

# 2do Trabajo De Investigación / Informe Ejecutivo:

## “A rapidly converging artificial bee colony algorithm for portfolio optimization”

Víctor Álvarez  
Universidad Técnica  
Federico Santa María  
Vitacura, Chile  
victor.alvarezb@usm.cl

Constantino Mavrakis  
Universidad Técnica  
Federico Santa María  
Vitacura, Chile  
constantino.mavrakis@usm.cl

Matías Otth  
Universidad Técnica  
Federico Santa María  
Valparaíso, Chile  
matias.otth@sansano.usm.cl

Gabriel Vergara  
Universidad Técnica  
Federico Santa María  
San Joaquín, Chile  
gabriel.vergara@usm.cl

### I. INTRODUCCIÓN

La siguiente investigación tiene como objetivo replicar los resultados del artículo “A rapidly converging artificial bee colony algorithm for portfolio optimization” en una base de datos diferente y actualizada.

Este trabajo consta de determinar la combinación óptima de activos para conformar un portafolio. Para esto, se toma como base el modelo denominado “Mean-Variance” de Markowitz el cual, permite la maximización de retornos y minimización del riesgo esperado. Por lo tanto, se suele recurrir a solucionar un problema de optimización de portafolio (PO).

Si bien, MV es una formulación usada para resolver PO, carece de elementos que permitan resolver un problema de características reales, como lo son las limitaciones en las cantidades de acciones en un portafolio o un rango de proporciones a invertir. En consecuencia, la literatura ha permitido reformular MV al considerar restricciones que acoten las limitaciones mencionadas, obteniendo así el modelo “Cardinality Constrained Mean-Variance” (CCMV).

A partir de lo anterior, el artículo de referencia busca resolver CCMV-PO. Sin embargo, debido a la naturaleza restricta y no lineal de CCMV-PO, este puede presentar mayores complejidades al momento de resolverlo para problemas de gran envergadura. Por lo tanto, heurísticas de resolución eficientes son necesarias para enfrentar estas dificultades. Una de estas son los algoritmo genético (GA).

Básicamente, GA permite simular las características evolutivas presentes en la biología, para obtener nuevas soluciones mediante operaciones de selección, cruce y mutación de soluciones anteriores en el cual, se busca la convergencia al óptimo tener múltiples generaciones. Como se puede apreciar, esta heurística permite una resolución iterativa para un dominio grande, con menor información y sin las limitaciones. Sin embargo, también posee mayores

tiempos de resolución y la posibilidad de caer en óptimos locales. En consecuencia, metodologías basados en insectos se han estudiado para favorecer el enfoque de estas técnicas. La que se va a usar en este trabajo es la optimización de CCMV-PO mediante la heurística “Artificial Bee Colony”(ABC). Esta replica el comportamiento de una colonia para buscar y explotar fuentes de alimentos mediante la simulación de tres tipos de abejas: trabajadoras, observadoras y exploradoras.

Tal como se menciona a inicios de la introducción, en esta investigación se pretende replicar a modo general, los resultados obtenidos en el texto de referencia. En este, los autores utilizan ABC para resolver el problema CCMV-PO como algoritmo evolutivo. Cabe mencionar que ellos utilizan una leve variación en el comportamiento de las abejas que permiten el traspaso de información con más eficiencia entre sus pares y la regeneración de soluciones antiguas. Estos cambios fueron difíciles de aplicar en esta investigación debido a la reciente experiencia con estos algoritmos. Por lo tanto, no se van a considerar en este artículo.

A partir de lo anterior, se va a aplicar la heurística ABC para optimizar un portafolio mediante CCMV-PO en una base de datos reciente de seis diferentes mercados. Además, se van a sensibilizar los hiperparámetros de ABC para comparar el comportamiento obtenido con el trabajo de referencia. Posteriormente, se comparará mediante métricas de evaluación, el rendimiento de ABC contra GA.

En la siguiente sección se va a detallar la metodología aplicada. Posteriormente, se van a presentar los modelos utilizados. Después, se describirá el caso experimental y se entregarán los resultados obtenidos. Finalmente, se concluirá la investigación.

## II. METODOLOGÍA

### II-A. Datos

Uno de los cambios más importantes que se realizan en esta investigación con respecto al artículo de referencia es la base de datos utilizada.

Los autores utilizan datos recolectados desde marzo de 1992 a septiembre de 1997 en la base provista por “OR-Library”. De aquí, seleccionan siete índices distintos: Hang Seng con 31 stocks, DAX100 con 85 stocks, FTSE100 con 89 stocks, SP100 con 98 stocks, Nikkei225 con 225 stocks, XU030 con 30 stocks y XU100 con 39 stocks. Estos fueron escogidos por ellos ya que, les permite comparar con otras investigaciones realizadas con la misma información.

Para esta investigación, “OR-Library” no proporciona información actualizada de estos índices. Por lo tanto, la base de datos utilizada se conforma por los obtenidos en la plataforma “Yahoo Finance”. Hay que aclarar de que se logran recolectar a lo más, 30 stocks para cada índice por limitaciones plataforma. Además, no se logra recolectar información para FTSE100, SP100, Nikkei225, Xu030 y XU100. En consecuencia, se cambian estos índices y se reducen la cantidad de mercados a analizar.

A partir de lo anterior, se recolectan los precios diarios desde enero de 2010 hasta diciembre de 2018. Los índices seleccionados son seis: Hang Seng con 25 stocks, DAX con 27 stocks, FTSE con 30 stocks, NASDAQ con 22 stocks, CAC con 28 stocks, IBEX con 29 stocks.

### II-B. Cardinality Constrained Mean-Variance Model

El modelo clásico MV propuesto por Markowitz, busca minimizar la desviación estándar como medida de riesgo de un portafolio. Si bien, su formulación entrega las herramientas básicas para resolver un problema de portafolio, carece de aspectos que simulen limitaciones reales. Por lo tanto, **Chang et al.** introducen el modelo CCMV que enriquece la formulación al considerar aspectos como aversión al riesgo y la cardinalidad de un portafolio.

$$\text{Min } \lambda \left[ \sum_{i=1}^N \sum_{j=1}^N x_i x_j \sigma_{ij} \right] - (1 - \lambda) \left[ \sum_{i=1}^N x_i \mu_i \right] \quad (1)$$

$$\text{s.t. } \sum_{i=1}^N x_i = 1 \quad (2)$$

$$\sum_{i=1}^N z_i = K \quad (3)$$

$$\varepsilon_i z_i \leq x_i \leq \delta_i z_i, \quad i = 1, \dots, N \quad (4)$$

$$z_i \in \{0, 1\}, \quad i = 1, \dots, N \quad (5)$$

$N$  es la cantidad de activos dentro de un índice de mercado.  $x_i$  es una variable que representa la proporción a invertir en

un activo  $i$ .  $\sigma_{ij}$  es la covarianza entre los retornos de los activos  $i$  y  $j$ .  $\mu_i$  es la media de los retornos del activo  $i$ .  $z_i$  es una variable binaria que representa si se invierte o no en un activo  $i$ .  $K$  es la cantidad deseada de activos dentro de un portafolio.  $\varepsilon_i$  y  $\delta_i$  son los límites mínimos y máximos que puede tomar  $x_i$ .  $\lambda$  es la aversión al riesgo que un inversionista puede tener dentro de un portafolio.

