos. Los no asegurados, en cambio, no tienen este reclamo sobre activos colaterales y la recuperación que se obtiene de la exposición (valor pendiente del principal y de los pagos de intereses) puede ser menor
 - A diferencia de los bonos, que son instrumentos comerciados públicamente, los préstamos son acuerdos privados entre un prestatario y un prestamista, por lo que hay una variedad de contratos con diferentes características legales. Esto hace difícil su valoración a través de principios de *fair-value*, por lo que se suele utilizar el valor nominal en el balance y enfoques de *fair-value* en el nivel 3
 - Los bonos son instrumentos públicamente comerciados emitidos por compañías y gobiernos que permiten al emisor recaudar fondos del mercado financiero
 - Los bonos emitidos por compañías se denominan bonos corporativos, mientras que los bonos emitidos por gobiernos se conocen como bonos del tesoro (o *treasuries*) o bonos soberanos

- La estructura de la recompensa es similar a la de los préstamos: el emisor del bono se compromete a hacer una serie de pagos de intereses a los compradores de los bonos y repagar el principal en el vencimiento
 - Estos pagos de intereses se denominan cupones, los cuales pueden ser fijos en la emisión del bono (bonos de cupones fijos) o variables dependiendo de tipos del mercado (bonos de interés variable). Normalmente se suele utilizar el LIBOR como referencia
 - Existen también bonos convertibles, los cuales permiten que el poseedor convierta el bono en acciones del emisor en determinados momentos del tiempo. Estos normalmente ofrecen intereses menores que los bonos corporativos porque se ofrece la oportunidad de participar en el crecimiento futuro de la compañía
 - El poseedor está sujeto al riesgo del tipo de interés, el riesgo de *spread* (riesgo de cambios en el *spread* crediticio) y el riesgo de fallida (riesgo de que no se paguen los intereses o el principal). Concretamente, el *spread* crediticio es la diferencia entre el rendimiento del bono y el rendimiento de un bono sin riesgo equivalente, por lo que un incremento en el *spread* hace que el valor de mercado del bono caiga (hay mayor riesgo de fallida)
- Una proporción significativa de transacciones con derivados se lleva a cabo en mercados OTC, y no hay una *clearing house* central que garantice el cumplimiento de las obligaciones contractuales. Esto hace que se esté sujeto al riesgo crediticio de contraparte (de que no cumpla con las obligaciones)
- Este tipo de riesgo fue muy importante durante la crisis financiera de 2008, dado que muchas instituciones que estuvieron involucradas en transacciones de derivados empeoraron su calidad crediticia o hicieron fallida
 - No obstante, la medición y la gestión del riesgo de contraparte es muy difícil. En el caso *swap*, las consecuencias de la fallida de una parte pueden variar dependiendo del subyacente, de modo que la exposición al riesgo de crédito no se puede saber *a priori* y no se sabe tampoco a quién pertenece la exposición
 - El riesgo de contraparte tiene que tenerse en cuenta en la valoración, por lo que se han desarrollado varias formas de ajustes crediticios del valor. Además, este riesgo se tiene que controlar utilizando técnicas de mitigación del riesgo tales como

la contratación con *netting* (el valor total de las transacciones es lo único sujeto al riesgo de contraparte, lo cual reduce el riesgo por la compensación entre transacciones) o la colateralización (se añade un colateral en el contrato como garantía, cuyo valor se ajusta dinámicamente con el cambio del valor de las transacciones)

- Los derivados crediticios son instrumentos principalmente usados para la cobertura y el comercio del riesgo crediticio. A diferencia de los otros productos, la recompensa prometida de un derivado crediticio está relacionada a un evento crediticio afectando a una o más compañías
 - Los mayores participantes en el mercado de los derivados crediticios son los bancos, las compañías de seguro y los fondos de inversión. Los bancos comerciales son típicamente compradores netos de protección contra eventos crediticios, mientras que otros como *hedge funds* y bancos de inversión actúan como compradores y vendedores
- Los *swaps* de fallida crediticia o *credit default swaps* (CDS) es un contrato entre dos partes: el comprador de la protección y el vendedor de la protección. Las recompensas de este activo crediticio están relacionadas a la fallida de una entidad de referencia



- Los CDS son muy importantes en el mercado de los derivados crediticios, y el mercado de los CDS escritos en referencia a corporaciones grandes es muy líquido
- Si la entidad de referencia experimenta un evento de fallida antes del vencimiento T de un contrato, el vendedor de protección realiza un pago de fallida al comprador de la protección, el cual imita la pérdida de un bono que es emitido por esa empresa de referencia (el activo de referencia). A esta parte de un CDS se le llama pierna de pago de fallida o *default payment leg*, y es la parte que se encarga de dar protección financiera al comprador contra la fallida

- Como compensación, el comprador de la protección hace pagos periódicos (normalmente trimestrales o semestrales) al vendedor de la protección (la pierna de pago de primas o *premium payment leg*). Después de la fallida de la empresa de referencia, el comprador deja de hacer los pagos
 - No hay pago inicial y los pagos están cotizados en la forma de un porcentaje anual x^* del valor nominal del activo de referencia, llamado diferencial del CDS o *CDS spread* (justo o según cotización de mercado)
 - En particular, las partes tienen que acordar la definición precisa del evento de fallida y del procedimiento para determinar el tamaño del pago de fallida en el caso de que ocurra en la entidad de referencia. Asociaciones como la ISDA han permitido que se estandaricen algunos términos
- Los inversores entran en el mercado de las CDS por varias razones, algunas de las cuales son las siguientes:
- Inversores en bonos con una gran exposición crediticia a la entidad de referencia pueden comprar CDS para asegurarse frente a pérdidas por un evento de fallida de la empresa, de modo que tienen motivos aseguradores. Esto es más fácil que reducir la posición en el bono debido a que los mercados de CDS son más líquidos que los de bonos corporativos y las posiciones en CDS se suelen pagar rápidamente
 - Se asume que los inversores con posiciones descubiertas o *naked* en CDS (posiciones en las que el comprador de la protección no es propietario del bono de la entidad de referencia) están especulando en el ensanchamiento del *credit spread* de la entidad de referencia. Estas posiciones son similares a las de vender bonos en corto, y los motivos especulativos son tan importantes en el mercado como los aseguradores
 - A diferencia del seguro, no se requiere que el comprador de la protección tenga intereses aseguradores (tener el bono de la entidad de referencia)
 - Los CDS son herramientas útiles de gestión del riesgo. Debido a la liquidez de los mercados CDS, estos son un instrumento natural subyacente para muchos derivados de riesgo complejo, y los modelos de valoración de carteras para estos derivados normalmente se calibran a los *spreads* de CDS cotizados

- Hay mucho debate sobre los riesgos del mercado de CDS, sobretodo con respecto al gran volumen de posiciones descubiertas y si estas deben ser limitadas o no
 - Tomando estas posiciones, los especuladores pueden hacer caer los precios de los bonos emitidos y hacer que la profecía se vuelva autocumplida
 - Un CDS se comercia en el mercado OTC y no se garantizan las transacciones por una *clearing house*. Por lo tanto, una posición puede estar sujeta a una cantidad sustancial de riesgo de contraparte, particularmente si la transacción no está respaldada por suficiente colateral
 - Hay preocupación de que los mercados de CDS hayan creado una nueva forma de dependencia entre instituciones financieras, de modo que la fallida de instituciones sistémicamente importantes cree una cascada de fallidas en el sector financiero debido al riesgo de contraparte
- Una nota vinculada a crédito es una combinación entre un derivado crediticio y un bono que paga cupones que se vende en un paquete fijo
 - Los pagos de cupones (y a veces el repago del principal) se reducen si una tercera parte (entidad de referencia) experimenta un evento de fallida durante la vida útil del contrato, de modo que el comprador de la nota proporciona protección crediticia de la nota
 - Desde el punto de vista legal, una nota vinculada a crédito se trata como una inversión de renta fija, por lo que los inversores que no pueden entrar en una transacción directamente con derivados crediticios (como compañías de seguros) pueden vender protección crediticia al comprar esas notas
 - Un inversor comprando una nota vinculada a crédito paga el precio primero, de modo que el comprador de la protección (el emisor) es protegido contra pérdidas causadas por la fallida del vendedor
- Independientemente del instrumento crediticio, el riesgo de pérdidas crediticias se ve afectado por tres cantidades generalmente relacionadas, llamadas parámetros de capital: la exposición en fallida, la probabilidad de fallida y la pérdida dada la fallida

- La exposición en fallida o *exposure at default* (EAD) es la pérdida potencial que la institución tiene en caso de fallida expresada en términos monetarios
 - Si se da un préstamo o se emite un bono, la exposición es relativamente fácil de determinar, dado que es el principal en juego. No obstante, hay más incertidumbre relacionada a los pagos de intereses o cupones y en productos como las líneas de crédito
 - Para derivados OTC, la exposición al riesgo de contraparte es más difícil de cuantificar, dado que es una variable estocástica dependiente del momento en el que la contraparte esté en situación de fallida y la evolución del valor del derivado llegado a ese punto. Sin embargo, el uso de colateral puede reducir el riesgo porque mitiga pérdidas
 - En la práctica, el concepto sirve para describir la exposición es la exposición en fallida, que reconoce que la exposición para varios instrumentos depende del momento exacto de fallida
- La probabilidad de fallida o *probability of default* (PD) es la probabilidad o la plausibilidad de que haya un incumplimiento en un instrumento como préstamos
 - Cuando se mide el riesgo de pérdidas en un horizonte temporal determinado, es necesario estimar la probabilidad de que los obligados estén en situación de fallida en el horizonte temporal, la cual es la probabilidad de fallida
 - Esta cantidad está relacionada a la calidad crediticia del obligado, hay diversos modelos para cuantificar esta calidad crediticia
 - Para instrumentos en donde la pérdida depende del momento en el que se hace fallida, el riesgo de fallida se describe por toda la distribución de probabilidad de posibles momentos de fallida (no solo por la probabilidad de fallida en un horizonte temporal dado)
- La pérdida dada la fallida o *loss given default* (LGD) es la magnitud de la pérdida probable en la exposición, expresado como un porcentaje de esta exposición
 - En el evento de fallida, es poco probable que toda la exposición se convierta en pérdidas (hay mitigadores directos del riesgo como garantías). En el caso de fallida con bonos, por ejemplo,

los acreedores recuperan parte de las pérdidas liquidando los activos de la empresa

- La pérdida dada la fallida es la proporción de la exposición (EAD) que se espera perder en el evento de fallida, o su converso, la recuperación (en porcentaje) a través de reestructuración de la deuda y la venta de activos
 - Normalmente, en el proceso de renegociación de la deuda, este porcentaje seguiría una distribución bimodal (con dos modas), dado que en la práctica se puede recuperar toda la exposición o no recuperar nada de esta
- Es importante ver como las tres cantidades son dependientes entre sí. Aunque es común intentar modelar las tres en términos de variables estocásticas independientes, no es realista hacerlo
- A cada individuo o entidad se le asocian los tres parámetros de capital
 - Las EL son proporcionales a estos tres parámetros, mientras que las UL son proporcionales a la LGD, la EAD y a una función de la PD

$$EL \propto PD * LGD * EAD$$

$$UL \propto f(PD) * LGD * EAD$$

- Además, estas tres cantidades son insumos clave para la fórmula de Basel para el enfoque basado en calificaciones internas o *internal-rating-based approach* (IRB) y los modelos estándar (*standard models* del Basel) usados para determinar los requerimientos de capital para carteras con riesgo crediticio
- Hay varias maneras de cuantificar la calidad crediticia o el riesgo de fallida de los obligados, pero estas se pueden dividir en dos filosofías: se puede hacer una descripción de la calidad crediticia a través de calificaciones o puntuaciones crediticias basadas en datos empíricos, o se pueden usar precios de mercado para inferir la calidad crediticia del obligado (si su capital se comercia públicamente)
- Las calificaciones (o *ratings*) o puntuaciones (o *scores*) crediticias tienen una función similar, dado que ambas medidas permiten ordenar a los obligados acorde a su nivel de riesgo crediticio y mapear ese riesgo para estimar la probabilidad de fallida

- Las calificaciones crediticias tienden a expresarse en términos en una escala categórica ordenada, mientras que las puntuaciones crediticias se expresan en términos de puntos en una escala métrica
 - Las puntuaciones crediticias normalmente se utilizan por clientes particulares y se basan en tablas de puntuaciones que el banco desarrolla a través de análisis extensivos de los datos históricos. La idea básica es que el riesgo de fallida se modele como una función de variables financieras, demográficas y de comportamiento que describen a los obligados, y se evalúa el binomio cliente-operación
 - Las puntuaciones se pueden clasificar en las puntuaciones crediticias de apertura o concesión (puntuación inicialmente dada a una entidad) y las puntuaciones crediticias de comportamiento (puntuación finalmente dada a una entidad basada en su comportamiento durante el periodo de tiempo concreto)
- En el enfoque de migración de crédito, cada empresa es asignada a una calificación crediticia en un punto en el tiempo y se especifica la probabilidad de moverse de una calificación crediticia a otra en un horizonte concreto (probabilidad de transición)
- Las probabilidades de transición se presentan típicamente en forma de matriz, las cuales se denominan matrices de transición y se estiman a partir de datos históricos y con métodos estadísticos estándar (máxima verosimilitud). En esta matriz, las filas indican el *rating* actual de las empresas o entidades y las columnas indican el *rating* a final de año

