

El problema de selección de carteras: modelos y algoritmos

Enriqueta Vercher González

Departamento de Estadística e Investigación Operativa.
Universitat de València

Facultad de CC. Económicas y Empresariales.
Universidad de Málaga
4 de Noviembre de 2016

Preliminares

- El problema de seleccionar una **cartera de inversión** tiene como objetivo determinar qué cantidad hemos de invertir en cada activo de manera satisfactoria para el inversor, teniendo en cuenta la **incertidumbre** asociada al comportamiento de los mercados financieros.
- La primera formulación matemática del problema de seleccionar una cartera se debe a Harry Markowitz, quien en 1952 estableció que:

Selección de una cartera

*El proceso de seleccionar una cartera puede ser dividido en dos fases. La **primera etapa** comienza con la observación y la experiencia, y concluye con las estimaciones (convicciones) sobre el futuro comportamiento de los activos disponibles. La **segunda etapa** comienza con la información relevante de la primera etapa acerca del rendimiento futuro y finaliza con la elección de la cartera.*

Números fuzzy y distribución de posibilidad

Dada una cartera $X = (x_1, x_2, \dots, x_N)$, la lógica fuzzy permite también modelizar la incertidumbre del rendimiento mediante cantidades fuzzy o distribuciones de posibilidad/credibilidad. Alternativamente:

- La incertidumbre del rendimiento del activo j -ésimo puede modelarse, por ejemplo, mediante **números fuzzy de tipo LR**, $\tilde{R}_j = (a_{lj}, a_{uj}, c_j, d_j)_{LR}$ para $j = 1, 2, \dots, N$, y el rendimiento total de la cartera como combinación de los rendimientos de los activos individuales:

$$\tilde{R}_P(x) = \sum_{j=1}^N x_j \tilde{R}_j,$$

- La incertidumbre del rendimiento de la cartera X puede aproximarse directamente mediante una **cantidad fuzzy**, P_X , que incorpore la relación contemporánea que guardan entre si los rendimientos de los N activos mediante una única función de pertenencia.
- Dado que un número fuzzy induce una **distribución de posibilidad** que coincide con su función de pertenencia (Zadeh'78), podemos estimar los valores esperados de rendimiento y riesgo de una cartera mediante los momentos posibilísticos de $\tilde{R}_P(x)$ o de P_X , respectivamente. También podemos aproximar una **distribución de credibilidad** a los datos históricos.

Optimización lineal semi-infinita

Para rendimientos fuzzy LR con funciones de referencia tipo potencia, \tilde{R}_j , y un nivel esperado de rendimiento trapezoidal \tilde{R}_f , el modelo (FMDR) tiene una estructura semi-infinita:

Modelo de PSIL

$$\begin{aligned}
 \text{(FMDR)} \quad & \text{Min} \quad \sum_{j=1}^N (B_j - A_j + s_{A_j} \frac{p_j}{p_j+1} + s_{B_j} \frac{q_j}{q_j+1}) x_j \\
 & \text{s.a.} \quad \sum_{j=1}^N (A_j - s_{A_j} (1 - \alpha)^{\frac{1}{p_j}}) x_j \geq R_l - c_0 (1 - \alpha) \quad \alpha \in [h, 1] \\
 & \quad \sum_{j=1}^N (B_j + s_{B_j} (1 - \alpha)^{\frac{1}{q_j}}) x_j \geq R_u + d_0 (1 - \alpha) \quad \alpha \in [h, 1] \\
 & \quad \sum_{j=1}^N x_j = 1 \\
 & \quad 0 \leq x_j \leq u_j
 \end{aligned}$$

- El cumplimiento de la restricción fuzzy se controla mediante el valor de $h \in [0, 1]$, permitiendo la incorporación de la opinión del inversor.
- El conjunto factible de este problema es acotado y su resolución se alcanza aplicando condiciones de optimalidad KKT bajo la cualificación de restricciones de Slater para [programación semi-infinita lineal](#) (Vercher'08).