En (1) se busca minimizar la varianza del riesgo y maximizar los retornos esperados. Es claro que la aversión al riesgo  $\lambda$  juega un rol importante en la función de objetivo ya que, va a permitir darle un peso a los riesgos o retornos, para un inversionista. En (2) se establece que las proporciones de todos los activos deben sumar 1 en el portafolio. En (3) se restringe la cardinalidad de elementos que debe contener el portafolio. Finalmente, en (4) y (5) se establece el dominio de las variables  $x_i$  y  $z_i$ .

### II-C. Algoritmos genéticos

Los algoritmos genéticos son heurísticas que replican el principio de evolución Darwiniana. En este, se sigue el principio de “los mejores individuos sobrevivirán” lo que, corresponde al proceso evolutivo de reproducción.

Básicamente en GA, se trabaja con una población inicial de candidatos que generalmente se llaman cromosomas. Cada uno de estos cromosomas son una representación de una combinación de variables que entregan un resultado en específico. Este resultado se denomina función de ajuste el cual, suele ser la función objetivo de un problema. Posteriormente, los cromosomas pueden cambiar su contenido mediante tres operadores genéticos: selección, cruce y mutación. En consecuencia, se obtienen nuevos cromosomas en lo que se desea obtener una función de ajuste mayor a la inicial y se conforma una nueva generación. Finalmente, se repite este proceso de tal manera de que los mejores cromosomas converjan al óptimo en múltiples generaciones. El algoritmo entonces puede resumirse en los siguientes pasos.

- **Paso 1:** Generar una población inicial de forma aleatorio.
- **Paso 2:** Evaluar la función de ajuste de cada cromosoma.
- **Paso 3:** Aplicar los operadores genéticos para cruzar la población.
- **Paso 4:** Actualizar la población con las nuevas generaciones.
- **Paso 5:** Repetir el paso 2 hasta que se consiga un criterio de termino.

Para el caso de esta investigación, se definirán los elementos utilizados.

#### 1. Población inicial

Para la población inicial de cromosomas, se escoge una cantidad de 100.

## 2. Función de ajuste

La función de ajuste va a ser representada por la ecuación (1) que minimiza el riesgo.

## 3. Cromosomas

Los cromosomas van a ser representados por un por una tupla de las  $k$  proporciones  $x_i$  del portafolio, el identificador de la generación y el valor de la función de ajuste.

## 4. Selección

La selección de los cromosomas se realiza mediante un torneo binario. Los cromosomas se seleccionaran aleatoriamente y el que posea mejor función de ajuste sera seleccionado.

## 5. Cruce

El cruce entre dos posibles cromosomas corresponde a generar un nuevo cromosoma mediante a la captura de elementos en la tupla de los cromosomas padres. De este manera, se logra evaluar una nueva función de ajuste que tendrá el cromosoma hijo. En específico para esta investigación se utilizó cruce aritmético. Este operador va a ocurrir con una tasa de cruce  $t_c = 0,4$ .

## 6. Mutación

La mutación corresponde a cambios aleatorios que se producen a un o varios elementos internos de un cromosoma. Para el caso de este trabajo, se va a cambiar las proporción de una acción seleccionado o se va a cambiar el identificador de la acción y mantener la proporción original. Este operador va a ocurrir con una tasa de mutación  $t_m = 0,6$ , pero en las etapas finales esta tasa disminuye a 0,1.

## 7. Reemplazo

Se va a utilizar la estrategia “Elite population replacement strategy”. En esta, sólo permanece la población de cromosomas con los mejores resultados de la función de ajuste dentro de la población original. La población elite con respecto a la población inicial es el 30 % superior.

## 8. Criterio de Termino

Para el criterio de finalización del algoritmo, se definen 1000 generaciones para cada índice de mercado.

problemáticas.

La heurística ABC es un tipo de variación de GA que busca imitar el comportamiento de una colonia de abejas al buscar fuentes de alimentación. Por lo tanto, se modela las habilidades de búsqueda, comunicación y explotación mediante a abejas artificiales. Existen tres tipos de abejas: empleadas, observadoras y exploradoras. Las primeras dos abejas juegan el rol de seleccionar fuentes de comidas con más nutrientes para su explotación mientras que, la tercera abeja se encarga de localizar nuevas fuentes de comidas cuando las ya utilizadas son consumidas.

El algoritmo inicia al generar fuentes de comida aleatorias. Después, cada abeja empleada se moverá entre dos fuentes de comida aleatoria. Aquí, se toma una medición del néctar (valor de solución) en la fuente seleccionada y comparten su información con las abejas observadoras. Posteriormente, estas ultimas decidirán si ir hacia esa fuente u otra con una con una probabilidad dada. Estás mejorarán la solución y la explotaran. Una vez consumidas las fuentes de comidas, las abejas exploradoras toman el rol de memorizar las mejores fuentes obtenidas y buscar nuevas en el dominio de solución. Este proceso se repite hasta cumplir con un criterio de finalización.

El algoritmo se ABC utilizado en está investigación se puede resumir en los siguientes pasos.

- **Paso 1:** Generar fuentes de alimentación aleatoriamente.
- **Paso 2:** En caso de que una fuente no cumpla con las restricciones iniciales, repararla .
- **Paso 3:** Enviar a la abejas empleadas a las fuentes de comida para seleccionar una de forma aleatoria y recolectar información.
- **Paso 4:** Enviar las abejas observadoras a seleccionar, mediante probabilidades, dos fuentes de comidas, una con la información provista por la empleada y otra aleatoria.
- **Paso 5:** Mejorar y seleccionar las mejores fuentes de comidas hasta cumplir con un criterio de salida.
- **Paso 6:** Cuando se cumple el criterio de salida, llamar a las abejas exploradoras para buscar nuevas fuentes de comidas.
- **Paso 7:** Repetir desde el paso 2 hasta que se cumpla el criterio de finalización.

Para el caso de esta investigación, se definirán los siguientes elementos utilizados.