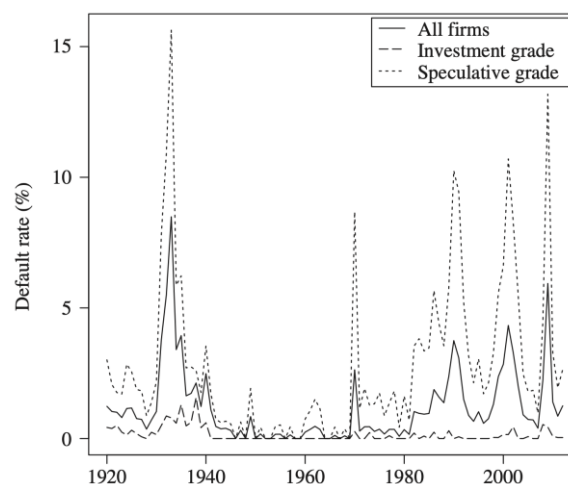
Initial rating	Rating at year-end (%)									WR
	Aaa	Aa	A	Baa	Ba	B	Caa	Ca-C	Default	
Aaa	87.20	8.20	0.63	0.00	0.03	0.00	0.00	0.00	0.00	3.93
Aa	0.91	84.57	8.43	0.49	0.06	0.02	0.01	0.00	0.02	5.48
A	0.06	2.48	86.07	5.47	0.57	0.11	0.03	0.00	0.06	5.13
Baa	0.039	0.17	4.11	84.84	4.05	7.55	1.63	0.02	0.17	5.65
Ba	0.01	0.05	0.35	5.52	75.75	7.22	0.58	0.07	1.06	9.39
B	0.01	0.03	0.11	0.32	4.58	73.53	5.81	0.59	3.85	11.16
Caa	0.01	0.02	0.02	0.12	0.38	8.70	61.71	3.72	13.34	12.00
Ca-C	0.00	0.00	0.00	0.00	0.40	2.03	9.38	35.46	37.93	14.80

- Como se puede ver, esta matriz no es cuadrada porque no tiene las probabilidades de la fila *default*. Esto se debe a que como las entidades en esta categoría son insolventes, entonces no es necesario hacer ninguna calificación crediticia

- En este enfoque se asume que la calificación crediticia actual determina completamente las probabilidades de falla, de modo que las probabilidades de falla se pueden leer directamente de la tabla
- Normalmente se realiza una corrección para tener en cuenta los retiros de calificaciones (transiciones al estado WR). La más sencilla sería dividir las probabilidades de los otros estados en cada fila entre 1 menos la probabilidad final en esa fila, lo cual asume implícitamente que el retiro de la clasificación no contiene ninguna información sobre la probabilidad de subida, rebaja o falla del obligado
- Las agencias de calificación crediticia también proporcionan una matriz de probabilidades de falla acumuladas para horizontes temporales más amplios (como si fuera una función de distribución de probabilidad para las filas de la matriz)

Initial rating	Term						
	1	2	3	4	5	10	15
Aaa	0.00	0.01	0.01	0.04	0.11	0.50	0.93
Aa	0.02	0.07	0.14	0.26	0.38	0.92	1.75
A	0.06	0.20	0.41	0.63	0.87	2.48	4.26
Baa	0.18	0.50	0.89	1.37	1.88	4.70	8.62
Ba	1.11	3.08	5.42	7.93	10.18	19.70	29.17
B	4.05	9.60	15.22	20.13	24.61	41.94	52.22
Caa-C	16.45	27.87	36.91	44.13	50.37	69.48	79.18

- Un hecho empírico conocido es que las tasas de incumplimiento tienden a variar con el estado de la economía, siendo altas en recesiones y bajas en expansiones



- Las tasas de transición se suelen estimar con datos de S&P o Moody's, pero también se pueden utilizar promedios históricos

de horizontes temporales más largos que contienen varios ciclos de negocio

- Las empresas están interesadas en la calidad crediticia promedio a través del ciclo cuando atribuyen *rating* crediticio a una empresa en particular. No obstante, las probabilidades de incumplimiento del enfoque de migración de crédito son estimaciones de la probabilidad de incumplimiento media, independiente del contexto económico actual
- Hay veces que uno está interesado en estimaciones de las probabilidades de incumplimiento en un punto en el tiempo de modo que se refleje el contexto macroeconómico actual (tales como la valoración de un préstamo a corto plazo). Es posible realizar ajustes a las probabilidades estimadas por el método anteriormente visto, siendo posible a través de usar los datos de las acciones como fuente de información alternativa (como en el modelo EDF)
- El cálculo de las probabilidades de migrar de una calificación crediticia a otra requeriría bastantes datos, pero a través de la suposición de que la serie temporal de las calificaciones crediticias cumplen la propiedad de Markov, es posible simplificar la obtención de estas probabilidades
 - Siendo $(R_t)_{t \geq 0}$ un proceso estocástico con tiempo discreto que toma valores en $S = 0, 1, 2, \dots, n$, siendo este un conjunto que representa los estados de calificación de la solvencia creciente (un valor mayor es una mejor calificación) y 0 es el *default*
 - Por lo tanto, $(R_t)_{t \geq 0}$ modela la evolución de las calificaciones crediticias de los deudores
 - Si se asume que $(R_t)_{t \geq 0}$ es una cadena de Markov, entonces las probabilidades condicionales de transición de un estado a otro cumplen la propiedad de Markov para toda $t \geq 1$ y para toda $j, r_0, r_1, \dots, r_{t-2}, k \in S$. De este modo, las probabilidades condicionales de transición solo dependen de su estado en $t - 1$ (no depende de su historia $0, 1, 2, \dots, t - 2$)
$$P(R_t = k | R_0 = r_0, R_1 = r_1, \dots, R_{t-1} = j) = P(R_t = k | R_{t-1} = j)$$
 - Esta suposición se ha criticado, dado que hay evidencia empírica sobre el efecto *momentum* y la pegajosidad en historiales de calificaciones crediticias. No obstante, se sigue usando debido a que conlleva modelos más prácticos en donde la teoría se entiende bien y permite utilizar estimadores naturales para las probabilidades de transición

- Además, si se sigue un proceso de Markov, entonces solo es necesario datos de un periodo para obtener las probabilidades de falla de los siguientes periodos
- Si la cadena de Markov es estacionaria, entonces quiere decir que la probabilidad de que $R_t = k$ dado $R_t = j$ debe ser la misma para toda $t \geq 1$ y para todos los estados de calificaciones j y k

$$P(R_t = k | R_{t-1} = j) = P(R_1 = k | R_0 = j)$$

- En este caso, se puede definir una matriz de transición $\mathbf{P} = (p_{jk})$ de tamaño $(n + 1) \times (n + 1)$, cuyos elementos son $p_{jk} = P(R_t = k | R_{t-1} = j)$ para cualquier $t \geq 1$
- De este modo, se pueden utilizar argumentos de probabilidad condicional básicos para obtener las ecuaciones de Chapman-Kolmogorov, las cuales dicen que para cualquier $t \geq 2$ y cualquier $j, k \in S$ se cumple la siguiente igualdad:

$$\begin{aligned} P(R_t = k | R_{t-2} = j) &= \\ &= \sum_{l \in S} P(R_t = k | R_{t-1} = l) P(R_{t-1} = l | R_{t-2} = j) = \sum_{l \in S} p_{lk} p_{jl} \end{aligned}$$

- Una implicación de estas ecuaciones es que la matriz de probabilidades de transición durante dos periodos de tiempo viene dada por $\mathbf{P}^2 = \mathbf{P} \times \mathbf{P}$. De este modo, la matriz de probabilidades de transición para T periodos de tiempo sería \mathbf{P}^T
- Para estimar \mathbf{P} , se supone que se observan o se obtiene información sobre la calificación crediticia de las compañías en los momentos $0, 1, \dots, T$. Esta información normalmente se relaciona con una cohorte de compañías fluctuante, en donde solo unas cuantas tienen un historial completo de calificaciones durante $[0, T]$ (algunas compañías se añaden, otras quiebran, otras no quieren calificación etc.)
 - En el caso en que las compañías no quieran una calificación crediticia, se asume que esto ocurre independientemente del riesgo de falla o de migración de calificación (lo cual puede no ser verdad)
 - En este caso, N_{tj} es el número de compañías que se han calificado como j en el momento t para $t = 0, 1, 2, \dots, T - 1$ y $j \in S - \{0\}$ y para las cuales su calificación está disponible en el momento $t + 1$. Además, N_{tjk} es el subconjunto de compañías que se han calificado como k en $t + 1$

- Bajo la suposición en tiempo discreto de la propiedad Markoviana, los experimentos multinomiales independientes se dan en cada momento t . En cada experimento, se puede interpretar que las N_{tj} compañías calificadas como j se asignan de manera aleatoria entre las calificaciones $k \in S$ acorde a probabilidades p_{jk} que satisfacen $\sum_{k=0}^n p_{jk} = 1$, de modo que la función de verosimilitud es la siguiente:

$$L((p_{jk}); (N_{tj}), (N_{tjk})) = \prod_{t=0}^{T-1} \left(\prod_{j=1}^n N_{tj}! \left(\prod_{k=0}^n \frac{p_{jk}^{N_{tjk}}}{N_{tjk}!} \right) \right)$$

- Utilizando el método de los multiplicadores de Lagrange para la *log-likelihood* restringida a $\sum_{k=0}^n p_{jk} = 1$, se obtiene que el estimador de máxima verosimilitud es el siguiente:

$$\hat{p}_{jk} = \frac{\sum_{t=0}^{T-1} N_{tjk}}{\sum_{t=0}^{T-1} N_{tj}}$$

- Una manera alternativa de poder estimar las probabilidades, más simple y directa, es a través de utilizar una función logarítmica de verosimilitud conceptualmente equivalente a la anteriormente vista
- Suponiendo que se da un punto de inicio arbitrario del proceso de Markov en $R_1 = r_1$, entonces se puede obtener la siguiente función logarítmica de verosimilitud a través de una generalización de un árbol de decisión para R :

$$L(\mathbf{p}|R_1, R_2, \dots, R_T) = \prod_{i=1}^T \prod_{j=1}^T p_{ij}^{n_{ij}}$$

- En este caso \mathbf{p} es el vector de probabilidades para cada entrada de la matriz \mathbf{P} , mientras que n_{ij} es el número de ocurrencias de la calificación del estado i seguido de la calificación del estado j (el número de empresas que caen en esa celda en la matriz de empresas con calificaciones crediticias)
- A partir de esta, se pueden obtener los estimadores de máxima verosimilitud a través de resolver por multiplicadores de Lagrange un problema de maximización restringido a que las probabilidades de cada fila en la matriz tienen que respetar $\sum_{k=0}^n p_{jk} = 1$

$$\sum_{j=1}^T p_{ij} = \sum_{j=1}^T \frac{n_{ij}}{\lambda_i} \Rightarrow \lambda_i = \sum_{j=1}^T n_{ij}$$

$$\Rightarrow \hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^T n_{ij}} \text{ for } i = 1, 2, \dots, k$$

- No obstante, el asumir un proceso de Markov discreto conlleva una serie de problemas, los cuales se pueden solucionar si se utiliza una cadena de Markov en tiempo continuo
 - En la práctica, los cambios de calificación tienden a darse en fechas en un espacio de tiempo irregular, se ignora la información de transiciones intermedias entre dos periodos t y $t + 1$ y se obtienen estimaciones no muy útiles
 - Aunque los datos se pueden aproximar a través de series (o datos de panel) con separación temporal regular, habrá una pérdida de información inevitable
 - Si un deudor hace una transición en un periodo de tiempo $[t, t + 1]$, este solo será registrado como una entidad que ha migrado de su calificación inicial en t a su final en $t + 1$, de modo que se ignora esta información sobre la transición intermedia
 - El procedimiento de estimación para una cadena en tiempo discreto tiende a resultar en estimaciones escasas o *sparse* para las matrices de transición, con muchos ceros como entradas. Sin embargo, eso no es realista, dado que se puede dar una transición de ese tipo (por muy extrema que sea) y la probabilidad no debería ser nula
 - Por lo tanto, es más satisfactorio modelar las transiciones de calificaciones como un fenómeno en tiempo continuo, y estas probabilidades de transición no se modelan de manera directa, sino que se dan en términos de tasas de transición. La relación entre las tasas de transición y las probabilidades de transición se puede explicar de la siguiente manera:
 - Se asume que sobre cualquier paso del tiempo pequeño de duración δt la probabilidad de transición de un *rating* j a k se da aproximadamente por $\lambda_{jk} \delta t$ para alguna constante $\lambda_{jk} > 0$, que es la tasa de transición entre el *rating* j y el k . La probabilidad de mantenerse en la calificación j se da por $1 - \sum_{k \neq j} \lambda_{jk} \delta t$

- Si se define la matriz Λ como una matriz con entradas λ_{jk} fuera de la diagonal y $-\sum_{k \neq j} \lambda_{jk}$ en la diagonal, se pueden resumir las probabilidades de transición implícitas para un paso del tiempo pequeño δt en la matriz $(I_{n+1} - \Lambda \delta t)$
- Se consideran transiciones en el intervalo $[0, t]$ y se denota la matriz de probabilidades de transición como $P(t)$. Si se divide el periodo de tiempo en N pasos de tiempo pequeños de medida $\delta t = t/N$ para una N lo suficientemente grande, la matriz de probabilidades de transición se puede aproximar por la siguiente matriz:

$$P(t) \approx \left(I_{n+1} + \frac{\Lambda t}{N} \right)^{t/N}$$

- Esta última matriz converge a la matriz exponencial de Λt cuando $N \rightarrow \infty$

$$\left(I_{n+1} + \frac{\Lambda t}{N} \right)^{t/N} \rightarrow e^{\Lambda t} \text{ as } N \rightarrow \infty$$

