Develando estrategias de mercado: minería de datos aplicada al análisis de mercados financieros

José Luis Gordillo-Ruiz[1], Enrique Martínez-Miranda[2] y Christopher R. Stephens[3]

[1] Dirección de Cómputo y Tecnologías de la Información y la Comunicación, UNAM, D.F., México

[2] C3 - Centro de Ciencias de la Complejidad, UNAM, D.F., México

[3] Instituto de Ciencias Nucleares, C3 - Centro de Ciencias de la Complejidad, UNAM, D.F., México

jlgr@super.unam.mx, enrique_mayhem@yahoo.com.mx, stephens@nucleares.unam.mx

Resumen. En los últimos años se han venido desarrollando marcos de estudio que intentan describir a los mercados financieros con mayor apego a la realidad que los marcos tradicionales, excesivamente simplificadores. En estos marcos se incluyen herramientas conceptuales y de análisis como evolución, sistemas complejos y minería de datos, entre otras. En particular, la minería de datos proporciona herramientas para extraer información a partir de la gran cantidad de datos que se generan del funcionamiento de los mercados financieros. En este trabajo, se presenta una metodología para inferir, a partir de los datos de un mercado, si participantes con resultados similares tienen estrategias similares y así intentar entender porque ciertos agentes son exitosos. Por decirlo de alguna manera, usamos estas herramientas para tratar de encontrar "huellas" de las estrategias de los agentes en las series de tiempo que son generadas a partir de su actividad en los mercados. Esta metodología puede verse a su vez como una conversión del problema a uno de clasificación, en donde se pretende corroborar que agentes con ganancias similares se encuentran en una misma región de un espacio discreto multidimensional, constituido por variables derivadas de los datos de las operaciones del mercado.

Palabras clave: Minería de datos, estrategias de mercado, análisis bayesiano, evolución, adaptación, predicción.

Inferring Market Strategies: Applying Data-Mining to Analysis of Financial Markets

Abstract. It has become increasingly common to model financial markets using frameworks which better capture their behavior than the excessively simplistic traditional frameworks. Key concepts in these new frameworks are evolution, complex systems and data mining, each with their associated characteristic analysis. In particular, data mining provides extremely useful tools for potentially extracting knowledge from the huge quantity of data available in financial markets. In this paper we present a new methodology for inferring, using market data, whether or not agents with similar performance are using similar trading strategies and by that to try to understand why certain agents are more successful than others. Put another way, we use data mining to look for "footprints", in the time series of price, that characterize the distinct trading strategies, and that are generated by their trading activity. One way to look at this is as a classification problem, where we try to classify agents with similar performance, determining if they are found in the same region of a discrete, multi-dimensional space composed of variables that are derived from the market data.

Keywords: Data mining, trading strategy, Bayesian analysis, evolution, adaptation, prediction.

1 Introducción

Hoy en día, debido a la cantidad y accesibilidad de datos electrónicos existente, la minería de datos se ha convertido en un campo de gran importancia y utilidad [12]. Una de las actividades humanas que genera datos en vastas cantidades son los mercados financieros; por ejemplo, en mercados como el NASDAQ en los Estados Unidos, se generan varios gigabytes al día, aun si sólo se consideran los datos más básicos del

estado dinámico del sistema. Estos datos pueden utilizarse para tratar de entender la dinámica de un mercado, desde la perspectiva de sistemas complejos adaptativos. En este artículo mostramos cómo la minería de datos puede servir para entender mejor el comportamiento de los mercados financieros.

El comportamiento de un mercado financiero está en gran medida determinado por las estrategias de sus participantes, o agentes, y sus posibilidades de aplicarlas, de modo que entender la forma en que estas estrategias interactúan es de gran relevancia para entender la dinámica de los mercados [8]. Hay un sin fin de posibles estrategias que un agente puede usar, desde lo más sofisticado, como utilizar herramientas de la Inteligencia Artificial para indicar cuándo comprar o vender, hasta lo más simple, como lanzar una moneda. Diferentes estrategias irán asociadas con diferentes grados de éxito. Desafortunadamente, a pesar de la vasta cantidad de datos que se puede obtener de la operación de un mercado, en la práctica es imposible determinar las estrategias que los participantes aplican para tomar sus decisiones (muchas veces éstas no son ni siquiera fácilmente descriptibles en forma algorítmica). A su vez, nadie en un mercado financiero dirá voluntariamente que estrategia usa, al menos no en forma detallada, porque otro inversionista podría aprovechar esa información para hacer ganancias. Así pues, es todo un problema determinar si las ganancias de un agente se deben a que tiene una estrategia superior en el mercado o simplemente son por suerte.

A pesar del hecho de que no existe un acceso directo a las estrategias de los participantes, si se cuenta con series de tiempo que describen sus operaciones. Así, en un mercado, cada participante genera una serie de tiempo de sus transacciones que, en cada tiempo dado, es única. Nuestra hipótesis es que en esta serie de tiempo está implícita una "huella" de las estrategias usadas por el inversionista, y que es posible clasificar a grupos de agentes con "huellas" similares. En cierto sentido, este es un problema similar al problema en la genética de inferir genotipos a partir de fenotipos.