### 1. Función Néctar

La función que representa el néctar de cada solución es inspirada por la ecuación (1). Sin embargo, se agregan las fuentes de comidas  $f \in F$  y las iteraciones  $t \in M$ .

## II-D. Artificial Bee Colony

Tal como se menciona en la introducción, la implementación de GA es útil para resolver problemas complejos como CCMV-PO. sin embargo, esta misma heurística puede presentar ineficiencias en los tiempos de resolución o convergencias en mínimos locales. Es por esto que otros algoritmos son desarrollados para enfrentar estas

$$nc_f^t = \lambda \left( \sum_{i \in S_f^t} \sum_{j \in S_f^t} x_{fi}^t x_{fj}^t \sigma_{ij} \right) - (1 - \lambda) \sum_{i \in S_f^t} x_{fi}^t \mu_i \quad (6)$$

Es claro que no todos los activos dentro de un índice de mercado sean seleccionados para ser parte de la solución néctar. Por lo tanto, va a existir un subconjunto  $S \subset N$  de las acciones seleccionadas del total posible en ese mercado. A partir de esto, es necesario definir otro subconjunto que represente las acciones seleccionadas para una fuente de comida e iteración en específica,  $S_f^t$ .

## 2. Representación de solución

Las soluciones van a ser representadas por una tupla de las  $k$  proporciones  $x_i$  del portafolio y el valor de su función néctar.

## 3. Abejas empleadas

Estas abejas son representadas por la sigla  $E$  las cuales, se dedican a explorar las fuentes de comidas existentes e informar sobre su valor nutricional a las abejas observadoras. Su mecanismo de selección de fuentes de comida es aleatorio.

## 4. Abejas observadoras

Estas abejas son representadas por la sigla  $O$  las cuales, se dedican a seguir el descubrimiento de las abejas empleadas con una probabilidad la cual, depende del valor nutricional de la solución encontrada. El criterio de selección es bajo ruleta. Posterior a la selección se dedican a explotar y mejorar las fuentes de comidas.

## 5. Abejas exploradoras

Estas abejas poseen el rol de memorizar las fuentes de comidas ya consumidas y buscar nuevas soluciones. Su trabajo inicia cuando, las otras abejas no logra mejorar las soluciones después de un cierta cantidad de iteraciones, *limit*. En otras palabras, el rol de estas abejas es no estancarse en mínimos locales.

## 6. Fuentes de comidas

Las soluciones o fuentes de comidas son representadas por sigla  $F$ . Estas son posibles soluciones del problema CCVM-PO que son accesibles por las abejas en cada iteración y su cantidad depende netamente de la cantidad de abejas en la colonia.

$$F = E + O + 10 \quad (7)$$

## 7. Criterio de finalización

Para el criterio de termino del algoritmo, se definen una cantidad  $M$  de iteraciones las cuales, van a ser sensibilizadas en el desarrollo experimental.

Como se puede apreciar, ABC posee características de diversificación e intensificación de soluciones.

Finalmente, el algoritmo base utilizado en la investigación proviene de la librería de python “pyMetaheuristic” creado por “Valdecy Pereira” y la “University of Chinese Academy of Sciences”.

## II-E. Métricas de rendimiento

En este artículo, se utilizan las siguientes métricas propuestas en el trabajo de referencia para facilitar la comparación. Estas son: “Mean Euclidean Distance” (MEUCD), “Variance of Return Error” (VRE) y “Mean Return Error” (MRE). Además, se utiliza el test estadístico “Wilcoxon Signed Rank” (WS) para visualizar la diferencia estadísticas entre los técnicas utilizadas.

### 1. MEUCD

$$MEUCD = \frac{\sum_{i=1}^{\mathcal{E}} \sqrt{(X_i^* - x_i) + (Y_i^* - y_i)}}{\mathcal{E}} \quad (8)$$

Donde  $(X_i^*, Y_i^*)$  es el punto más cercano de EFS con respecto a EFH,  $(x_i, y_i)$ .

### 2. VRE

$$VRE = \frac{\sum_{i=1}^{\mathcal{E}} 100 |X_i^* - x_i| / x_i}{\mathcal{E}} \quad (9)$$

Donde  $(X_i^*, Y_i^*)$  es el punto más cercano de EFS con respecto a EFH,  $(x_i, y_i)$ .

### 3. MRE

$$MRE = \frac{\sum_{i=1}^{\mathcal{E}} 100 |Y_i^* - y_i| / y_i}{\mathcal{E}} \quad (10)$$

Donde  $(X_i^*, Y_i^*)$  es el punto más cercano de EFS con respecto a EFH,  $(x_i, y_i)$ .

### 4. WS

Este test evalúa las diferencias estadísticas entre modelos, plantea una hipótesis nula de mediana igual a 0, que equivaldría a que métodos comparados no demuestran diferencias significativas de uno con respecto a otro. En caso de determinarse ciertas asimetrías en las diferencias, esto asumiría la superioridad en rendimiento de un modelo con respecto a otro.

$$WS = \sum_T I_+(d_t) \text{rank}(|d_t|), \quad I_+(d_t) = 1, \quad \text{si} : d_t > 0 \quad (11)$$

### III. DESARROLLO EXPERIMENTAL Y RESULTADOS.

Esta investigación engloba dos experimentos: El primero busca determinar los parámetros óptimos para ABC mientras que en el segundo, se busca comparar el rendimiento de ABC y GA .

Para ambos experimentos, se utilizan los siguientes datos de aversión al riesgo: Hang Seng con  $\lambda = 0,93$ , DAX con  $\lambda = 0,96$  , FTSE  $\lambda = 0,66$  , NASDAQ con  $\lambda = 0,86$  , CAC con  $\lambda = 0,86$  , IBEX con  $\lambda = 0,72$ . Cada uno de estos parámetros fueron optimizados para diversificar los portafolios resultantes ya que, se desea evitar la inversión mayoritaria en solo un activo.

Además, todos los resultados son computados con lo siguientes valores:  $K = 8$ ,  $\varepsilon_i = 0,01 (i = 1, \dots, N)$  y  $\delta_i = 1 (i = 1, \dots, N)$ . Por otro lado, se ejecuta los experimentos cinco veces y los resultados presentados corresponden al promedio de estos.

#### III-A. Experimento 1: Sensibilización de parámetros ABC

En esta sección, se busca sensibilizar los factores correspondientes al numero de abejas empleadas  $E$ , observadoras  $O$ , la cantidad de alimentos  $F$ , el limite para llamar a las abejas exploradoras  $limit$  y la cantidad de iteraciones del algoritmo  $M$ .

Para iniciar la sensibilización, se decide partir con los parámetros iniciales usados en el texto de referencia:  $E = 5$ ,  $O = 5$ ,  $F = 20$ ,  $limit = 200$  y  $M = 1000$ .