- Esta formulación permite obtener un método para calcular las probabilidades de transición para cualquier horizonte temporal t en términos de la matriz Λ , llamada matriz generadora

- Una cadena de Markov en tiempo continuo con matriz generadora Λ se puede construir de la siguiente manera: un deudor se mantiene en el estado j por un tiempo que se distribuye exponencialmente con un parámetro λ_{jj} equivalente al negativo del elemento diagonal de la matriz generadora

$$\lambda_{jj} = -\sum_{k \neq j} \lambda_{jk}$$

- Cuando se da una transición, la nueva calificación se determina por un experimento multinomial en el que la probabilidad de transición de un estado j a un estado k es $\lambda_{jk}/\lambda_{jj}$
- Esta construcción también permite obtener estimadores naturales para la matriz Λ . Como λ_{jk} es la tasa de migración instantánea de j a k , se puede estimar con la siguiente fórmula, en donde $N_{jk(t)}$ es el número total de transiciones observadas de j a k en el periodo $[0, T]$ e $Y_j(t)$ es el número de deudores con calificación j en el momento t :

$$\hat{\lambda}_{jk} = \frac{N_{jk}(t)}{\int_0^T Y_j(t) dt}$$

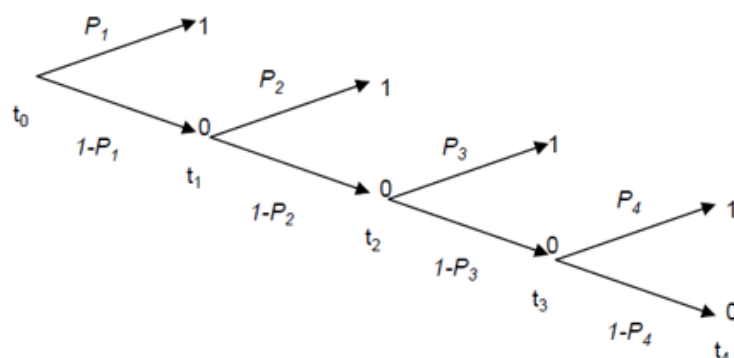
- En esta formulación, el denominador representa el tiempo total que todas las compañías de la base de datos han pasado en el estado j . Este estimador es un análogo en tiempo continuo del estimador de máxima verosimilitud visto anteriormente, por lo que se puede demostrar que este estimador es el de máxima verosimilitud para las intensidades de transición para una cadena de Markov homogénea en tiempo continuo

El riesgo crediticio en la práctica: la PD

- El evento fundamental a modelizar en la medición del riesgo de crédito es el evento de fallida, la cual se considera así si se encuentra en situación de gestión de mora o peor. A través de conceptualizar el evento como una variable de Bernoulli, es posible desarrollar modelos prácticos para la PD
 - La definición de fallida es acorde a los requerimientos de Basel II, que considera morosas a todas las operaciones incumplidas con más de 90 días, así como los clientes de los que se dude que puedan hacer frente a sus pagos (dudosos subjetivos)
- El evento de fallida es discreto, por lo que se puede conceptualizar a través de una variable aleatoria de Bernoulli. A partir de esta se puede definir la PD es la probabilidad de que una operación o cliente entren en estado de incumplimiento *fallida* de sus obligaciones crediticias en un periodo temporal prefijo

$$X_t = \begin{cases} 1 & \text{with prob.} = PD \\ 0 & \text{with prob.} = 1 - PD \end{cases}$$

- La probabilidad de fallida o PD depende del tiempo. Por lo tanto, se habla de probabilidad (acumulada) de fallida a diferentes horizontes temporales, como ya se ha visto



- Este parámetro tiene aplicación en diversos ejercicios relacionados con la gestión del riesgo de crédito, entre los que destacan el cálculo de provisiones contables, el cálculo del capital regulatorio, la gestión, admisión y seguimiento de clientes, y los *stress test*
 - En función de la aplicación para la que esté destinada el parámetro, la estimación de éste deberá satisfacer unas características u otras
 - El parámetro se puede ajustar a un momento concreto del ciclo económico, a un ciclo económico concreto o a un momento futuro, de modo que se reflejarían las expectativas futuras
 - El parámetro puede mostrar situaciones de estrés económico o situaciones sin ningún estrés
 - El parámetro se puede medir en diferentes periodos temporales
- Algunos conceptos asociados a la PD que son importantes son los siguientes:
 - El umbral de materialidad es el límite en términos de la exposición existente en el momento de entrada en fallida por encima del cual quede garantizada la representatividad de las fallidas consideradas y la estabilidad de las estimaciones
 - El periodo de cura es el intervalo de tiempo que debe aguardarse tras la salida de fallida de una operación /cliente para que pueda volver a considerarse en situación normal (típicamente son 12 meses)
 - La dependencia temporal de la PD se refiere a que la PD depende del horizonte al que se calcula. Por defecto se toma la de un año, pero para horizontes mayores habría que utilizar matrices de transición
 - Es habitual encontrarse con correlaciones entre los incumplimientos de las contrapartidas de las carteras. A la hora de asignar una probabilidad de incumplimiento, se suele asignar una probabilidad promediada para todo un ciclo económico (PD acíclica)
- Una serie temporal de frecuencia de fallidas observadas u *observed default frequency* (ODF) es un vector de variables que representan el porcentaje de acreditados que al inicio de cada periodo se encuentran en una situación contable normal y al final del mismo periodo acaban

estando clasificadas en situación dudosa. La fórmula a partir de la cual se calcula una serie ODF es la siguiente:

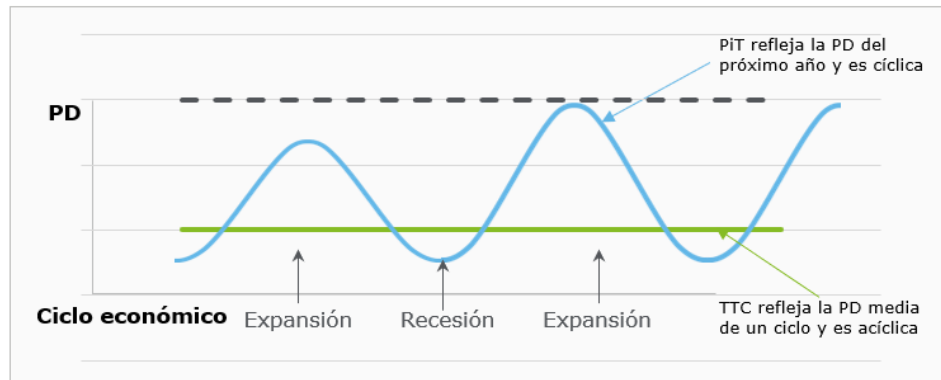
$$ODF = \frac{n^{\circ} \text{ bad operations}}{n^{\circ} \text{ bad operations} + n^{\circ} \text{ good operations (adjusted)}}$$

- El número de operaciones malas es el número de operaciones en situación normal al inicio del periodo que al final del periodo están clasificadas en situación dudosa. En cambio, el número de operaciones buenas ajustadas es el número de operaciones en situación normal y que no se han pasado a situación dudosa, ajustadas por la porción del periodo que estas han estado vigentes
- Como medida conservadora, y a fin de no subestimar la frecuencia, el peso de los contratos que no finalizan el periodo en situación dudosa se ajusta mediante el porcentaje de meses que estos contratos han tenido vigencia durante el periodo. Por tanto, si una operación ha estado vigente durante todo el periodo bajo análisis, esta tendrá un peso de 1, y si es menor, tendrá un peso menor a 1
- A partir de la PD estimada para un segmento homogéneo de riesgo y con la ayuda del teorema de Bayes, es posible ajustar la PD del segmento completo al grado deudor de cada una de las contrapartes que forman dicho segmento
 - Para obtener la PD de un grado de deudor a partir de la PD de todo el segmento se usa la siguiente fórmula:

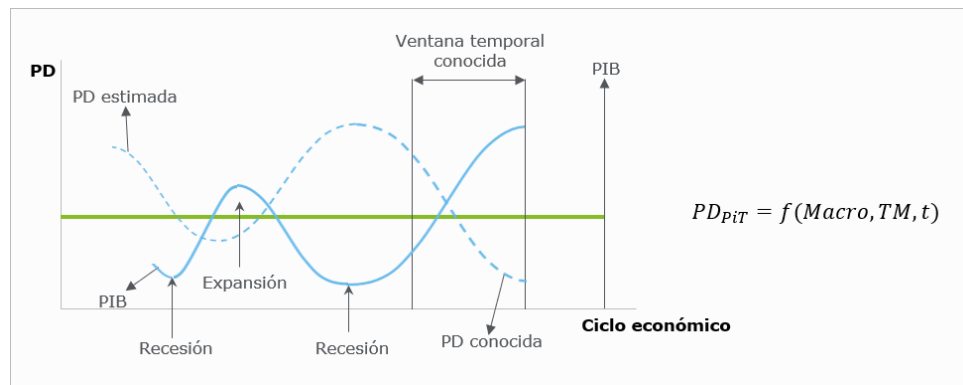
$$PD_{rating} = \frac{PD_{segment} \frac{M_{rating}}{M_{segment}}}{PD_{segment} \frac{M_{rating}}{M_{segment}} + (1 - PD_{segment}) \frac{B_{rating}}{B_{segment}}}$$

- En este caso, M denota el número de operaciones malas y B denota el número de operaciones buenas
- A partir de la PD estimada para un segmento homogéneo de riesgo y con la ayuda del teorema de Bayes, es posible ajustar la PD del segmento completo al grado deudor de cada una de las contrapartes que forman dicho segmento
- La probabilidad de fallida se puede aplicar principalmente a dos usos: a la determinación de capital (IRB) y a la determinación de provisiones (IFRS9). Ambos usos utilizan enfoques determinados que se tienen que estudiar y analizar

- En función de la naturaleza de las variables utilizadas en un modelo de estimación y de la información utilizada en el cálculo de la PD, se pueden distinguir modelos *Point-in-Time* (PiT) y modelos *Through The Cycle* (TTC)

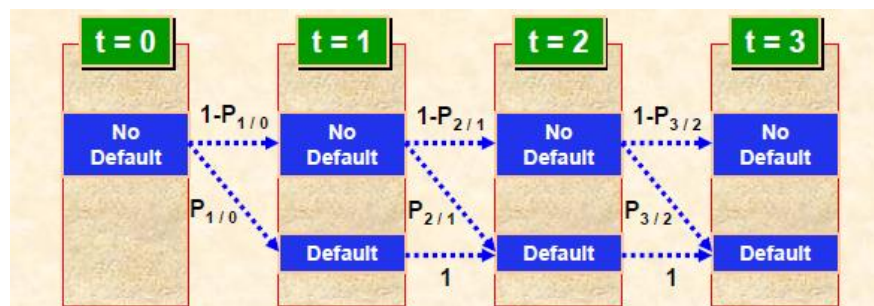


- Un modelo de calificación es PiT cuando es servir como herramienta de ordenación de rangos o *rank ordering* ajustado al ciclo económico, de modo que las probabilidades de impago asociadas a cada nivel de calificación cambian en el tiempo
 - Las PD se basan en las observaciones más recientes (en caso de que la evolución histórica de la PD presente una cantidad de inercia razonable) o empleo de covariantes que reflejen el ciclo económico (variación porcentual del PIB, incremento de ingresos, incremento de deudas, etc.)
 - Un modelo de calificación es TTC cuando establece una relación entre el PD y el nivel de calificación unívoca e independiente de cualquier factor externo
 - Las PD se basan en una tendencia central o tasa de incumplimiento promedio de la cartera en un ciclo económico completo
- A través de la PD conocida en un horizonte temporal y la serie histórico de una variable macroeconómica (PIB) se puede construir un modelo para obtener una PD estimada durante el ciclo económico



- La PD PiT está en función de las variables macroeconómicas, de la tasa de morosidad y del tiempo
- Las diferentes tipologías de PD con las que se trabajan son las siguientes:
 - La probabilidad de fallida acumulada es la probabilidad de que una contrapartida incumpla entre el momento actual $t = 0$ y algún año futuro T

$$p_T^{accum} = 1 - S_T = 1 - \prod_{t=1}^T (1 - p_t)$$



- La probabilidad de fallida acumulada es la probabilidad de que una contrapartida incumpla entre el momento actual $t = 0$ y algún año futuro T
- Anteriormente se ha hablado sobre los detalles probabilísticos de la PD, de modo que se puede entrar más en detalle sobre el proceso de calibración práctico que se realiza para este parámetro