Buscaremos clasificar a los agentes según sus ganancias, con el propósito de entender cómo y por qué ganan, ¿por suerte o por una estrategia inteligente? La metodología propuesta está inspirada de la minería de datos, y consiste en primero clasificar a los agentes en términos de un conjunto de características. Como características usamos diferentes variables observables en el mercado, como número de operaciones efectuadas, número de compras, etc., y después generamos grupos a partir de estas clasificaciones. Usando algunos diagnósticos estadísticos básicos identificamos las variables más correlacionadas con las ganancias. Finalmente, construimos un modelo predictivo usando el método de Bayes ingenuo para predecir cuales agentes tendrán el mayor éxito. Las clasificaciones pueden verse como dimensiones de un espacio discreto, mientras que las medidas son utilizadas para describir la posición de los agentes en este espacio. Posiciones similares representan "huellas" similares. El proceso completo se ilustra en la Fig. 1.

Para probar esta metodología, se divide el periodo de funcionamiento del mercado en dos partes. En el primer periodo se determinan grupos de agentes a partir de sus ganancias y se utiliza la metodología para describir estos grupos y la correlación entre sus series de tiempo y sus ganancias. Se construye un modelo predictivo usando este periodo como conjunto de entrenamiento. En el segundo periodo se vuelve a repetir el análisis, y se utiliza la metodología para tratar de explicar cómo ha evolucionado el mercado. Con el modelo predictivo usamos este periodo como conjunto de prueba para probar la calidad del modelo.

1.1 El mercado utilizado

Se ha escogido aplicar el análisis de mercados mediante metodologías de minería de datos a un mercado experimental que fue realizado en torno a las elecciones parlamentarias del estado alemán de Branderburgo [10]. Este experimento fue llevado a cabo entre los meses de agosto y septiembre de 2004, y en él participaron 108 personas con la libertad de operar en 8 emisoras (los partidos contendientes en la elección). No hubo dinero real invertido en este mercado experimental.

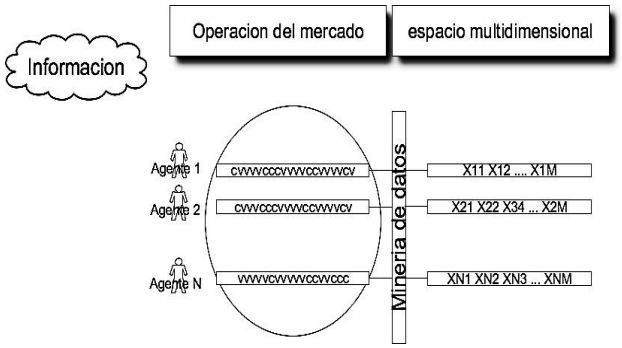


Fig. 1. Representación esquemática de la metodología utilizada**.** CVV representa la secuencia de compras (C) y ventas (V) de un agente, mientras que Xij representan las características como, por ejemplo, el volumen promedio por transacción

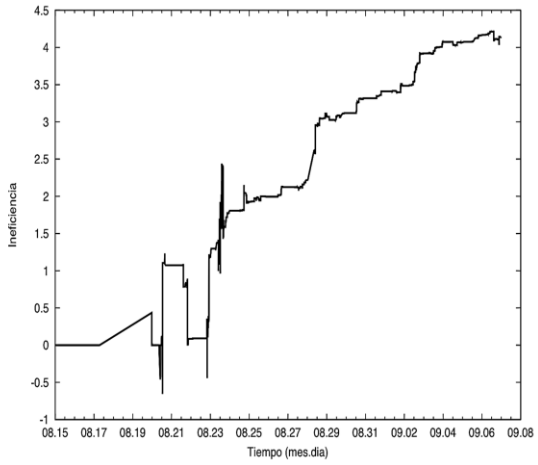


Fig. 2. Ineficiencia del Grupo1 vs. el resto. Primer periodo

Una de las teorías más conocidas de mercados es la Teoría del Mercado Eficiente E.F. Fama [4], que esencialmente indica que la información se difunde homogénea e instantáneamente entre los participantes. De esta teoría se desprende que en un mercado eficiente es imposible tener ganancias excesivas consistentemente, puesto que la información que permite obtener ganancias se difunde de forma automática e instantánea a todo el mercado.

En trabajos anteriores se ha desarrollado una métrica conocida como *ineficienci*a [11], que es una medida de confiabilidad estadística de que existe un grupo de participantes que tiene ganancias sistemáticas en un mercado. La existencia de ineficiencia en un mercado invita a analizarlo y tratar de determinar cuáles son las estrategias que están utilizando los agentes ganadores.

En la Fig. 2 se muestra la gráfica de ineficiencia obtenida del primer periodo del mercado estudiado[1]. Para determinar esta ineficiencia, los participantes del mercado fueron organizados en 10 grupos, definidos a partir de sus ganancias hasta la finalización del primer periodo. Es evidente que dado este sesgo en la selección de grupos, la ineficiencia en este periodo debe existir. Sin embargo, lo que es de notarse es que es sistemática, por lo que intentaremos descubrir si los agentes ganadores tienen estrategias similares, así como si son estrategias ganadoras en el otro periodo del mercado.