Primero, se busca determinar la proporción óptima de abejas tipo  $E$  y  $O$ . Por lo tanto, se generan 10 casos al cambiar solo las cantidades de abejas de la siguiente forma (E,O): (10,0), (9,1), (8,2), (7,3), (6,4), (5,5), (4,6), (3,7), (2,8), (1,9), (0,10). Para esto, se ejecuta la heurística ABC, se recolecta los resultados de la función objetivo y se normalizan para su comparación.

En la **Figura 1** se muestran los resultados obtenidos de la sensibilización. Se puede apreciar el promedio el cual, muestra la proporción óptima para las abejas empleadas que es 6 y para las observadoras es 4.

Lo anterior tiene sentido ya que, la existencia de ambas abejas permite al comunicación eficiente de las soluciones encontradas. Sin embargo, también es necesario de que exista un balance preciso en las proporciones para maximizar los resultados. Esto se puede ver al analizar los rendimientos en los casos extremos. Para los primeros tres casos donde hay una mayor cantidad de abejas empleadas, los rendimientos obtenidos son los peores ya que, existe la posibilidad de que las abejas observadoras no sean las suficientes para explotar el traspaso de información y recaen en la incertidumbre al seleccionar las fuentes de nutrición. Por otro lado, al aumentar la proporción de abejas observadoras después

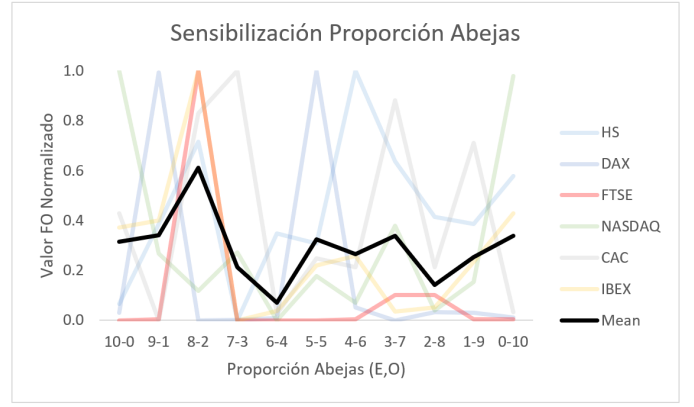


Figura 1. Resultados al obtener los valores normalizados de la función objetivo para los seis mercados estudiados. Además, se entrega el promedio de todas las proporciones.

del punto óptimo, la calidad de la solución comienza a degradarse lo que, permite aseverar de que tener una cantidad de saludable de abejas empleadas es igual de importante para el algoritmo.

Al comparar con el texto de referencia, llama la atención que la proporción (5,5) no entrega los mejores resultados. Esto se puede deber a los siguientes tres factores que diferencian este trabajo con el de referencia: La base de datos estudiada es sustancialmente distinta, la cantidad de veces que se ejecuta cada caso es menor y que el algoritmo ABC utilizado es levemente ineficiente. De todas maneras, la proporción de abejas que se va a utilizara para los próximas sensibilizaciones es (6,4).

La sensibilización del parámetro  $limit$  es la siguiente que se va a realizar. Es importante ver su comportamiento ya que, indica el numero de iteraciones necesarias para que las abejas exploradoras sean invocadas. Por lo tanto, se espera que para valores bajos de  $limit$ , estas abejas sean mayormente utilizadas. En consecuencia, se toman los siguientes valores para el análisis: 0, 50, 100, 150, 200, 250, 300, 350 y 400.

En la **Figura 2** se aprecia que el valor óptimo de iteraciones  $limit$  es 50. Esto quiere decir que el algoritmo ABC se beneficia de una utilización recurrente de las abejas exploradoras lo que, refleja un escape de los mínimos locales más a menudo. Por otro lado, es claro que si se utilizan estas abejas en cada iteración, no se logra una convergencia a la optimalidad del problema. También se logra visualizar que al aumentar este parámetro hasta 300, se observa una leve degradación del rendimiento. Sin embargo, al pasar este punto, se vuelve a mejorar el rendimiento.

Lo anterior no permite entregar una conclusión concreta sobre el comportamiento del rendimiento cuando se aumenta considerablemente el valor  $limit$ . Aunque, según la literatura, un crecimiento en este parámetro debiese degradar el rendimiento de la heurística ya que, se puede caer en

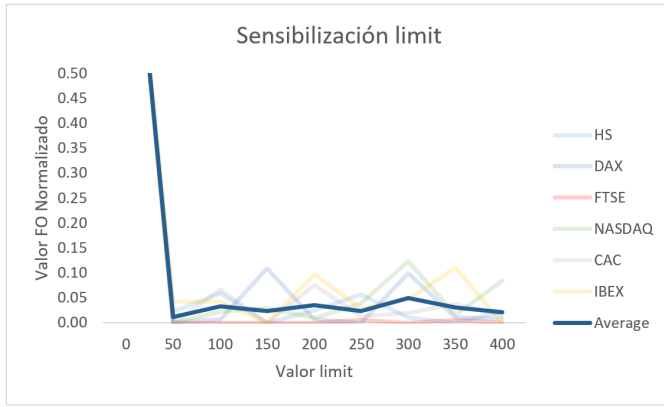


Figura 2. Resultados al obtener los valores normalizados de la función objetivo para los seis mercados estudiados. Además, se entrega el promedio de los límites utilizados para cada caso.

mínimos locales al no explorar nuevas soluciones. Es posible que los resultados obtenidos sean justificados por la cantidad reducida de acciones dentro de cada índice de mercado. En consecuencia, el problema que está resolviendo el algoritmo posee menores dificultades y logra encontrar soluciones buenas en las primeras iteraciones. De todas maneras, se va a fijar *limit* en 50 para las futuras sensibilizaciones.

El siguiente parámetro a analizar es la cantidad de abejas que va a poseer la colonia. La intuición inicial es que, un aumento en la cantidad de abejas debiese significar en una mejor convergencia en los resultados y posee un mayor tiempo de resolución. Esto también es obtenido en el artículo de referencia. Para el caso de esta investigación, dado a que el número de elementos dentro de cada mercado llega hasta un máximo de 30 activos, se varía la cantidad de abejas desde  $(E, O) = (6, 4)$  hasta  $(18, 12)$ . Por lo tanto, se realizan cinco casos y se presentan a continuación.