El riesgo crediticio en la práctica: la LGD

El riesgo crediticio en la práctica: la EAD

Los modelos estructurales de incumplimiento: el modelo de Merton

- En los modelos del valor de una empresa o estructurales de fallida, uno postula un mecanismo para la fallida de una empresa en términos de su relación entre el activo y el pasivo
 - Normalmente, la fallida ocurre cuando una variable estocástica (o en modelos dinámicos, un proceso estocástico), generalmente representando el valor de los activos, cae por debajo de un nivel, representando el pasivo
 - El tipo de pensamiento que inducen estos modelos ha sido muy influyente en el análisis del riesgo crediticio y en el desarrollo de soluciones industriales
 - El modelo más importante es el modelo de Merton, el cual ha llevado a la creación de varios modelos posteriores y a aplicaciones industriales
- El modelo propuesto por Merton en 1974 es el prototipo para todos los modelos del valor de una empresa. Este modelo es una descripción estilizada de lo que es una fallida, dado que es mucho más complejo y el marco legal actual permite a una empresa no hacer automáticamente necesario liquidar la empresa en caso de fallida, pero es un buen punto por el que comenzar
 - En este modelo se considera una empresa cuyo valor de sus activos sigue un proceso estocástico $(V_t)_{t=0}^{\infty}$ y que se financia únicamente de capital y de deuda
 - El capital consiste de acciones que no pagan dividendos, mientras que la deuda consiste de bonos de cupón cero con vencimiento T y valor nominal B , los cuales solo se pueden emitir una vez (no se pueden emitir nuevos)
 - Los valores del capital y la deuda en el momento t se denotan por S_t y B_t respectivamente. La fallida ocurre si la empresa no realiza un pago a sus prestamistas, lo cual solo puede ocurrir en el momento T , y se pueden distinguir dos casos en ese momento:
 - Si $V_T \geq B$, el valor nominal de los activos de la empresa es igual o excede el del pasivo, de modo que los prestamistas reciben B

y los accionistas reciben el valor residual $S_T = V_T - B$ y no hay fallida

- Si $V_T < B$, el valor nominal de los activos es menor al del pasivo, de modo que no se puede cumplir con las obligaciones financieras de la empresa. Debido a que los accionistas no tienen interés en emitir más acciones (dado que iría todo a los prestamistas), la empresa se declara en quiebra y los activos pasan a ser gestionados por los prestamistas, los cuales liquidan la empresa y reciben $B_T = V_T$ (mientras que los accionistas no reciben nada)
- Una debilidad de este modelo es que, por muy cerca que esté V_T de B , si $V_T \geq B$ entonces no se produce una situación de fallida. En la realidad, esto no suele suceder así, y se han hecho modelos más realistas, tales como los modelos de primer paso de tiempo o los modelos con saltos

- Un modelo de primer paso o *first-passage-time model* es un tipo de modelo donde se da una situación de fallida la primera vez que cruza el umbral de deuda B (el momento τ), el cual se interpreta como el valor medio del pasivo

$$\tau \equiv \inf\{t \geq 0 : V_t \geq B\}$$

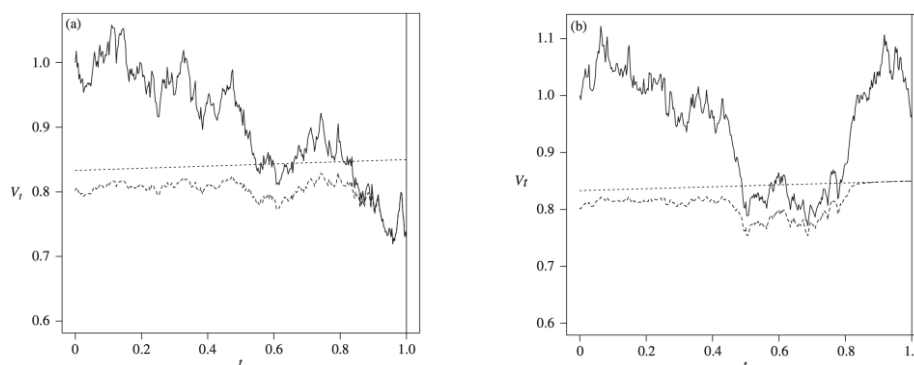
- Otros desarrollos técnicos en el área provienen de utilizar modelos de tipos de interés estocásticos sin posibles incumplimientos y modelos con saltos de difusión o *model with diffusion jumps*, que considera saltos en el valor de los activos de la empresa V_t a través de brownianas fraccionales
- Otra extensión interesante del modelo de Merton son los modelos con un umbral endógeno de incumplimiento. En este caso, el umbral de incumplimiento B no se fija *a priori*, si no que se determina endógenamente por consideraciones estratégicas de los accionistas
- Finalmente, modelos estructurales con información incompleta del valor de los activos o del pasivo proporcionan un buen vínculo entre los enfoques estructurales y de forma reducida para el modelaje del riesgo de crédito
- Estas dos situaciones permiten establecer las siguientes relaciones entre el valor de las acciones y la deuda:

$$S_T = \max(V_T - B, 0) = (V_T - B)^+$$

$$B_T = \min(V_T, B) = B - (B - V_T)^+$$

- La primera ecuación implica que el valor de las acciones de la empresa en T equivale a la recompensa de una acción de compra europea sobre V_t , mientras que la segunda implica que el valor de la deuda de la empresa en su vencimiento es el valor nominal de esta menos la recompensa de una opción de venta europea sobre V_t (ambas opciones con precio de ejercicio B)
- La interpretación de las acciones y de los bonos como opciones es útil para explicar posibles conflictos de interés entre accionistas y prestamistas
 - Manteniéndose todo igual, un incremento en la volatilidad del subyacente de una opción hace que aumente su valor, por lo que los accionistas tienen incentivos en tomar proyectos arriesgados en la empresa, mientras que, debido a que los prestamistas tienen una posición corta en la opción de venta, estos querrán ver la volatilidad reducida y preferirán proyectos menos arriesgados
- En el modelo de Merton se asume que, bajo la medida de probabilidad P del mundo real o físico, el proceso $(V_t)_{t=0}^{\infty}$ sigue un modelo de difusión de una moción browniana geométrica (o de Black-Scholes) de la siguiente forma:

$$dV_t = \mu_V V_t dt + \sigma_V V_t dW_t \quad \text{where } \mu_V \in \mathbb{R} \text{ \& } \sigma_V > 0$$



- En este caso, μ_V se puede interpretar como el rendimiento de los activos de la entidad bajo la probabilidades reales o físicas (a diferencia del uso del rendimiento en el mundo neutral al riesgo, que sería la tasa libre de riesgo r)
- La ecuación se puede resolver de manera explícita para encontrar el valor V_T

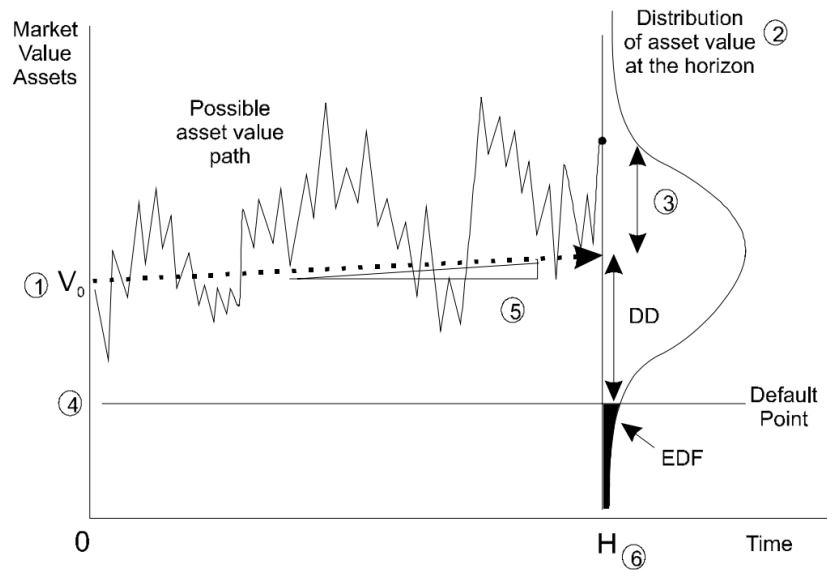
$$V_t = V_0 e^{(\mu_V - \frac{\sigma_V^2}{2})t + \sigma_V dW_t}$$

- Debido a que $W_t \sim N(0, T)$, se puede deducir que el valor de los activos $\ln V_t \sim N(\ln V_0 + (\mu_v - \sigma_v/2)T, \sigma_v^2 T)$, por lo que se puede encontrar una expresión para calcular la probabilidad de fallida

$$PD = P(V_t \leq B) = P(\ln V_t \leq \ln B) =$$

$$= \Phi\left(\frac{\ln B - \ln V_0 - (\mu_v - \frac{\sigma_v}{2})T}{\sigma_v \sqrt{T}}\right) = \Phi\left(\frac{\ln\left(\frac{B}{V_0}\right) - (\mu_v - \frac{\sigma_v}{2})T}{\sigma_v \sqrt{T}}\right)$$

- La probabilidad de fallida guarda una relación positiva con B y una negativa con V_0 y μ_v , mientras que, para $V_0 > B$, la volatilidad σ_v guarda una relación positiva. Esto tiene sentido económico, dado que un mayor valor inicial o tendencia del valor de los activos creará más distancia con el umbral de fallida, pero una mayor volatilidad del valor hace que haya más probabilidad de estar por debajo del umbral



- En el contexto del modelo de Merton, uno puede fijar el valor de los instrumentos cuya recompensa dependa del valor de los activos de la empresa V_T en T , tales como las acciones de la empresa o los instrumentos de deuda
 - En el análisis de la valoración de activos en el contexto del modelo de Merton se utilizan las siguientes suposiciones:
 - El tipo de interés sin riesgo r es determinístico y no es negativo, de modo que $r \geq 0$

- El proceso estocástico del valor de los activos de la empresa $(V_t)_{t=0}^{\infty}$ es independiente de la manera en que la empresa se financia, y en particular, independiente del nivel B
- El valor de los activos de la empresa $(V_t)_{t=0}^{\infty}$ se puede comerciar en un mercado sin fricciones, y las dinámicas del valor de los activos viene dada por una moción browniana geométrica
- Algunos comentarios que se pueden hacer sobre las suposiciones anteriores son los siguientes:
 - Esta independencia de $(V_t)_{t=0}^{\infty}$ de la estructura financiera es cuestionable, dado que para grandes niveles de deuda (y, por tanto, una alta probabilidad de fallida) pueden afectar adversamente a la capacidad de realizar las actividades principales de la empresa, afectando así al valor de los activos (costes indirectos de la bancarrota)
 - Aunque hay muchas empresas que comercian sus acciones, el valor de los activos no se puede observar ni comerciar de manera completa (sin fricciones)
- Considerando un reclamo sobre el valor de los activos de la empresa con vencimiento T y recompensa $h(V_T)$, se pueden valorar las acciones de la empresa y los bonos (debido a las relaciones anteriormente vistas)
 - Gracias a las suposiciones hechas, se puede ver como el *fair value* $f(t, V_t)$ de este reclamo en un momento $t \leq T$ puede ser calculado usando la valoración neutral al riesgo para una medida de probabilidad Q (en el mundo neutral al riesgo)

$$f(t, V_t) = E_Q[e^{-r(T-t)} h(V_T) | \mathcal{F}_t]$$

- Debido a que $S_T = (V_T - B)^+$, el valor de las acciones se puede fijar a través de la fórmula de Black-Scholes para opciones de compra europea con precio de ejercicio B y vencimiento T

$$S_t = C^{BS}(t, V_t; r, \sigma_V, B, T) = V_t \Phi(d_{t,1}) - B e^{-r(T-t)} \Phi(d_{t,2})$$

$$\text{where } d_{t,1} \equiv \frac{\ln\left(\frac{V_t}{B}\right) + \left(r + \frac{\sigma_V^2}{2}\right)(T-t)}{\sigma_V \sqrt{T-t}} \text{ \& } d_{t,2} \equiv d_{t,1} - \sigma_V \sqrt{T-t}$$

- Debido a que se asume un tipo de interés constante, el precio en $t \leq T$ de un bono de cupón cero sin riesgo con vencimiento T y con valor nominal 1 es $p_0(t, T) = \exp[-r(T-t)]$. Como $B_T = B - (B - V_t)^+$, se puede utilizar la fórmula de Black-Scholes

para una opción de venta europea con precio de ejercicio B y vencimiento T

$$B_t = Bp_0(t, T) - P^{BS}(t, V_t; r, \sigma_V, B, T) \quad \text{where}$$

$$P^{BS}(t, V_t; r, \sigma_V, B, T) = Be^{-r(T-t)}\Phi(-d_{t,2}) - V_t\Phi(-d_{t,1})$$

$$\Rightarrow B_t = Be^{-r(T-t)}\Phi(d_{t,2}) - V_t\Phi(-d_{t,1})$$

- Es interesante calcular cuál es la volatilidad de las acciones de la empresa bajo las suposiciones hechas, por lo que se puede definir la siguiente cantidad:

$$v(t, V_t) = \frac{V_t}{C^{BS}(t, V_t)} \frac{\partial C^{BS}(t, V_t)}{\partial V} = \frac{V_t C_V^{BS}(t, V_t)}{C^{BS}(t, V_t)}$$

- En el contexto de la apreciación de opciones, esta cantidad representa la elasticidad de una opción de compra europea con respecto al precio del subyacente. En este contexto, la cantidad mide el cambio porcentual del valor del capital por el cambio porcentual en el valor de los activos subyacentes
- Aplicando la fórmula de Itô en $S_t = C^{BS}(t, V_t; r, \sigma_V, B, T)$ y usando la definición de la elasticidad v , se obtiene la siguiente fórmula:

$$dS_t = \left(C_t^{BS} + \mu_V V_t C_V^{BS} + \frac{\sigma_V}{2} V_t^2 C_{VV}^{BS} \right) dt + \sigma_V V_t C_V^{BS} dW_t$$

$$\Rightarrow dS_t = \left(C_t^{BS} + \mu_V V_t C_V^{BS} + \frac{\sigma_V}{2} V_t^2 C_{VV}^{BS} \right) dt + \sigma_V C^{BS} v(t, V_t) dW_t$$

