Aplicación de Cópulas al Arbitraje Estadístico con Pares

Análisis Cuantitativo

1. Introducción

El arbitraje estadístico con pares (*pairs trading*) tradicional se basa en modelos de cointegración o regresión. Las cópulas ofrecen un enfoque más sofisticado al modelar la **estructura de dependencia completa** entre dos activos, capturando relaciones no lineales y dependencias de cola que los métodos tradicionales pasan por alto.

2. Marco Teórico

2.1. Definición de Cópula

Una cópula C es una función de distribución multivariada con marginales uniformes en el intervalo $[0, 1]$. Según el **Teorema de Sklar**, para dos variables aleatorias X y Y con funciones de distribución conjunta F_{XY} y marginales F_X y F_Y, existe una cópula C tal que:

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) \tag{1}$$

La densidad de la cópula se define como:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \tag{2}$$

donde $u = F_X(x)$ y $v = F_Y(y)$.

3. Proceso de Implementación

3.1. Paso 1: Preprocesamiento de Datos

3.1.1. Transformación de Precios

Para un par de activos A y B, calculamos los rendimientos logarítmicos:

$$r_{A,t} = \ln(P_{A,t}) - \ln(P_{A,t-1}) \tag{3}$$

$$r_{B,t} = \ln(P_{B,t}) - \ln(P_{B,t-1}) \tag{4}$$

4. Transformación a Uniformes

4.1. Marginales Empíricas

Para cada activo, estimamos la distribución marginal de rendimientos:

$$\hat{F}_A(r) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{r_{A,t} \leq r\}} \tag{5}$$

4.2. Transformación a Uniformes

Aplicamos la transformada integral de probabilidad:

$$u_t = \hat{F}_A(r_{A,t}), \quad v_t = \hat{F}_B(r_{B,t}) \tag{6}$$

Donde:

- $r_{A,t}, r_{B,t}$: **Rendimientos observados** (datos originales)
- $\hat{F}_A, \hat{F}_B$: **Funciones de distribución acumulada** empíricas o paramétricas
- u_t, v_t: **Variables uniformes** en [0,1] resultantes de la transformación

4.3. Paso 2: Modelado con Cópulas

4.3.1. Familias de Cópulas Common

- **Gaussiana**:

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{7}$$

- ***t* de Student**:

$$C_{\rho,\nu}(u, v) = t_{\rho,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v)) \tag{8}$$

- **Clayton**:

$$C_\theta(u, v) = \left(\max\{u^{-\theta} + v^{-\theta} - 1, 0\}\right)^{-1/\theta} \tag{9}$$

- **Gumbel**:

$$C_\theta(u, v) = \exp\left(-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{1/\theta}\right) \tag{10}$$

4.3.2. Estimación de Parámetros

Los parámetros se estiman mediante máxima verosimilitud:

$$\hat{\theta} = \arg\max_\theta \sum_{t=1}^{T} \ln c(u_t, v_t; \theta) \tag{11}$$

4.4. Paso 3: Señales de Trading

4.4.1. Probabilidades Condicionales

Definimos las señales usando derivadas de la cópula:

$$\begin{aligned}\text{Probabilidad condicional para venta:} \quad & P(V > v|U = u) = 1 - \frac{\partial C(u,v)}{\partial u} \\ \text{Probabilidad condicional para compra:} \quad & P(V \leq v|U = u) = \frac{\partial C(u,v)}{\partial u}\end{aligned}$$

4.4.2. Reglas de Trading

- **Señal VENTA** (Vender A/Comprar B):

$$P(V > v|U = u) < \tau_{\text{low}} \tag{12}$$

- **Señal COMPRA** (Comprar A/Vender B):

$$P(V \leq v|U = u) < \tau_{\text{low}} \tag{13}$$

donde τ_{low} es un umbral típicamente entre 0.05 y 0.10.

4.5. Paso 4: Gestión de Riesgos

4.5.1. Salida por Take-Profit

$$P(V > v|U = u) \approx 0{,}5 \quad \text{o} \quad P(V \leq v|U = u) \approx 0{,}5 \tag{14}$$

4.5.2. Salida por Stop-Loss

- **Basado en probabilidad**: La probabilidad condicional se vuelve más extrema
- **Basado en precio**: Pérdida máxima definida sobre el spread

5. Ejemplo Numérico

5.1. Escenario

- Par: Acción A vs. Acción B
- Cópula: t-Student con $\rho = 0{,}8$, $\nu = 5$
- Umbral: $\tau_{\text{low}} = 0{,}05$

5.2. Señal de Trading

$$
\begin{aligned}
u_t &= 0{,}98 \quad \text{(A tiene rendimiento muy alto)} \\
v_t &= 0{,}60 \quad \text{(B tiene rendimiento moderado)} \\
P(V > 0{,}60 | U = 0{,}98) &= 0{,}03 < 0{,}05
\end{aligned}
$$

Decisión: Vender A/Comprar B

6. Consideraciones Prácticas

6.1. Ventajas

- Captura dependencias no lineales y de cola
- interpretación probabilística sólida de las señales
- Mejor gestión del riesgo en eventos extremos

6.2. Desafíos

- Complejidad en la estimación y selección del modelo
- Requiere suficiente data histórica
- Mayor costo computacional