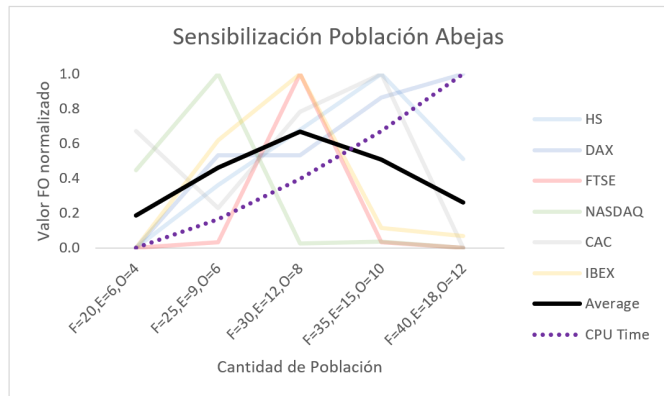


Figura 3. Resultados al obtener los valores normalizados de la función objetivo para los seis mercados estudiados. Además, se entrega el promedio de las cantidades de abejas para cada caso y su fuentes de comida.

En la **Figura 3**, se puede apreciar un comportamiento distinto al esperado. La mejor solución se encuentra al

definir la colonia con los parámetros  $E = 6$ ,  $O = 4$  y  $F = 20$ . Al aumentar el tamaño de la población, se obtiene un comportamiento parabólico en donde, los mejores dos resultados se obtienen en los casos extremos. Es decir, el peor rendimiento se encuentra en  $E = 12$ ,  $O = 8$  y  $F = 30$ . Si se decide aumentar o disminuir la colonia, se consigue mejoras en el rendimiento. Con respecto a los tiempos computacionales, se observa una tendencia creciente que se alinea con lo esperado.

Se debe recalcar que no se esperaba conseguir el comportamiento observado. Sin embargo, distintos factores pueden dar indicios de sus motivos.

Primero, las fuentes de alimentos se posicionan aleatoriamente en cada ejecución del código, lo que significa que las cinco corridas realizadas no son suficientes para generar cambios significativos en los valores semillas en los casos donde se encuentran tendencias crecientes. Análogamente, para el caso de población reducida, se encuentran valores semillas que catalizan el rendimiento de la función objetivo en gran parte de los mercados.

Segundo, al utilizar una base de datos diferente al del texto de referencia, fue necesario obtener nuevos valores de  $\lambda$  que, pueden afectar el rendimiento de los portafolios. Para el caso de Hang Seng y DAX, ambos poseen tasas superiores al 90 %. En consecuencia, en (1) se entrega un peso superior a la volatilidad y no se logra reducir el valor de la función objetivo al aumentar la cantidad de la colonia.

Tercero, definir solo cinco escenarios para el tamaño de las abejas puede no ser suficiente para ver el comportamiento general de la sensibilización. Por lo tanto, se debería evaluar para una mayor cantidad de abejas a las expuestas. Sin embargo, Se puede observar que los tiempos computacionales son crecientes y se vuelven impedimentos considerables para solucionar el problema para una tamaño superior de la colonia.

Al considerar los factores descritos, una profundización de la investigación es necesaria. Por lo tanto, en la sección del experimento 2, se van a utilizar dos poblaciones de abejas que son: baja y alta densidad. Mayores detalles se van a entregar en esa sección. Mientras tanto, se va a utilizar el resultado óptimo para la sensibilización restante, es decir, considerar una población de baja densidad con  $E = 6$ ,  $O = 4$  y  $F = 20$ .

El último parámetro que queda por sensibilizar es la cantidad de iteraciones  $M$ . Este define cuantas generaciones va a poseer la heurística ABC. Se toman los siguientes valores: 1000, 2000 y 4000. Se espera obtener resultados alineados con el texto de referencia. Es decir, un aumento en las cantidades de generaciones no debería generar una mejora significativa en los resultados pero si un aumento de los tiempos computacionales.

Cuadro I  
SENSIBILIZACIÓN DEL PARÁMETRO  $M$ , CON SUS RESPECTIVOS VALORES DE FUNCIÓN OBJETIVO Y TIEMPOS COMPUTACIONALES

Base de datos		Función Objetivo			Tiempos (s)		
		M = 1000	M = 2000	M = 4000	M = 1000	M = 2000	M = 4000
HS	Media	<b>-2.991.E-05</b>	-2.989.E-05	-2.956.E-05	<b>117.10</b>	223.93	436.06
	Std. Dev	4.214.E-07	<b>3.031.E-07</b>	8.490.E-07	5.43	<b>5.09</b>	17.32
	Max	-2.926.E-05	-2.935.E-05	<b>-2.823.E-05</b>	<b>124.20</b>	229.74	464.62
	Min	<b>-3.057.E-05</b>	-3.017.E-05	-3.055.E-05	<b>108.77</b>	217.10	416.88
DAX	Media	<b>-1.474.E-05</b>	-1.400.E-05	-1.467.E-05	<b>115.51</b>	232.86	462.71
	Std. Dev	7.283.E-08	<b>1.321.E-06</b>	7.710.E-08	<b>6.73</b>	10.58	24.09
	Max	<b>-1.462.E-05</b>	-1.136.E-05	-1.457.E-05	<b>122.89</b>	249.19	492.46
	Min	<b>-1.485.E-05</b>	-1.477.E-05	-1.474.E-05	<b>106.45</b>	217.69	420.01
FTSE	Media	<b>-1.885.E-04</b>	-1.882.E-04	-1.877.E-04	<b>107.85</b>	220.34	437.56
	Std. Dev	1.141.E-07	5.584.E-07	<b>1.078.E-06</b>	<b>6.55</b>	9.74	28.21
	Max	<b>-1.883.E-04</b>	-1.871.E-04	-1.857.E-04	<b>118.64</b>	229.40	485.77
	Min	<b>-1.886.E-04</b>	<b>-1.886.E-04</b>	<b>-1.886.E-04</b>	<b>100.47</b>	206.10	402.57
Nasdaq	Media	-3.152.E-05	-3.106.E-05	<b>-3.161.E-05</b>	<b>97.92</b>	210.08	422.54
	Std. Dev	3.441.E-07	5.979.E-07	<b>1.156.E-06</b>	<b>5.54</b>	15.43	13.06
	Max	<b>-3.098.E-05</b>	-3.014.E-05	-2.941.E-05	<b>103.93</b>	231.56	439.26
	Min	-3.205.E-05	-3.162.E-05	<b>-3.273.E-05</b>	<b>89.40</b>	193.84	402.94
CAC	Media	-2.828.E-05	<b>-2.883.E-05</b>	-2.723.E-05	<b>114.78</b>	236.64	460.12
	Std. Dev	<b>1.276.E-06</b>	4.303.E-07	2.211.E-06	<b>5.98</b>	6.90	19.99
	Max	-2.605.E-05	<b>-2.823.E-05</b>	-2.284.E-05	<b>121.38</b>	246.84	494.24
	Min	<b>-2.958.E-05</b>	-2.948.E-05	-2.872.E-05	<b>105.68</b>	226.29	439.34
IBEX	Media	-8.884.E-05	-8.798.E-05	<b>-8.959.E-05</b>	<b>118.15</b>	230.46	471.84
	Std. Dev	6.246.E-07	7.799.E-07	<b>2.995.E-07</b>	<b>3.01</b>	10.58	28.52
	Max	-8.820.E-05	-8.671.E-05	<b>-8.916.E-05</b>	<b>122.31</b>	247.16	513.34
	Min	-8.988.E-05	-8.879.E-05	<b>-8.991.E-05</b>	<b>115.31</b>	214.21	430.33