- Debido a esto, se puede ver como la volatilidad del capital de la empresa (de las acciones) es una función determinística $\sigma_S(t, V_t)$ de la siguiente forma:

$$\sigma_S(t, V_t) = v(t, V_t) \sigma_V$$

- Debido a que la elasticidad de una opción de compra europea es siempre mayor a uno, la volatilidad $\sigma_S(t, V_t)$ siempre será mayor a σ_V
- A partir de las suposiciones, también es posible calcular las probabilidades de fallida físicas y en el mundo neutral al riesgo en el modelo de Merton

- Un resultado básico de las matemáticas financieras muestra que bajo la medida de probabilidad neutral al riesgo Q , el proceso $(V_t)_{t=0}^{\infty}$ satisface la siguiente ecuación diferencial estocástica $dV_t = rV_t dt + \sigma_V V_t d\tilde{W}_t$ para una moción browniana Q estándar
- Por lo tanto, la probabilidad de fallida en el mundo neutral al riesgo viene dada por la fórmula anteriormente vista en el modelo de Merton (con r en vez de μ_V):

$$q = Q(V_T \leq B) = \Phi\left(\frac{\ln\left(\frac{B}{V_0}\right) - \left(r - \frac{\sigma_V^2}{2}\right)T}{\sigma_V \sqrt{T}}\right)$$

- Si se comparan estas probabilidades con las físicas dadas por el modelo de Merton (la probabilidad $p = P(V_T \leq B)$), se puede establecer la siguiente relación, en donde el término de corrección $(\mu_V - r)/\sigma_V$ es la *Sharpe ratio* de V :

$$q = \Phi\left(\Phi^{-1}(p) + \frac{\mu_V - r}{\sigma_V} \sqrt{T}\right)$$

- La fórmula de transición anterior a veces se usa en la práctica con tal de pasar de probabilidades físicas a probabilidades neutras al riesgo. No obstante, esta fórmula solo es válida si se mantienen las suposiciones hechas en el contexto del modelo de Merton

Los modelos estructurales de incumplimiento: modelos prácticos

- Existen muchos modelos prácticos industriales que se derivan del modelo de Merton, y un ejemplo importante de este tipo de modelos es el modelo EDF de empresa pública, usado por *Moody's Analytics*
 - El acrónimo EDF proviene de *expected default frequency* o frecuencia de incumplimiento esperada, el cual es un estimador de la probabilidad de incumplimiento física de una empresa en un periodo de tiempo anual
 - La metodología propuesta por *Moody's Analytics* se origina por trabajar previo de KMV en los años 90, por lo que también se conoce como el modelo KMV
 - Para presentar el desarrollo de este modelo uno se debería concentrar las ideas básicas, dado que los detalles sobre la implementación y el calibrado es propietario y estos procedimientos pueden cambiar en el tiempo

- En el modelo de Merton clásico, la probabilidad de incumplimiento en un horizonte de un año se da por la probabilidad de que el valor de los activos esté por debajo del umbral B , representando el pasivo de la empresa

- Bajo las suposiciones del modelo, la probabilidad de incumplimiento es una función del valor actual de los activos V_0 , la tasa de deriva anualizada μ_V y la volatilidad σ_V del proceso del valor de los activos y el umbral B . Asumiendo $T = 1$, se obtiene la infiere la siguiente expresión:

$$EDF_{Merton} = 1 - \Phi \left(\frac{\ln V_0 - \ln B + \left(\mu_V - \frac{1}{2} \sigma_V^2 \right)}{\sigma_V} \right)$$

- El modelo de EDF público tiene una estructura similar al EDF del modelo de Merton, pero $1 - \Phi$ se reemplaza por una función empíricamente estimada, B se reemplaza por un umbral \tilde{B} representando la estructura del pasivo de la empresa de manera más precisa y el término $\left(\mu_V - \frac{1}{2} \sigma_V^2 \right)$ en el numerador se omite a veces. Además, el valor actual de los activos de la empresa V_0 y la volatilidad σ_V se infieren de la información del capital
- Los modelos que se basan en el valor de la empresa normalmente se apoyan en el valor de mercado V_0 de esta
 - Esto tiene sentido debido a que el valor de mercado es una medida futurística que refleja las expectativas de los inversores sobre el negocio y sus futuros flujos de caja
 - No obstante, en contraste con las suposiciones del modelo de Merton, en la mayoría de casos no hay mercado para los activos de una empresa, de modo que este valor no es directamente observable. Además, el valor de mercado puede diferir mucho de su valor contable, de modo que la información contable y los datos del balance contable no son muy útiles para inferir V_0
 - Por lo tanto, los modelos EDF públicos se suelen basar en un enfoque indirecto e infiere los valores V_t en cada momento t usando los valores del capital de la empresa S_t , que son más fáciles de observar y permite obtener estimaciones para V_0 y σ_V (muy importante por su efecto en las probabilidades de incumplimiento)
- El procedimiento para estimar las cantidades se basa en el modelo de Merton. Según las suposiciones hechas en el modelo, el precio del capital de la empresa en t es igual al precio de una opción *call*:

$$S_t = C_{BS}(t, V_t; r, \sigma_V, B, T)$$

- Obviamente, en un punto fijo de t , la ecuación anterior tiene solo dos incógnitas (V_0 y σ_V)
- Para poder atravesar esta dificultad, uno puede utilizar un proceso iterativo: en el primer paso se usa la ecuación anterior con un valor inicial $\sigma_V^{(0)}$ y $V_t^{(0)}$ para inferir las series temporales del valor de los activos a partir de las del capital; en el segundo, se vuelve a estimar $\sigma_V^{(1)}$ a partir de la serie temporal, y se construye otra serie $\{V_t^{(1)}\}$ usando la ecuación anterior y se estima $\sigma_V^{(1)}$. Todo este proceso se itera n veces, hasta que las estimaciones de las volatilidades $\sigma_V^{(n-1)}$ y $\sigma_V^{(n)}$ sean lo suficientemente parecidas
- En el modelo EDF público, la estructura del capital se modela de una manera más sofisticada que en el modelo de Merton: hay varias clases de pasivo (largo y corto plazo), el modelo permite pagos intermedios de cupones o dividendos y el incumplimiento puede ocurrir en cualquier momento. Además, el punto de incumplimiento \tilde{B} se determina con un análisis más detallado de la estructura temporal del pasivo de la empresa
- Por lo tanto, el capital de la empresa no se da por la ecuación anterior, si no que será dada por una función diferente $f(t, V_t, \sigma_V)$ que se tiene que calcular numéricamente
- En el modelo de Merton, el incumplimiento ocurre si el valor de los activos de la empresa está bajo un cierto umbral B del valor del pasivo, y con la suposición de valores distribuidos log-normales, esto lleva a probabilidades de incumplimiento de la forma EDF_{Merton}
 - Esta relación entre el valor de los activos y la probabilidad de incumplimiento puede ser demasiado simplista para describir realísticamente las probabilidades. Por ejemplo, los valores de los activos no se tienen que distribuir log-normal pero puede tener colas gruesas, y pueden haber pagos intermedios que causen el incumplimiento en cualquier momento
 - Por estas razones, en el modelo EDF público, se introduce una variable de estado intermedia en un paso intermedio, llamada distancia hasta el incumplimiento o *distance-to-default*. Esta medida a veces se interpreta como el número de desviaciones estándar de lejanía hasta el umbral de incumplimiento

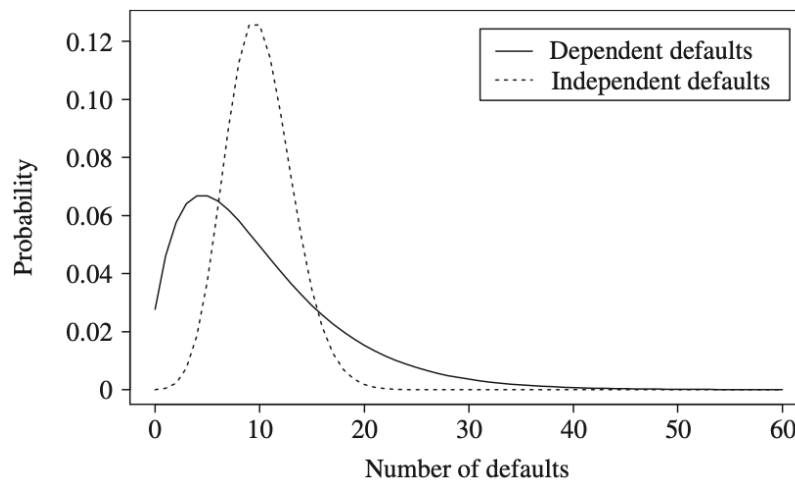
$$DD = (\log V_0 - \log \tilde{B})/\sigma_V$$

- En algunas versiones del modelo EDF, \tilde{B} se modela como la suma de los pasivos que se pagan dentro de un año y medio de la deuda a largo plazo
- En la metodología EDF, se asume que la DD clasifica las empresas debido a que una mayor DD hace que la empresa tenga menos probabilidades de incumplimiento
- La forma funcional de la relación entre la EDF y la DD se determina empíricamente usando una base de datos históricos sobre incumplimientos de la empresa para estimar la proporción de empresas con una DD en un rango pequeño que incumplen en un horizonte temporal de un año (la EDF empíricamente estimada)
- El mapeado de DD a EDF exhibe colas gruesas, dado que para empresas de mucha calidad (con una DD alta) la EDF empíricamente estimada es mucho mayor que la que da la EDF en el modelo de Merton

Variable	J&J	RadioShack	Notes
Market value of assets V_0	236 bn	1834 m	Determined from time series of equity prices
Asset volatility σ_V	11%	24%	
Default threshold \tilde{B}	39 bn	1042 m	Short-term liabilities and half of long-term liabilities
DD	16.4	2.3	Given by $(\log V_0 - \log \tilde{B})/\sigma_V$
EDF (one year)	0.01%	3.58%	Determined using empirical mapping between DD and EDF

La gestión del riesgo de carteras de crédito: modelos de umbral

- El tema de mayor importancia en la gestión del riesgo de crédito es modelar la estructura de dependencia de los eventos de fallida, dado que esta tiene un impacto profundo en la cola superior de la distribución de pérdidas de una cartera crediticia grande
 - Si se compara la distribución de pérdidas de dos carteras, una en la que la fallida de las empresas es independiente una de la otra, y una en la que tienen una dependencia débil, entonces se puede ver que hay diferencias entre ambas distribuciones



- Se puede ver claramente como la distribución de pérdidas de la cartera con dependencia débil está más sesgada a la izquierda que la de la cartera con independencia, haciendo que la cola superior sea sustancialmente más gruesa
- Hay diversas razones económicas por las que se espera dependencia entre eventos de falla de diferentes deudores
 - La más importante es que la salud financiera de las empresas varía con factores macroeconómicos aleatorios (como el crecimiento), así que, si varias de estas empresas son afectadas por factores macroeconómicos comunes, esto crearía dependencia
- Los modelos de umbral son modelos de un solo periodo para riesgo crediticio de cartera que se inspiran en los modelos del valor de la empresa vistos anteriormente
 - El atributo característico es la idea de que la falla de una compañía i puede ocurrir en el periodo $[0, T]$ si alguna variable aleatoria crítica X_i se encuentra bajo un umbral determinístico d_i
 - La variable X_i puede tener diferentes interpretaciones, aunque de manera más abstracta se suele interpretar como una variable latente que representa la calidad crediticia o la habilidad de repago del deudor i
 - La dependencia entre fallas proviene de la dependencia entre componentes del vector $\mathbf{X} = (X_1, X_2, \dots, X_m)'$. Las distribuciones asumidas para \mathbf{X} pueden ser completamente generales y tener una u otra cópula

- Para poder explicar estos modelos, se introduce una notación para los modelos de cartera de un solo periodo que se utilizará a lo largo del desarrollo

- Se considera una cartera de m deudores y un horizonte temporal fijo T
- Para $1 \leq i \leq m$, se necesita que la variable aleatoria R_i sea un indicador de estado para i en T y se asume que R_i toma valores discretos en el conjunto $\{0, 1, 2, \dots, n\}$ representando, por ejemplo, calificaciones crediticias. En el momento $t = 0$, los deudores están en un estado diferente al de fallida (diferente a cero)
- Como solo se está interesado en si ocurre la fallida o no, se crea una variable Y_i donde $Y_i = 1$ si, y solo si, $R_i > 0$, y $Y_i = 0$ si, y solo si, $R_i = 0$

$$Y_i = \begin{cases} 0 & \text{when } R_i = 0 \\ 1 & \text{when } R_i > 0 \end{cases}$$

- El vector aleatorio $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)'$ es un vector de indicadores de fallida para la cartera, y $p(\mathbf{y}) = P(Y_1 = y_1, \dots, Y_m = y_m)$ para $\mathbf{y} \in [0, 1]^m$ es la función de masa de probabilidad conjunta (probabilidad de fallida), mientras que las probabilidades de fallida marginales son $p_i = P(Y_i = 1)$ para $i = 1, 2, \dots, m$
- Las correlaciones de fallida se definen como las correlaciones de los indicadores de fallida, las cuales se obtienen de la siguiente manera:

$$E(Y_i) = 1 \times p_i + 0 \times (1 - p_i) = p_i$$

$$\Rightarrow \text{Var}(Y_i) = E(Y_i^2) - E(Y_i)^2 = p_i - p_i^2$$

$$\text{Cov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j) = E(Y_i Y_j) - p_i p_j \quad \text{for } i \neq j$$

$$\Rightarrow \rho(Y_i, Y_j) = \frac{E(Y_i Y_j) - p_i p_j}{\sqrt{(p_i - p_i^2)(p_j - p_j^2)}} \quad \text{for } i \neq j$$