Rendimiento fuzzy de una cartera

Rendimiento fuzzy de una cartera: P_X

Dada una cartera $X = (x_1, x_2, \dots, x_N)$ se calculan los rendimientos históricos, $\{r_t(X)\}_{t=1}^T$:

$$r_t(X) = \sum_{j=1}^N r_{jt} x_j$$

donde el rendimiento r_{jt} del activo j -ésimo se calcula como $r_{jt} = \frac{p_{j(t+1)} - p_{jt}}{p_{jt}}$, siendo p_{jt} el precio del activo j en el periodo t . La incertidumbre sobre el rendimiento se modeliza mediante un número fuzzy de tipo LR: $P_X = (p_l, p_u, c, d)_{L\pi R\rho}$.

- En lugar de agregar las incertidumbres individuales de los activos, aproximamos directamente la incertidumbre del rendimiento de la cartera.
- Los **rendimientos** $r_t(X)$ incorporan información implícita sobre las relaciones contemporáneas entre los diferentes activos (Bermúdez *et al'*07).
- Los valores de los **momentos de la distribución de posibilidad** no son parámetros conocidos en el problema de optimización, alternativamente se evalúan una vez conocidos los valores de la variable de decisión X .

Algoritmo genético para el modelo CCMDR

El algoritmo genético trabaja con **carteras factibles** e incorpora un criterio de dominancia. Maneja **dos poblaciones**: la actual siempre del mismo tamaño M , y el conjunto de soluciones no-dominadas que puede tener diferentes tamaños. Cada nueva generación contiene todas las soluciones no dominadas de la generación anterior.

- Aplica una estrategia **elitista**, selecciona las carteras más cercanas a la **frontera superior** a las que aplica un operador mutación aleatorio (con pesos proporcionados por un procedimiento de aprendizaje).
- Operador de **mutación aleatorio-ponderado**. Para las carteras de la élite se modifican ligeramente las proporciones asociadas a dos activos, $x_i > 0$ y $x_j \geq 0$. El activo i se selecciona aleatoriamente, la probabilidad de selección de j depende del número medio de veces en que haya sido seleccionado en la élite en generaciones anteriores. En el nuevo individuo, las proporciones son: $x'_i = x_i - p_0$ y $x'_j = x_j + p_0$, con p_0 tal que se mantienen la **cardinalidad y la factibilidad**. Se aplica un número determinado de veces en cada cartera de la élite.
- El algoritmo para si la distancia entre dos fronteras superiores sucesivas es menor que una constante, o si llega al máximo de generaciones permitido.

Modelo de media-riesgo lateral-asimetría (MDRS)

Se incorpora un tercer objetivo (coeficiente de asimetría) cuando los rendimientos no están simétricamente distribuidos alrededor de la media. La formulación matemática del modelo es (Vercher|Bermúdez'13):

$$\begin{aligned}
 \text{(MDRS)} \quad & \max \quad \tilde{E}(P_X) = \frac{\rho_u + \rho_l}{2} + \frac{d}{2} \frac{\rho}{\rho+1} - \frac{c}{2} \frac{\pi}{\pi+1} \\
 & \min \quad \tilde{w}(P_X) = \rho_u - \rho_l + d \frac{\rho}{\rho+1} + c \frac{\pi}{\pi+1} \\
 & \max \quad S(P_X) = \frac{\mu_3(P_X)}{w(P_X)^3} \\
 \text{s.a.} \quad & \sum_{j=1}^N x_j = 1 \\
 & 0 \leq l_j \leq x_j \leq u_j \quad j = 1, \dots, N \\
 & K_l \leq c(X) \leq K_u
 \end{aligned}$$

- Se ha resuelto parcialmente este problema mediante el genético anterior para dos problemas bi-objetivo (DRS_M) y (MDR_S), en donde el tercer objetivo se incorporaba al modelo como restricción con una cota inferior.
- Para aproximar el frente óptimo de Pareto del modelo MDRS se han utilizado tres algoritmos evolutivos: NSGA-II (Deb et al|02), MOEA/D (Zhang|Li'07) y Global WASF-GA (Ruiz|Saborido|Luque'16).
- Se han adaptado operadores de mutación, recombinación y reparación para mantener la factibilidad de las soluciones generadas por los algoritmos EMO (Saborido et al'16).