En el **Cuadro I** se muestran los resultados obtenidos. Se consiguen, en la mayoría de los casos, valores óptimos cuando se recurre a solo 1000 generaciones de la función objetivo. Por otro lado, no se observa una mejora consistente al aumentar el numero generaciones. Es probable que este resultado sea influenciado por las cinco corridas que se están haciendo para el algoritmo. Sin embargo, al comparar los resultados entre los valores  $M$ , no existen mayores diferencias. Por lo tanto, permite intuir que al aumentar la cantidad de veces de ejecución del algoritmo, se obtendrán resultados que convergen.

Al visualizar los tiempos computaciones, queda claro que los menores tiempos se obtienen cuando hay 1000 generaciones. Esto permite aseverar que la calidad de solución es buena con los parámetros obtenidos y no es necesario aumentar la cantidad de iteraciones. En consecuencia, Se escoge la cantidad de generaciones  $M$  igual a 1000.

Finalmente, Los resultados de la sensibilización se pueden resumir en lo siguiente:  $E = 6$ ,  $O = 4$ ,  $F = 20$ ,  $limit = 50$  y  $M = 1000$ .

### III-B. Experimento 2: Comparación de rendimientos de ABC y GA.

En esta sección se va a comparar el modelo propuesto con los obtenidos usando GA. Para casos prácticos y según

los resultados obtenidos en la sección anterior, se van a utilizar dos tipos de poblaciones de abejas. El primero con una población baja de abejas (ABC-LP) los cuales, sus parámetros  $F$ ,  $E$ ,  $O$  son 6, 4 y 20 respectivamente. El segundo con una población alta de abejas (ABC-LP) los cuales, sus parámetros  $F$ ,  $E$ ,  $O$  son 40, 18 y 12.

Con respecto a los parámetros GA utilizados, se decide comparar en igualdad de condiciones. Por lo tanto, se define una población inicial de 100 cromosomas y 1000 iteraciones. En la población elite se definen las tasas de mutación y cruce en 0,4 y 0,6. Además, en la etapa final, se reduce la tasa de mutación en 0,1.

La comparación entre estos algoritmos es realizada según la curva de eficiencia restringida (CCEF). Debido a que se trabaja con datos nuevos en está investigación, fue necesario calcularla manualmente con un solver GRG y variando los valores de  $\lambda$  en 0,1 para reducir los tiempos computacionales agregados. Cabe mencionar que en el texto de referencia, utilizan la librería OR-Library la cual, posee los datos de la CCEF calculados en su base de datos.

En la **Figura 4** se muestran la CCEF estándar y de los modelos utilizados. Lo primero que se puede destacar es que en todos los mercados, GA posee resultados de alto riesgo y menores retornos. Por otro lado, En los casos de HS, FTSE, NASDAQ, CAC e IBEX, ambos modelos ABC logran un

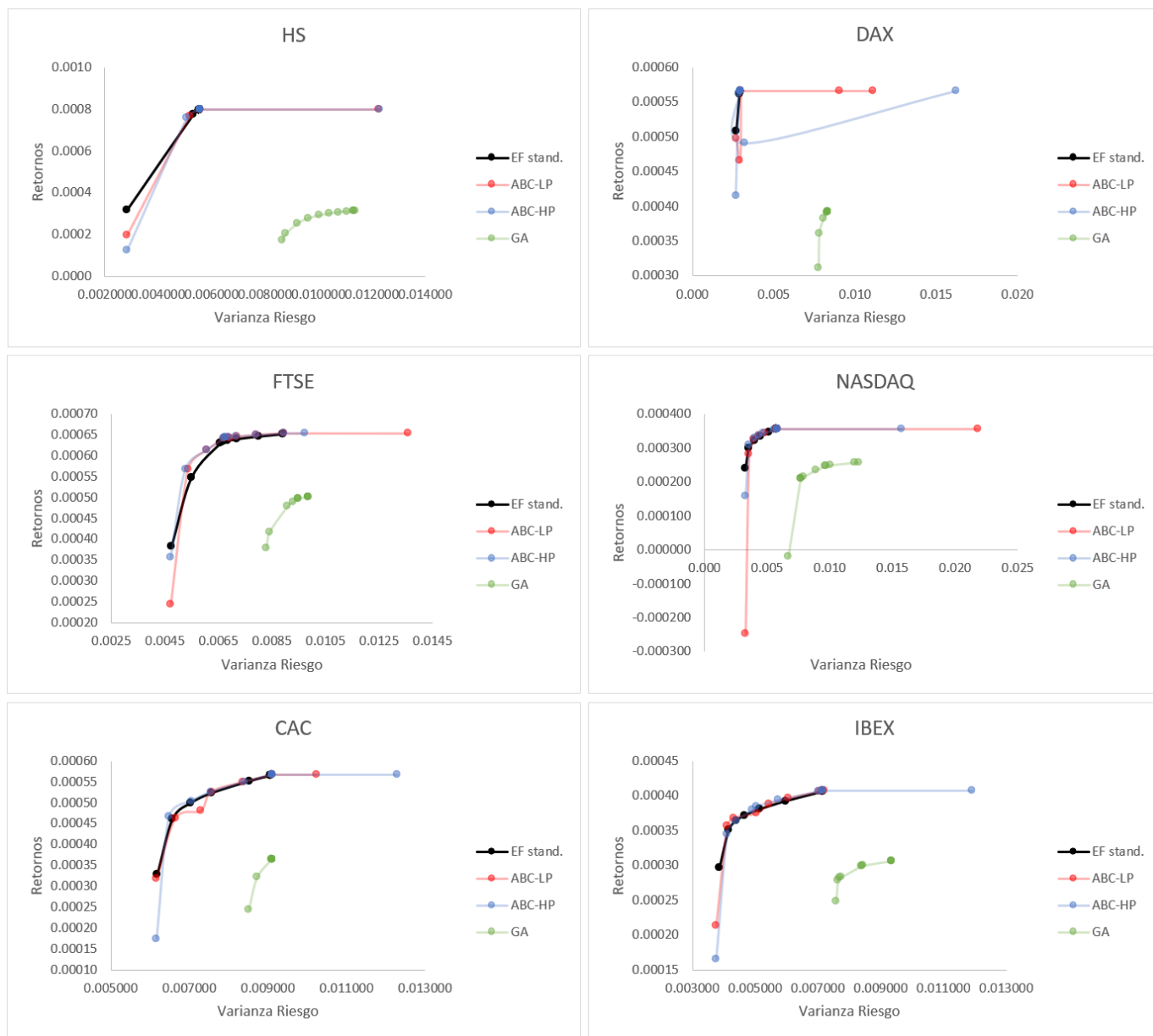


Figura 4. Resultados al obtener los valores normalizados de la función objetivo para los seis mercados estudiados. Además, se entrega el promedio de las cantidad de abejas para cada caso y su fuentes de comida.