- Se tiene en cuenta el número de deudores en situación de fallida en el momento T con la variable aleatoria $M \equiv \sum_{i=1}^m Y_i$

- El *loss given default* (LGD) se modela por la cantidad aleatoria $\delta_i e_i$, donde e_i representa la exposición en fallida o EAD) y $\delta_i \in [0,1]$ representa una proporción aleatoria de la exposición que se pierde en el evento de fallida. Por tanto, la pérdida general se denota como $L \equiv \sum_{i=1}^m \delta_i e_i Y_i$
- Para simplificar el análisis, se puede asumir que el vector de indicadores de estado y el vector de indicadores de fallida son vectores aleatorios intercambiables
 - Esto hace que se formalice matemáticamente la noción de grupos homogéneos que se usa en la práctica
- Siendo $\mathbf{X} = (X_1, X_2, \dots, X_m)'$ un vector m -dimensional y $D \in \mathbb{R}^{m \times n}$ una matriz determinística con elementos d_{ij} tales que, para cada i , los elementos de la fila i forman un conjunto de umbrales estrictamente crecientes $d_{i0} < d_{i1} < \dots < d_{in} < d_{i(n+1)}$ (donde $d_{i0} = -\infty$ y $d_{i(n+1)} = \infty$ para todos los deudores) y $R_i = j$ si, y solo si, $d_{ij} < x_i \leq d_{i(j+1)}$ para toda i y j , entonces (\mathbf{X}, D) define un modelo de umbral para el vector de estados $\mathbf{R} = (R_1, R_2, \dots, R_m)'$
 - El vector \mathbf{X} es el vector de variables críticas y se denotan sus funciones de distribución marginales por $F_{X_i}(x) = P(X_i \leq x)$. La fila i de D contiene los umbrales críticos para la empresa i
 - Por definición, la fallida ocurre cuando $X_i \leq d_{i1}$, de modo que la probabilidad de fallida de la compañía i se da por $p_i = F_{X_i}(d_{i1})$. Si solo se trabaja con un modelo en donde solo se tiene en cuenta la fallida, se escribe $d_{i1} = d_i$ y se denota el modelo de umbral por (\mathbf{X}, \mathbf{d})
- Es posible generalizar el modelo de Merton para una cartera de m empresas. Asumiendo que un proceso del valor de los activos multivariante \mathbf{V}_t con $\mathbf{V}_t = (V_{t,1}, V_{t,2}, \dots, V_{t,m})'$ sigue un proceso geométrico browniano m -dimensional con vector de deriva $\boldsymbol{\mu}_V = (\mu_1, \mu_2, \dots, \mu_m)'$, vector de volatilidades $\boldsymbol{\sigma}_V = (\sigma_1, \sigma_2, \dots, \sigma_m)'$ y matriz de correlación instantánea \mathbf{P}

$$dV_{t,i} = \mu_i V_{t,i} dt + \sigma_i V_{t,i} dW_{t,i} \text{ where } \rho(dW_{t,i}, dW_{t,j}) = \rho_{ij}$$

$$\text{for } i, j = 1, 2, \dots, m \text{ and } t \geq 0$$

- Para toda i , el valor de los activos $V_{T,i}$ tiene la siguiente forma, donde $\mathbf{W}_T = (W_{T,1}, W_{T,2}, \dots, W_{T,m})'$ es un vector aleatorio normal multivariante satisfaciendo $\mathbf{W}_T \sim N_m(\mathbf{0}, T\mathbf{P})$:

$$V_{T,i} = V_{0,i} \exp \left[\left(\mu_i - \frac{\sigma_i^2}{2} \right) T + \sigma_i W_{T,i} \right]$$

- En su forma más básica, el modelo de Merton es un modelo solo de fallida, en donde una empresa está en situación de fallida si $V_{T,i} < B_i$ y B_i es el valor del pasivo de la empresa i . Escribiendo $\mathbf{B} = (B_1, B_2, \dots, B_m)'$, la representación del modelo de umbral es (V_T, \mathbf{B})
- Debido a que en un modelo de umbral el evento de incumplimiento es invariante bajo transformaciones estrictamente crecientes de variables críticas y umbrales, este modelo es equivalente a un modelo de umbral (\mathbf{X}, \mathbf{d}) que satisface las siguientes ecuaciones:

$$X_i = \frac{\ln V_{T,i} - \ln V_{0,i} - \left(\mu_i - \frac{1}{2} \sigma_i^2 \right) T}{\sigma_i \sqrt{T}}$$

$$d_i = \frac{\ln B_i - \ln V_{0,i} - \left(\mu_i - \frac{1}{2} \sigma_i^2 \right) T}{\sigma_i \sqrt{T}}$$

- Las variables transformadas satisfacen $\mathbf{X} \sim N_m(\mathbf{0}, \mathbf{P})$ (debido a que son combinación lineal de variables normales) y su cópula es la cópula gaussiana C_P^{Ga}
- Asumiendo que $X_i = \sqrt{\beta_i} \tilde{F}_i + \sqrt{1 - \beta_i} \varepsilon_i$, donde \tilde{F}_i y $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ son variables normales estándar independientes y que $\beta_i \in [0,1]$ para toda i , se tiene un modelo multifactorial de Merton

$$X_i = \sqrt{\beta_i} \tilde{F}_i + \sqrt{1 - \beta_i} \varepsilon_i$$

- En esta formulación, \tilde{F}_i es un factor que representa las variables sistemáticas, que están correlacionadas, y ε_i son las variables idiosincráticas. Por lo tanto, β_i se puede considerar como la sensibilidad de X_i al riesgo sistemático (la parte de la varianza de X_i explicada por la variable sistémica)
- Se asume que las variables sistemáticas tienen la forma $\tilde{F}_i = \mathbf{a}_i' \mathbf{F}$, donde \mathbf{F} es un vector de factores comunes que se distribuyen $\mathbf{F} \sim N_p(\mathbf{0}, \Omega)$ con $p > m$, donde Ω es una matriz de correlación (la matriz de varianzas y covarianzas). Normalmente \mathbf{F} representa los valores de los efectos industriales o regionales
- Debido a que $Var(X_i) = 1$ y a que \tilde{F}_i y $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ son variables normales estándar independientes, las correlaciones entre activos se dan por la siguiente expresión:

$$Var(\tilde{F}_i) = E[(\mathbf{a}_i' \mathbf{F})(\mathbf{a}_i' \mathbf{F})'] = E(\mathbf{a}_i' \mathbf{F} \mathbf{F}' \mathbf{a}_i) = \mathbf{a}_i' \Omega \mathbf{a}_i = 1$$

$$\Rightarrow Var(X_i) = \beta_i Var(\tilde{F}_i) + (1 - \beta_i) Var(\varepsilon_i) = \beta_i + 1 - \beta_i = 1$$

$$\Rightarrow \rho(X_i, X_j) = Cov(X_i, X_j) = \sqrt{\beta_i \beta_j} Cov(\tilde{F}_i, \tilde{F}_j) = \sqrt{\beta_i \beta_j} \mathbf{a}_i' \Omega \mathbf{a}_j$$

- Para poder construir el modelo, es necesario especificar \mathbf{a}_i , β_i y Ω para cada deudor, con la restricción añadida de que $\mathbf{a}_i' \Omega \mathbf{a}_i = 1$ para toda i . Debido a que Ω tiene $p(p-1)/2$ parámetros, el vector \mathbf{a}_i y los coeficientes β_i tienen un total combinado de $mp + m$ parámetros, aplicando m restricciones, de modo que la dimensión del problema de calibración es $mp + p(p-1)/2$ (el número de parámetros crece linealmente en m)
- El modelo factorial se puede expresar como los modelos multifactoriales vistos anteriormente, donde $B = DA$, $D = diag(\sqrt{\beta_1}, \dots, \sqrt{\beta_m})$, $A \in \mathbb{R}^{m \times p}$ es la matriz con la fila i dada por \mathbf{a}_i' y $\tilde{\varepsilon}_i = \sqrt{1 - \beta_i} \varepsilon_i$

$$\mathbf{X} = B\mathbf{F} + \tilde{\varepsilon}_i$$

- Normalmente se suele asumir que se tiene un solo factor $F = \tilde{F}_i$ y que todos los deudores tienen la misma varianza sistemática $\beta_i = \rho$, de modo que $\rho(X_i, X_j) = \rho$ para toda $i \neq j$ y se está en un modelo de equicorrelación

$$X_i = \sqrt{\beta_i} F + \sqrt{1 - \beta_i} \varepsilon_i$$

$$\Rightarrow \rho(X_i, X_j) = \sqrt{\beta_i \beta_j} Cov(\tilde{F}_i, \tilde{F}_j) = \sqrt{\rho^2} Var(F) = \rho$$

La gestión del riesgo de carteras de crédito: el modelo ASRF

- El modelo de Merton multifactorial es la base del enfoque IRB del acuerdo Basel II, de modo que se pueden utilizar enfoques basados en una versión equivalente de este modelo factorial para gestionar el riesgo crediticio. El marco general que se utiliza para el desarrollo de las técnicas usadas se basa en métricas básicas de gestión del riesgo y medidas crediticias
 - Los modelos se pueden construir considerando los siguientes elementos:
 - Hay una cartera de N deudores que se caracterizan por tres parámetros: la exposición en fallida E_n , la pérdida dada la fallida

L_n y la probabilidad de fallida P_n (todos estimados de datos de fallida empíricos)

- Para representar la incertidumbre sobre el futuro, se especifica un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y con una filtración $\mathcal{F}_{t \geq 0}$ satisfaciendo las condiciones usuales
- Fijando un horizonte temporal $T = 1$, se puede considerar un deudor n sujeto a fallida en el horizonte temporal fijo T y crear así una variable D_n que indique si el deudor está en situación de fallida (con valor 1) o no (con valor 0)

$$D_n = \begin{cases} 1 & \text{with } P(D_n = 1) = P_n \\ 0 & \text{with } P(D_n = 0) = 1 - P_n \end{cases}$$

- Se define \mathcal{L} como la pérdida en la cartera, la cual depende de las pérdidas de los deudores \mathcal{L}_n , definidas como $E_n L_n D_n$

$$\mathcal{L} = \sum_{n=0}^N \mathcal{L}_n = \sum_{n=0}^N E_n L_n D_n$$

- En general, el riesgo de crédito se puede dividir en pérdidas esperadas (EL), que se pueden predecir y gestionar fácilmente, y las pérdidas inesperadas (UL), las cuales son más complicadas de cuantificar

- Las pérdidas inesperadas se pueden clasificar como en riesgo sistémico, que nace de interdependencias entre deudores individuales en una cartera y de shocks, y en riesgo idiosincrático, que nace de shocks específicos para los deudores
- Este último tipo puede ser diversificado, mientras que el riesgo sistémico no puede ser eliminado (aunque si reducido a través de cambiar la exposición para reducir la correlación)

- La desviación de las pérdidas de las EL normalmente se mide a través de la desviación estándar de las pérdidas

- No obstante, como hay una probabilidad alta de que las pérdidas superen las EL de la cartera para más de una desviación estándar, mantener las UL como capital de riesgo para casos de dificultades financieras no es apropiado y se utilizan otros conceptos
- El Basel II recomienda el uso del VaR_α al 0.999 para el cálculo del requerimiento del capital, de modo que se tiene que reservar una cantidad equivalente al cuantil para cubrir pérdidas extremas

- El *expected shortfall* ES_α (o $TVaR_\alpha$) también se puede utilizar como medida del riesgo relevante
- También se puede considera como descomponer el riesgo medido por el VaR_α o el ES_α en las transacciones individuales. Considerando el principio de asignación, que es la derivada parcial de la medida de riesgo respecto a la exposición del deudor, es posible definir las siguientes contribuciones:

- La contribución al riesgo al VaR_α ($VaRC_{\alpha,n}$) de un deudor n a un nivel de confianza α se define de la siguiente manera:

$$VaRC_{\alpha,n} = E_n \frac{\partial VaR_\alpha}{\partial E_n}$$

- La contribución al riesgo al ES_α ($ESC_{\alpha,n}$) de un deudor n a un nivel de confianza α se define de la siguiente manera:

$$ES_{\alpha,n} = E_n \frac{\partial ES_\alpha}{\partial E_n}$$

- Ambas definiciones cumplen la condición aditiva, de modo que se cumplen las siguientes igualdades:

$$\sum_{n=1}^N VaRC_{\alpha,n} = VaR_\alpha \quad \sum_{n=1}^N ESC_{\alpha,n} = ES_\alpha$$