Métrica C

La comparación de los valores medios de la métrica C alcanzados entre las aproximaciones encontradas por las versiones clásicas de los métodos y las que utilizan los nuevos operadores, para las 50 ejecuciones, son:

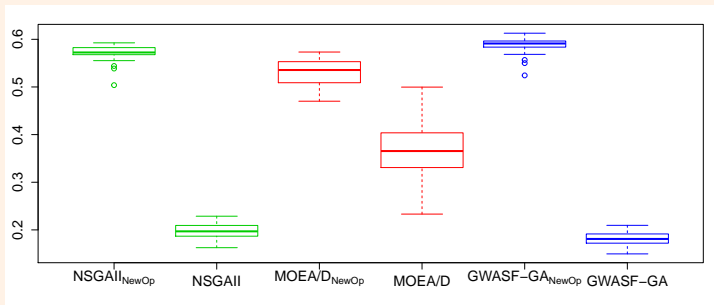
- $C(\text{NSGA-II}_{\text{Nuevo}}, \text{NSGA-II})=100$
- $C(\text{GWASF-GA}_{\text{Nuevo}}, \text{GWASF-GA})=100$
- $C(\text{MOEA/D}_{\text{Nuevo}}, \text{MOEA/D})=39.125$ y $C(\text{MOEA/D}, \text{MOEA/D}_{\text{Nuevo}})=4.394$ (test t: p-valor $< 10^{-10}$)

Comparación de los valores medios de la métrica C alcanzados entre las aproximaciones encontradas por las nuevas versiones de los algoritmos: $\text{NSGAII}_{\text{Nuevo}}$, $\text{MOEA/D}_{\text{Nuevo}}$ y $\text{GWASF-GA}_{\text{Nuevo}}$.

C(A,B)	$\text{NSGAII}_{\text{Nuevo}}$	$\text{MOEA/D}_{\text{Nuevo}}$	$\text{GWASF-GA}_{\text{Nuevo}}$
$\text{NSGAII}_{\text{Nuevo}}$	–	21.236	31.327
$\text{MOEA/D}_{\text{Nuevo}}$	25.999	–	28.571
$\text{GWASF-GA}_{\text{Nuevo}}$	14.884	11.020	–

Hipervolumen

Diagrama de caja del hipervolumen (HV) alcanzado en las 50 ejecuciones de los tres procedimientos (NSGAII, MOEA/D y GWASF-GA), con los operadores clásicos y con los nuevos operadores propuestos.



Hipervolumen

Comparaciones dos a dos de los hipervolumenes medios, según el algoritmo utilizado.

Algoritmo	Oper. clásicos		Oper. nuevos		p-valor
	Media	D. típica	Media	D. típica	
NSGAI	0.198	0.016	0.572	0.015	$< 2.2e - 16$
MOEA/D	0.369	0.056	0.529	0.028	$< 2.2e - 16$
GWASF-GA	0.182	0.014	0.588	0.015	$< 2.2e - 16$

La comparación de los HV medios obtenidos con los operadores nuevos mediante un test ANOVA (asumiendo homogeneidad de varianzas), presentan diferencias estadísticamente significativas ($p_v < 2.2e - 16$). Las comparaciones múltiples dos a dos:

	NSGAI _{Nuevo}	MOEA/D _{Nuevo}
MOEA/D _{Nuevo}	$< 1.0e - 5$	
GWASF-GA _{Nuevo}	$2.0e - 4$	$< 1.0e - 5$

Conclusiones

Para la **primera fase** del problema de selección de carteras, hemos propuesto:

- Considerar los rendimientos de cada cartera como conjunto de datos histórico, en lugar de considerar la agregación de la incertidumbre de los rendimientos de los activos individualmente.
- Aproximar la incertidumbre del rendimiento de una cartera dada mediante **números fuzzy de tipo LR**, con funciones de referencia de la familia de potencias, cuyos parámetros se obtienen a partir de los percentiles muestrales.

En cuanto a la **fase de selección** de carteras óptimas, proponemos:

- Aplicar algoritmos evolutivos (adaptados con nuevos operadores) para obtener la frontera de Pareto del problema de selección de carteras con **restricción de cardinalidad**, independientemente del procedimiento que haya sido utilizado para obtener la información en la primera fase.
- Proporcionar carteras eficientes que cumplen con los requisitos impuestos por el inversor, clasificadas según su perfil de aversión al riesgo.