comportamiento similar a CCEF estándar. Sin embargo, para el caso de DAX, no se logra visualizar una curva marcada. Esto puede estar justificado por la poca cantidad de puntos utilizados. Además, las curvas presentadas son solo una proyección de una función polinomial que une los puntos obtenidos. En consecuencia, es probable que en DAX existan curvas de eficiencias escalonadas.

Otro aspecto interesante es que, en algunos mercados, la CCEF estándar es sobrepasadas por los modelos ABC. Según la teoría, estas CCEF debiesen ser utópicas y las heurísticas deberían tratar de imitar su comportamiento por debajo de estas. En consecuencia, es probable que los resultados

obtenidos al resolver el problema original con GRG no son óptimos. De todas maneras, son muy próximos en su vecindad lo que sirve como buena aproximación.

Para complementar los resultados anteriores, se comparan las métricas descritas MEUCD, MRE y VRE; y presentan los valores de las funciones objetivos y tiempos conseguidos en el **Cuadro II**. Las métricas mencionadas son calculadas tomando como base CCEF estándar.

Con respecto a los valores de la función objetivo de CCVM-PO, las metodologías ABC-LP y ABC-HP logran tener resultados idénticos a CCEF en las mayoría de los



Cuadro II  
COMPARACIÓN DE MODELOS USADOS CON LA CCEF DE CADA MERCADO. SE MUESTRAN LAS MÉTRICAS MEUCD, VRE, MRE, FUNCIÓN OBJETIVO Y TIEMPOS DE RESOLUCIÓN

Base de datos	Metrica Rendimiento	CCEF	ABC-LP	ABC-HP	GA
HS	MEUCD	-	<b>0.01369</b>	0.01435	0.06407
	VRE	-	<b>11.64065</b>	11.86460	84.72079
	MRE	-	<b>3.99251</b>	6.16936	65.19760
	Valor FO	-0.00039	<b>-0.00039</b>	<b>-0.00039</b>	0.00466
	Tiempos (s)	115	115	666	<b>70</b>
DAX	MEUCD	-	0.01843	<b>0.01508</b>	0.07106
	VRE	-	44.16002	<b>41.71600</b>	176.79282
	MRE	-	<b>2.01759</b>	3.08905	32.49549
	Valor FO	-0.00028	<b>-0.00028</b>	<b>-0.00028</b>	0.00385
	Tiempos (s)	115	115	656	<b>68</b>
FTSE	MEUCD	-	0.01431	<b>0.01075</b>	0.01906
	VRE	-	6.05342	<b>2.31712</b>	6.80130
	MRE	-	4.52352	<b>1.86579</b>	26.60843
	Valor FO	-0.00030	<b>-0.00030</b>	<b>-0.00030</b>	0.00428
	Tiempos (s)	110	110	664	<b>66</b>
NASDAQ	MEUCD	-	0.01967	<b>0.01635</b>	0.05409
	VRE	-	26.52426	<b>16.88649</b>	57.68732
	MRE	-	19.68883	<b>4.06394</b>	40.65198
	Valor FO	-0.00017	<b>-0.00017</b>	<b>-0.00017</b>	0.00391
	Tiempos (s)	104	104	682	<b>67</b>
CAC	MEUCD	-	<b>0.01025</b>	0.01208	0.01234
	VRE	-	2.03641	3.77752	<b>0.65953</b>
	MRE	-	<b>0.91243</b>	4.73670	37.72874
	Valor FO	-0.00025	<b>-0.00025</b>	<b>-0.00025</b>	0.00429
	Tiempos (s)	115	115	757	<b>66</b>
IBEX	MEUCD	-	<b>0.00881</b>	0.01415	0.03281
	VRE	-	<b>1.72196</b>	7.66148	18.45414
	MRE	-	<b>3.28851</b>	4.60321	28.02642
	Valor FO	-0.00019	<b>-0.00019</b>	<b>-0.00019</b>	0.00387
	Tiempos (s)	115	115	732	<b>65</b>

casos mientras que, GA posee rendimientos con ordenes de magnitud menores. Si bien, los tiempos computacionales de GA son notoriamente menores, no son motivos para preferir este algoritmo sobre los otros. Sin embargo, es interesante ver que, a una misma cantidad de iteraciones, se consiguen alrededor de 50 % y 1000 % reducción en los tiempos computaciones con respecto a ABC-LP y ABC-HP. En consecuencia, la simulación original de las abejas posee ineficiencias que son bien abordadas en el texto de referencia.

Al ver el comportamientos de las métricas de error, es claro que GA posee mayores mientras que, ABC- LP y ABC- HP los menores. A partir de esto, se puede determinar que ambos modelos ABC son significativamente mejor que GA y poseen resultados similares a CCEF. Sin embargo, no queda claro si ABC-HP es superior a ABC-LP. Por lo tanto se, se utiliza Wilcoxon Signed Rank Test para determinar las diferencias estadísticas entre los modelos.

En el **Cuadro III**, se realiza el test WS al comparar ABC-HP con los otros modelos, Se demuestra que ABC-HP y ABC-LP no son significativamente distintos en todas sus métricas. Por lo tanto, ambas metodologías se pueden

considerar como equivalentes. Al comparar GA, se obtiene que el modelo ABC-HP es significativamente mejor en MEUCD, MRE y en los valores de la función objetivo lo que es razonable con los resultados previamente presentados. Sin embargo, en VRE no se obtiene una diferencia significativa. Esta métrica mide los errores en los riesgos y al revisar la **Cuadro 4**, se puede justificar estos resultados ya que se puede apreciar que los puntos de riesgos en GA se mueven dentro de los rangos de puntos obtenidos en los algoritmos ABC. Sin embargo, al considerar los otros aspectos mencionados, nuevamente queda claro que GA no es recomendable en esta investigación.