- El capital económico o *economic capital* (EC) se mantiene para pérdidas inesperadas que nacen del riesgo sistémico e idiosincrático
- Se supone que la exposición a la fallida E_n y la pérdida dada la fallida L_n son cantidades determinísticas. De este modo, las pérdidas esperadas se pueden calcular de la siguiente manera:

$$EL = E(\mathcal{L}) = \sum_{n=0}^N E_n L_n P_n$$

- El concepto de EC es un enfoque muy utilizado para los modelos de riesgo de crédito internos de los bancos, y es el capital que debe mantener el banco para reducir la probabilidad de fallida a un intervalo dado. Este depende de un nivel de confianza α , y se define como la diferencia entre el VaR_α y las pérdidas esperadas

$$EC = VaR_\alpha - EL$$

- El VaR se reduce por EL debido a la descomposición común del VaR, en donde parte se dedica a cubrir pérdidas esperadas y otra parte se dedica a cubrir pérdidas inesperadas
 - En el peor de los casos, un banco puede perder toda su cartera en un solo año, pero mantener un capital tan grande para un evento muy poco probable es económicamente ineficiente
 - Como los bancos quieren gastar la mayoría de su capital en inversiones beneficiosas, tienen incentivos a minimizar el capital que el banco mantiene para las pérdidas
 - Por lo tanto, el problema de la gestión del riesgo en una institución financiera es encontrar el balance entre mantener un capital suficiente para cumplir con todas las obligaciones, pero minimizando el capital económico para obtener más beneficios
 - Como en análisis de riesgo crediticio es deseable explicar el éxito económico de una empresa a través de influencias globales, se suelen utilizar modelos factoriales. El modelo ASFR utilizado normalmente se suele aplicar adaptando una versión multifactorial del modelo de Merton, la cual se presenta ahora
 - Para poder derivar la versión del modelo multifactorial de Merton usado en el ASFR, es necesario hacer las siguientes suposiciones:
 - Para el modelo se consideran N deudores, en donde se asume que cada uno tiene un préstamo con principal E_n y se fija un horizonte temporal $T > 0$
 - Se define como $V_{t,n}$ al valor de los activos de la contraparte n en el momento $t \leq T$, y se asume que para cada contraparte existe un umbral T_n tal que la contraparte n hace fallida en el periodo $[0, T]$ si $V_{t,n} < T_n$
 - Si T_n es el valor del pasivo de la contraparte, entonces $V_{t,n}$ se puede entender como la variable latente que determina la fallida y se puede definir una variable D_n para toda n tal que el evento 1 ocurre con probabilidad $p = P(V_{T,n} < T_n)$
- $$D_n = \mathbf{1}_{\{V_{T,n} < T_n\}} \sim \text{Bern}(1, p)$$
- Si T_n es el valor del pasivo de la contraparte, entonces $V_{t,n}$ se puede entender como la variable latente que determina la fallida y se puede definir una variable D_n para toda n tal que el evento 1 ocurre con probabilidad $p = P(V_{T,n} < T_n)$

- La variable r_n se define como $\ln(V_{T,n}/V_{0,n})$ y se interpreta como el rendimiento logarítmico de los activos
- El modelo factorial de Merton visto anteriormente también se suele presentar aplicando transformaciones cuadráticas en β_i (β_i^2) y cambiando la variable X_n por r_n (los rendimientos logarítmicos), de modo que el modelo queda de la siguiente manera:

$$r_n = \beta_n Y_n + \sqrt{1 - \beta_n^2} \varepsilon_n$$

- En esta formulación, Y_n es un factor que representa las variables sistemáticas, que están correlacionadas, ε_n son las variables idiosincráticas y r_n son los rendimientos logarítmicos de los activos. Por lo tanto, β_n se puede considerar como la sensibilidad de r_n al riesgo sistemático (la parte de la varianza de r_n explicada por la variable sistémica Y_n) y captura la correlación entre r_n e Y_n
- Se asume que las variables sistemáticas tienen la forma $Y_n = \mathbf{a}_n' \mathbf{X}$, donde \mathbf{X} es un vector de factores comunes que se distribuyen $\mathbf{X} \sim N_K(\mathbf{0}, \Omega)$ con $p > m$, donde Ω es una matriz de correlación (la matriz de varianzas y covarianzas). En este caso, $a_{n,k}$ representa la dependencia del deudor n en un sector regional o industrial k (representado por el factor X_k), y se asume que la suma cuadrada de estas $a_{n,k}$ es igual a uno

$$Y_n = \sum_{k=1}^K a_{n,k} X_k \quad \sum_{k=1}^K a_{n,k}^2 = 1$$

- Debido a que $\text{Var}(r_n) = 1$ y a que Y_n y $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ son variables normales estándar independientes, la varianza de los rendimientos es unitaria:

$$\text{Var}(Y_n) = \sum_{k=1}^K a_{n,k}^2 \text{Var}(X_k) = 1$$

$$\Rightarrow \text{Var}(r_n) = \beta_n^2 \text{Var}(Y_n) + (1 - \beta_n^2) \text{Var}(\varepsilon_n) = \beta_n^2 + 1 - \beta_n^2 = 1$$

- A partir de esto, es posible reescribir la variable D_n como de la siguiente manera, donde t_n es el umbral correspondiente a T_n después de cambiar $V_{T,n}$ por r_n :

$$D_n = \mathbf{1}_{\{r_n < t_n\}} \sim \text{Bern}(1, P_n)$$

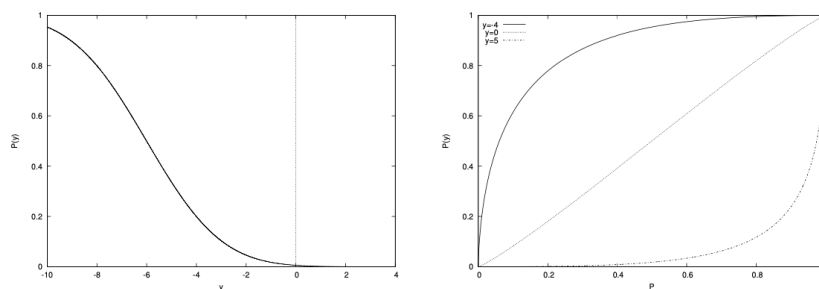
- Asumiendo un horizonte temporal $T = 1$, entonces se denota la probabilidad de falla a un año para n como P_n . Por lo tanto, se

tiene que $P_n = P(r_n < t_n)$ y, como $r_n \sim N(0,1)$, se puede definir $t_n = \Phi^{-1}(P_n)$

- Por lo tanto, la probabilidad de fallida para un deudor n condicional a una realización de Y_n (como y_n) se puede escribir de la siguiente manera:

$$P_n(y_n) = P(D_n = 1 | Y_n = y_n) = P(r_n < t_n | Y_n = y_n) = \\ = \Phi\left(\frac{t_n - \beta_n y_n}{\sqrt{1 - \beta_n^2}}\right)$$

- La probabilidad condicional de fallida tiene una relación negativa con el factor, mientras que tiene una relación positiva con t_n . Además, la probabilidad de fallida incondicional guarda una relación positiva con la condicional, aunque esta es exponencial para valores grandes de y_n y logarítmica para valores pequeños de y_n



- El Basel II presenta fórmulas de ponderación del riesgo para el cálculo del capital regulatoria para las pérdidas inesperadas basadas en el modelo asintótico de un solo factor de riesgo o *asymptotic single risk factor model* (ASFR), en el cual se puede adaptar el modelo de Merton multifactorial
 - Este modelo se construyó de manera que el capital requerido para cualquier préstamo con riesgo no dependa de la descomposición de la cartera particular a la que se aplica
 - A esta propiedad se le llama invarianza de cartera o *portfolio invariance*, y fue necesaria por razones de aplicabilidad apoyadas en cálculos simples y rápidos de requerimientos de capital
 - No obstante, la invarianza de cartera conlleva desventajas porque hace que el reconocimiento de efectos de diversificación sea difícil
 - Como juzgar si un préstamo se ajusta bien a la cartera existente requiere conocimiento específico de la cartera (contradiciendo

la propiedad), entonces se asume que la cartera está bien diversificada en el modelo ASFR

- Realmente, ningún banco real cumple con esta suposición, por lo que se espera que se tenga en cuenta la existencia de riesgo de concentración (segundo pilar)
- El modelo ASFR se basa en la ley de grandes números, dado que, si una cartera ordinaria consiste de un gran número de exposiciones de diferentes tamaños, pero se distribuyen de la misma manera, el riesgo idiosincrático se diversifica
 - Por lo tanto, los riesgos idiosincráticos se cancelan entre ellos y el riesgo sistemático es lo único que afecta a las exposiciones. A este tipo de carteras se les llama infinitamente granulares o *infinity fine-grained*, las cuales no existen en la realidad porque los bancos tienen un número finito de deudores y distribuciones desiguales de las exposiciones
 - Aunque la suposición pueda ser aproximadamente válida para algunas carteras, para la mayoría no es satisfactoria. Por lo tanto, cargos de capital cualesquiera calculados bajo esta suposición tienden a subestimar el capital requerido para una cartera real finita, y eso hace que los bancos tengan que tener en cuenta este riesgo idiosincrático no diversificado (en el segundo pilar)
- Considerando una cartera de N préstamos con riesgo, se puede expresar la pérdida como un porcentaje de la exposición total (y no en valor absoluto) de la siguiente manera:

$$s_n = \frac{E_n}{\sum_{n=1}^N E_n}$$

- Entonces, la pérdida porcentual de la cartera \mathcal{L} se expresa de la siguiente manera:

$$\mathcal{L} = \sum_{n=1}^N s_n L_n D_n$$

- Una cartera es infinitamente granular si satisface las siguientes condiciones:

$$\sum_{n=1}^N E_n \rightarrow \infty \text{ as } N \rightarrow \infty$$

$$\exists \zeta > 0 \text{ such that } \max(s_n) \text{ is } O\left(N^{-\frac{1}{2}-\zeta}\right)$$

- Las suposiciones son suficientes para garantizar de que la parte porcentual con la mayor exposición simple de la exposición total de la cartera se desvanece cuanto mayor es el número de exposiciones en la cartera
- De manera práctica, estas suposiciones son muy débiles y casi cualquier cartera grande de un banco las cumple
- En el modelo ASRF, todos los riesgos sistemáticos en una cartera se modelan por un solo factor de riesgo sistemático que suele representar el ciclo económico
 - Por lo tanto, las dos suposiciones del modelo son que las carteras son infinitamente granulares y que la dependencia entre exposiciones solo depende de un factor sistémico Y y de nada más
- Bajo todas las condiciones anteriores, el siguiente teorema se deriva directamente de la ley de grandes números: bajo las dos suposiciones del modelo ASFR, la pérdida porcentual de la cartera condicional para cualquier realización y del factor de riesgo sistémico Y satisface la siguiente condición

$$\mathcal{L} - E[\mathcal{L}|Y = y] \xrightarrow{a.s} 0 \text{ as } N \rightarrow \infty$$

- En términos intuitivos, este teorema expresa que mientras la parte porcentual de exposición de cada activo en la cartera tiende a cero, el riesgo idiosincrático de la cartera se diversifica perfectamente (dado que $E[\mathcal{L}|Y = y]$ no incluye el término ε)
- Esta cartera, cuando se tiende al límite $N \rightarrow \infty$, se denomina infinitamente granular o cartera asintótica
- Si se quiere un modelo práctico para el cálculo del capital regulatorio, primero es necesario encontrar un modo de derivar las probabilidades condicionales de fallida, lo cual se lleva a cabo a través de una adaptación del modelo de Merton
 - El comité del Basel tomó la suposición de que los factores de riesgo se distribuyen normalmente

- En este caso, el modelo tiene un solo factor de riesgo, cuya correlación con los rendimientos logarítmicos viene dada por la raíz de la sensibilidad logarítmica

$$r_n = \sqrt{\beta_n}Y + \sqrt{1 - \beta_n}\varepsilon_n$$

- Además, la correlación entre los rendimientos de los activos se puede expresar de la siguiente manera:

$$\begin{aligned}\rho(r_n, Y) &= \frac{Cov(r_n, Y)}{\sqrt{Var(r_n)}\sqrt{Var(Y)}} = E(r_n Y) - E(r_n)E(Y) = \\ &= E[(\sqrt{\beta_n}Y + \sqrt{1 - \beta_n}\varepsilon_n)Y] = \sqrt{\beta_n}E(Y^2) = \sqrt{\beta_n}\end{aligned}$$

- La esperanza condicional de las pérdidas totales de la cartera en el modelo viene dada por la siguiente expresión:

$$\begin{aligned}\rho(r_n, r_m) &= \frac{Cov(r_n, r_m)}{\sqrt{Var(r_n)}\sqrt{Var(r_m)}} = E(r_n r_m) - E(r_n)E(r_m) = \\ &= E[(\sqrt{\beta_n}Y + \sqrt{1 - \beta_n}\varepsilon_n)(\sqrt{\beta_m}Y + \sqrt{1 - \beta_m}\varepsilon_m)] = \\ &= \sqrt{\beta_n}\sqrt{\beta_m}\end{aligned}$$

- Para la cuantificación del riesgo de crédito, es posible utilizar el cuantil VaR_α a un nivel de confianza α . En el contexto de este modelo y teniendo en cuenta el teorema visto anteriormente, es posible obtener la siguiente propiedad:

$$VaR_\alpha - E[\mathcal{L}|Y = q_{(1-\alpha)}(Y)] \xrightarrow{a.s} 0 \text{ as } N \rightarrow \infty$$

- Esto se debe a que el cuantil- α de la distribución de las pérdidas totales \mathcal{L} de la cartera es el cuantil- $(1 - \alpha)$ de la distribución del factor Y
- Debido al teorema anterior, \mathcal{L} tiende casi seguramente a la esperanza condicional de \mathcal{L} para una realización de Y , de modo que esta esperanza también tiende en distribución a la distribución de \mathcal{L} . En consecuencia, si VaR_α es un cuantil de \mathcal{L} y la distribución de Y tiene una relación negativa con \mathcal{L} (las pérdidas aumentan cuanto menor es Y), entonces el cuantil- α de \mathcal{L} corresponderá a un cuantil- $(1 - \alpha)$ de la distribución de Y (aquel para obtener unas pérdidas iguales al VaR_α , ya que Y es lo único que determina las pérdidas en el modelo)

$$E(\mathcal{L}|Y = y) = \sum_{n=1}^N s_n L_n \Phi\left(\frac{t_n - \sqrt{\beta_n} y}{\sqrt{1 - \beta_n}}\right)$$

$$\Rightarrow E[\mathcal{L}|Y = q_{(1-\alpha)}(Y)] = \sum_{n=1}^N s_n L_n \Phi\left(\frac{t_n - \sqrt{\beta_n} [q_{(1-\alpha)}(Y)]}{\sqrt{1 - \beta_n}}\right)$$