Finalizando el análisis de resultados, se puede aseverar que ABC-LP es una heurística altamente más apropiada que ABC-HP, pues posee resultados que son considerados idénticos, con una disminución en los tiempos de resolución de 530 %.

#### IV. CONCLUSIONES

En esta investigación, se realiza un análisis comprensivo de dos algoritmos evolutivos para comparar sus rendimientos.

Cuadro III  
RESULTADOS DEL WILCOXON SIGNED RANK TEST: COMPARACIÓN DEL MODELO ABC-HP CONTRA LOS RESTANTES. SE EVALÚAN MÉTRICAS MEUCD, VRE, MRE Y FUNCIÓN OBJETIVO

Metrica Rendimiento	ABC-HP	ABC-LP	GA
MEUCD	Wins(+)	12	21
	Loss(-)	9	0
	Ties	0	0
	W-value	9	0
	p-value	0.844	0.031
VRE	Wins(+)	13	20
	Loss(-)	8	1
	Ties	0	0
	W-value	8	1
	p-value	0.688	0.063
MRE	Wins(+)	10	21
	Loss(-)	11	0
	Ties	0	0
	W-value	11	0
	p-value	1.000	0.031
Función Objetivo	Wins(+)	3	21
	Loss(-)	0	0
	Ties	0	0
	W-value	0	0
	p-value	0.346	0.031

La motivación surge por el texto de referencia que aplica una heurística ABC eficiente para resolver el problema CCVM-PO, en una base de datos estudiada por la literatura y comparándolo con otros algoritmos de resolución.

Tal como se menciona en la realización de este trabajo, se toma un enfoque más acotado ya que, se utiliza una base de datos actualizada en donde, no se poseen datos para comparar con otros artículos. Por lo tanto, se aplica el algoritmo estándar de ABC, se optimizan sus parámetros y se comparan con GA para verificar mejoras en la calidad de solución y tiempos computacionales.

La investigación conlleva en una fase inicial en la cual, se obtienen las aversiones de riesgo óptimas para diversificar un portafolio de ocho activos en seis mercados distintos. A partir de esto, se sensibilizan los parámetros de ABC en la segunda fase y se obtienen resultados interesantes.

Lo primero que se puede concluir es la necesidad de obtener una composición de abejas ideal para los datos que se desean analizar. Para el caso de esta investigación, la razón entre las abejas empleadas y observadoras fue 60\40. Por otro lado, peores resultados a una razón elevada de abejas empleadas pero, también cuando no se disponía ninguna. Por lo tanto, se concluye, que las existencias de ambas abejas es importante para la resolución de estos problemas ya que, permite la comunicación entre sus pares y una exploración eficiente de sus fuentes de comidas disponibles.

Posteriormente, se evalúa el impacto que tiene las abejas

exploradoras dentro del algoritmo al variar el parámetro *limit*. Se concluye que un valor de reducido de iteraciones es necesario para conseguir resultados óptimos, esto debido a que la heurística es beneficiada de escapar de los mínimos locales con mayor frecuencia. Esto se puede deber a que la base de datos utilizadas presenta una mayor volatilidad y existen una mayor de falsos óptimos en su dominio.

Al sensibilizar el tamaño de la población, se encuentran resultados interesantes y discrepantes con la intuición original. Se obtiene que los óptimos son encontrados en los casos extremos evaluados. Este es el único punto de la investigación en el cual no se logra encontrar una tendencia clara sobre el comportamiento de la colonias. Por lo tanto, se pone a prueba el rendimiento de poblaciones con baja y alta densidad en futuras pruebas. Sin embargo, los tiempos computacionales entregan un motivo para preferir el primer tipo de colonia. Una conclusión mas fáctica de este punto se entregará al final de esta sección.

Se finaliza el análisis de la segunda fase al verificar si un aumento en las generaciones entrega mejores resultados. De aquí, se concluye que la aseveración pasada no es del todo cierta ya que se consiguen una mayor calidad de solución y menores tiempos de resolución cuando se utiliza una población relativamente baja. Este mismo comportamiento es validado al comparar con el texto de referencia.

Con los parámetros óptimos obtenidos de la segunda fase, se comienza la tercera fase al comparar los resultados con la CCEF estándar y GA. Para esto, se separa el algoritmo ABC

en dos grupos: ABC-LP y ABC-HP.

A partir de los resultados, se obtiene que en términos de calidad en la solución, ABC-LP y ABC-HP poseen rendimientos que son equivalentes estadísticamente. Además, ambos logran mejores resultados y menores errores que GA. También se valida que las heurísticas ABC poseen resultados similares al CCEF estándar. Por lo tanto, se concluye que el algoritmo basado en las colonias de abejas, gracias a su aproximación realista con el comportamiento de seres que han evolucionado que benefician a su comunidad, es una opción recomendable para resolver problemas relacionados a la optimización de portafolios.

Finalmente, se concluye que entre ambos métodos ABC, El algoritmo con menor población es recomendable sobre el de alta densidad ya que, presenta resultados óptimos a una fracción del tiempo de resolución.

## V. TRABAJOS FUTUROS

A partir de los resultados obtenidos, es claro que no se logran resultados comparables con el artículo de referencia debido a que se trabaja con una base de datos diferente y actualizada. Sin embargo, igual se esperaba obtener semejanzas en algunos comportamientos de la sensibilización.

Por lo anterior, se recomienda para un futuro trabajo, extender las instancias para la sensibilización de parámetros. Es decir, aumentar la cantidad de poblaciones y el parámetro *limit* para visualizar si existen cambios en sus tendencias. Por otro lado, también, es recomendable aumentar la cantidad de veces que se corre cada algoritmo ya que, se utiliza un número limitado debido a las grandes cargas computacionales que representan ejecutar cada algoritmo.

En el texto de referencia se compara la heurística ABC con al menos siete otros algoritmos. Se plantea la necesidad de probar estas siete técnicas en la nueva base de datos utilizada. Es posible que, por factores económicos recientes, algunas de estas heurísticas logren un mejor o peor comportamiento. De todas formas, probar estas metodologías servirían para validar aún más los resultados en la literatura.

Finalmente, este problema es resuelto para un CCVM-PO. Por lo tanto, una posible extensión sería resolver un problema con diferente función objetivo y restricciones. Si se sigue esta línea, el último punto es donde se debería tomar un mayor tiempo de análisis ya que, el algoritmo ABC puede sufrir al generar fuentes de comidas no factibles. Por lo tanto, se debería generar heurísticas adicionales para corregir estos nodos de alimentación sin perjudicar los tiempos computacionales.