- A través del teorema y de la expresión anterior, y sabiendo que Y se distribuye normalmente, el VaR_α se puede expresar de manera similar. A este se le llama VaR_α^A y tiene la siguiente expresión:

$$VaR_\alpha^A = \sum_{n=1}^N s_n L_n \Phi\left(\frac{t_n - \sqrt{\beta_n} \Phi^{-1}(\alpha)}{\sqrt{1 - \beta_n}}\right)$$

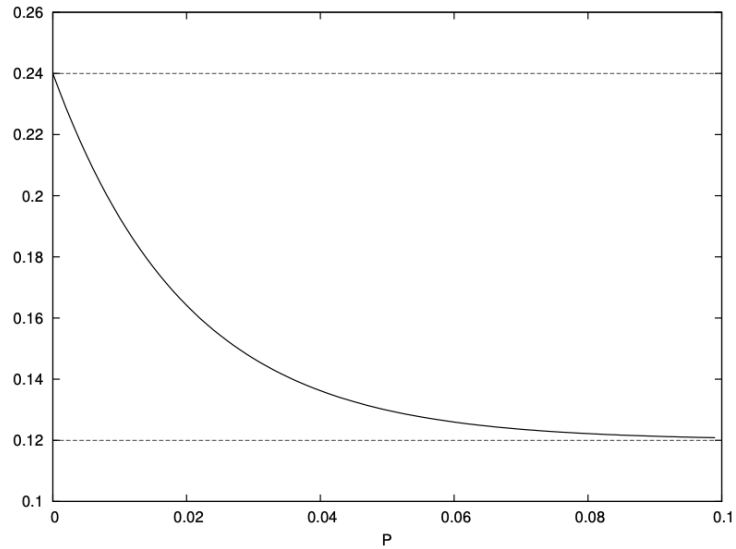
- Las contribuciones al VaR_α , por lo tanto, se puede calcular utilizando el VaR_α^A a través de la siguiente fórmula:

$$VaRC_{\alpha,n}^A = s_n \frac{\partial VaR_\alpha^A}{\partial s_n} = s_n L_n \Phi\left(\frac{t_n - \sqrt{\beta_n} \Phi^{-1}(\alpha)}{\sqrt{1 - \beta_n}}\right)$$

- El VaR_α no es una medida coherente del riesgo debido a que no es necesariamente una medida subaditiva. No obstante, como en el modelo ASFR se utiliza $E[\mathcal{L}|Y = q_{(1-\alpha)}(Y)]$, y la esperanza es un operador lineal, el VaR_α^A es perfectamente subaditivo
- El parámetro β_n describe el grado de la exposición del deudor al factor de riesgo sistemático por medio de la correlación entre activos, y está determinado por la clase de activos del deudor. En el Basel II, la correlación del valor de los activos para una clase de activos dada, representada por la probabilidad de fallida incondicional P_n , se da por la siguiente expresión:

$$\beta_n = 0.12 \frac{1 - e^{-50P_n}}{1 - e^{-50}} + 0.24 \left(1 - \frac{1 - e^{-50P_n}}{1 - e^{-50}}\right)$$

- Por lo tanto, se ve como la correlación β_n decrece cuando incrementa P_n . Esto se basa en evidencia empírica y en la intuición económica: cuanto mayor es P_n , mayores son los componentes del riesgo idiosincrático de un deudor (el riesgo de fallida depende menos del estado de la economía)



- Otro de los intereses principales a la hora de gestionar el riesgo bajo el modelo ASFR es poder obtener la distribución de pérdidas de una cartera infinitamente granular
 - Asumiendo una cartera homogénea, en donde todos los deudores tienen la misma probabilidad de fallida $P_n = P$ para $n = 1, 2, \dots, N$ y se utiliza una pérdida dada la fallida unitaria $L_n = 1$ para $n = 1, 2, \dots, N$, eso permite ver que se cumplen las siguientes propiedades:

- El parámetro β_n será el mismo para cualquier deudor $n = 1, 2, \dots, N$, dado que $P_n = P$

$$\beta_n = \beta$$

- Las pérdidas de la cartera condicionales a una realización de Y se pueden expresar de la siguiente manera:

$$E(\mathcal{L}|Y = y) = \sum_{n=1}^N s_n \Phi\left(\frac{t_n - \sqrt{\beta_n} \Phi^{-1}(\alpha)}{\sqrt{1 - \beta_n}}\right) = \Phi\left(\frac{t_n - \sqrt{\beta_n} \Phi^{-1}(\alpha)}{\sqrt{1 - \beta_n}}\right)$$

- Por lo tanto, la pérdida porcentual \mathcal{L} de la cartera tiende a la probabilidad de fallida condicional $P(y)$ de manera casi segura cuando $N \rightarrow \infty$. Esto hace que en una cartera infinitamente granular se puede describir la fracción de deudores en situación de fallida con $P(y)$

$$E(\mathcal{L}|Y = y) - P(y) \xrightarrow{a.s} 0 \quad \text{as } N \rightarrow \infty$$

- Denotando el número de deudores en situación de fallida en la cartera como \mathcal{L} , entonces se puede calcular la función de distribución de probabilidad en el modelo ASFR $F_{\mathcal{L}}^A$ de \mathcal{L} de la siguiente manera:

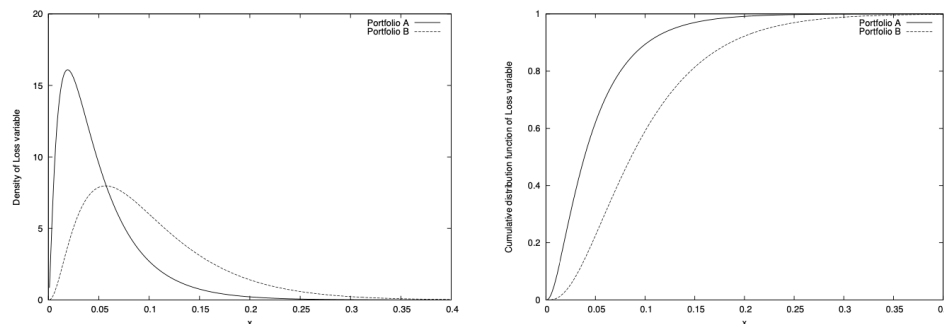
$$\begin{aligned}
F_L^A(x) &= P(\mathcal{L} \leq x) = P[P(y) \leq x] = \\
&= P \left[-y \leq \frac{1}{\sqrt{\beta}} [\Phi^{-1}(x)\sqrt{1-\beta} - \Phi^{-1}(P)] \right] = \\
&= \Phi \left(\frac{1}{\sqrt{\beta}} [\Phi^{-1}(x)\sqrt{1-\beta} - \Phi^{-1}(P)] \right) \text{ for } x \in [0,1]
\end{aligned}$$

- Diferenciando la función respecto x , se obtiene la función de densidad de probabilidad f_L^A

$$f_L^A = \frac{\partial F_L^A}{\partial x} = \sqrt{\frac{1-\beta}{\beta}} e^{-\frac{1}{2\beta}(\Phi^{-1}(x)\sqrt{1-\beta} - \Phi^{-1}(P))^2} e^{\frac{1}{2}(\Phi^{-1}(x))^2}$$

for $x \in [0,1]$

- Las funciones de densidad y de distribución de probabilidad muestra como, cuanto mayor es la probabilidad de falla, mayor será el riesgo de la cartera



- Esto se puede reflejar fácilmente a través de calcular el VaR_α^A para la cartera, ya que será mayor para una probabilidad de falla P mayor porque hace que las colas de la distribución sean mucho más gruesas
- Cuando se genera la distribución de pérdidas de la cartera, se combinan las distribuciones marginales de pérdidas de los deudores para formar una distribución multivariante que captura la dependencia de falla entre deudores
 - El enfoque IRB del Basel II utiliza la cópula Gaussiana de un factor para calcular el cuantil- α de la distribución de $E(\mathcal{L}|Y = y)$, la aproximación analítica del VaR_α en el modelo ASFR
 - Al aplicar esta cópula para calcular el capital regulatorio para el riesgo de crédito, se asume implícitamente que la distribución

gaussiana multivariante modela bien el riesgo de cola de las carteras de crédito, aunque realmente no lo hace bien

La gestión del riesgo de carteras de crédito: riesgo de concentración

- La experiencia histórica muestra que la concentración del riesgo en las carteras de activos han sido una de las causas principales de las fallidas de bancos, y se ha demostrado la relevancia del riesgo de concentración sectorial
 - Bajo el marco del modelo ASFR, se asume que las carteras de los bancos son infinitamente granulares, de modo que el capital económico solo depende del riesgo sistémico
 - Las carteras reales no son infinitamente granulares, de modo que cuando hay concentraciones de nombre de exposición, habrá un riesgo idiosincrático residual no diversificado, y la fórmula IRB subestimaré el capital económico
 - Además, la suposición de un solo factor de riesgo no permite una medición explícita del riesgo de concentración sectorial
 - Los riesgos de concentración suelen surgir de una distribución desigual de préstamos a un solo deudor (llamada concentración de nombre) o diferentes industrias o sectores regionales (llamada concentración de país o de sector)
 - Además, ciertas dependencias como vínculos directos de negocios entre diferentes deudores pueden aumentar el riesgo de crédito en una cartera debido a que la fallida de un deudor puede causar la fallida de un segundo deudor dependiente. A esto se le llama contagio de fallida, y está vinculado al riesgo de concentración de nombre y de país o de sector
 - Un índice o medida de concentración para una cartera de N préstamos debería satisfacer las siguientes propiedades:
 - La reducción de la exposición de un préstamo y un incremento de la exposición de un préstamo más grande no debe hacer decrecer el riesgo de concentración (principio de transferencia)
 - La medida de concentración debe tomar su valor mínimo cuando todos los préstamos tienen una medida equivalente (principio de distribución uniforme)
 - Si dos carteras, compuestas por el mismo número de préstamos, satisface que el tamaño agregado de los k préstamos más

grandes de la primera cartera es mayor o igual que el tamaño agregado de las k carteras más grandes de la segunda cartera para $1 \leq k \leq N$, entonces la misma desigualdad se debe mantener para las medidas de concentración de ambas carteras (criterio de Lorenz)

- Si dos o más préstamos se combinan, la medida de concentración no puede decrecer (superaditividad)
 - Considerando una cartera que consiste de préstamos de igual tamaño, la medida de concentración no debe incrementar con el incremento del número de préstamos (independencia de la cantidad de préstamos)
 - Garantizar un préstamo adicional de una cantidad relativamente pequeña no debe incrementar la medida de concentración (irrelevancia a pequeñas exposiciones)
 - Es posible demostrar que, si se cumple la primera y la última propiedad de las anteriores, entonces se satisfacen las seis propiedades
- Las medidas de concentración más utilizadas son la razón de concentración, la curva de Lorenz, el coeficiente de Gini y el índice de Herfindahl-Hirschman. Este último es el más usado, de modo que se desarrolla el análisis en base a este
 - El índice de Herfindahl-Hirschman (HHI) se define como la suma de las cuotas de mercado (denotadas por MS) al cuadrado de cada participante del mercado, medidas como fracciones de la cartera total

$$HHI = \sum_{i=1}^N (MS)_i^2$$

- Esta es una medida continua en donde el cero corresponde al caso infinitamente granular (cada participante tiene una cuota infinitesimal) y el uno corresponde al monopolio (solo hay un participante). Manteniendo todo lo otro igual, mientras más cerca esté el HHI de una cartera a uno, más concentrada estará la cartera
- Esta medida satisface las seis propiedades expuestas anteriormente. No obstante, ni el HHI ni ninguna de las medidas mencionadas anteriormente puede incorporar el efecto de las cualidades específicas de un acreditado (como las probabilidades de fallida), de modo que se han desarrollado métodos basados en modelos (como los anteriores)

- Normalmente se emplean simulaciones de Monte Carlo para poder generar la distribución de pérdidas empírica y así determinar el VaR , el ES o las contribuciones de riesgo para una cartera crediticia
- Suponiendo que se genera una distribución de pérdidas para una cartera de N deudores por una simulación de $\mathcal{L} = \sum_{i=1}^N \mathcal{L}_i$, el estimador por Monte Carlo del VaR_α es el estimador del cuantil

$$VaR_\alpha = \inf \{l : \hat{F}_\mathcal{L} \geq \alpha\} \quad \text{where} \quad \hat{F}_\mathcal{L} = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{\mathcal{L}_n \leq l\}}$$

- La distribución empírica $\hat{F}_\mathcal{L}$ es una función por partes, por lo que el estimador de cuantiles para el Monte Carlo puede obtenerse a través del siguiente algoritmo:
- Si se quiere calcular el ES_α es más directo, definido de la siguiente manera:

$$\widehat{ES}_\alpha = \frac{1}{1-\alpha} \frac{1}{N} \sum_{n=1}^N \mathcal{L}_n \mathbf{1}_{\{\mathcal{L}_n \leq \widehat{VaR}_\alpha\}}$$