Para esto, primero se realizará el análisis del comportamiento de los agentes en el primer periodo. De este análisis se desprende si los agentes con ganancias similares se agrupan en la misma región del espacio de parámetros, es decir, si existen características similares entre los agentes con ganancias similares. Después, el análisis se repite para el segundo periodo, y se trata de verificar si las estrategias se mantienen y generan los mismos resultados, o, en todo caso, se buscan procesos de adaptación y aprendizaje (cambio de fenotipo), y se observa si los fenotipos tienen resultados similares al periodo anterior, o si nuevos fenotipos han surgido que resulten también en comportamientos ganadores.

En la siguiente sección se describe con mayor detalle la metodología usada para analizar el mercado. En la Sección 3 se presentan los resultados de ambos periodos, y en la Sección 4 se presentan las conclusiones y trabajo futuro.

[1] El primer periodo comprende el intervalo del 17/08 al 06/09 y el segundo periodo el intervalo del 07/09 hasta el 20/09.

2 Definiciones de medidas, clasificaciones y score

Como ya se ha mencionado, de la operación de un mercado se genera una serie de tiempo para cada uno de los participantes. Esta serie de tiempo es la secuencia de las operaciones que cada agente realiza, tanto en posición (compra o venta), como en volumen y precio. Las series de tiempo son los datos sobre los que se realizará la minería. De alguna forma, son las características "fenotípicas" que observamos en los agentes.

2.1 Clasificación

De las series de tiempo de los agentes se pueden derivar datos como son: cuantas operaciones hicieron, cuantas compras, cuantas ventas, cual fue su ganancia, etc. Los agentes se caracterizan por estos criterios, de modo que cada agente puede caracterizarse por su posición, en un orden descendente, en cada una de estas variables. Para reducir aún más el tamaño del espacio de parámetros, es decir reducir los grados de libertad, dividimos las variables en deciles, de modo que cada agente se caracterice por el decil de su posición en cada clasificación.
La lista de características utilizadas es:

- X_1: Número total de transacciones realizadas.
- X_2: Número total de compras.
- X_3: Número de ventas.
- X_4: D, definida como el porcentaje de cambios de signo (compras seguidas de una venta o viceversa) con respecto al total de transacciones.
- X_5: Número de contrapartes, es decir, el número de agentes diferentes contra los que el agente hizo operaciones.
- X_6: Volumen total de acciones operadas durante el periodo.
- X_7: Volatibilidad del volumen operado.
- X_8: Volumen promedio de acciones por operación.
- Clase C: Ganancias medidas mediante *Sharpe ratio*[2]

A partir de las clasificaciones anteriores transformamos el problema a uno discreto: se crea un espacio de 10 dimensiones, en donde cada dimensión tiene 108 puntos (el valor de la posición de cada agente en cada dimensión o variable). Clasificaremos este espacio a partir de las medidas de correlación y *score* descritas en la siguiente sección.

2.2 Medidas de correlación y *score*

La pregunta es: ¿qué tienen en común, si hay algo, los agentes con ganancias similares? Primero, hay que buscar cuales características (X_i) tienen una fuerte correlación con las ganancias de los agentes. Esto puede determinarse mediante la medida $\varepsilon(C|X_i)$, definida como que da el significado estadístico de que nuestras observaciones desvían de la hipótesis nula P(C).

$$\epsilon(C|X_i) = \frac{N_{x_i}(P(C|X_i) - P(C))}{(N_{x_c}P(C)(1-P(C)))^{1/2}} \qquad (1)$$

La distribución relevante es un binomial que, si la muestra no es muy pequeña, puede ser aproximada por una distribución normal. En este caso $|\varepsilon(C|X_i)| > 2$ significa que la hipótesis nula está violada hasta un intervalo de confianza de 95%. Para determinar $\varepsilon(C|X_i)$, se ordena a los agentes según su *Sharpe ratio* en el primer periodo y se dividen en deciles. La clase C es el grupo de agentes de un cierto decil de *Sharpe ratio*, mientras que X_i representa un decil de una de las variables fenotípicas que se usa para caracterizar a los agentes. $\varepsilon(C|X_i)$ indica la confiabilidad estadística de que un miembro de X_i pertenezca a la clase C [3].

Otra pregunta interesante es: ¿qué tan diferentes son los agentes con un *Sharpe ratio*

[2] El Sharpe ratio [Sharpe, 1994] es una medida de las ganancias de un portafolio con respeto a un punto de referencia, como la taza de interés sin riesgo, y relativo al grado de riesgo del portafolio como medido por la desviación estándar de las ganancias. En el presente contexto se usa como medida de ganancia, las ganancias excesivas sistemáticas y el riesgo está medido por la desviación estándar de estas ganancias.

[3] Ver [Stephens, et al., 2006]

diferente? Esto puede determinarse mediante otra medida, nombrada como ε', y definida como

$$\epsilon' = \frac{(\langle x_i \rangle_c - \langle x_i \rangle_{c'})}{(\frac{\sigma^2_{i_c}}{n_c} + \frac{\sigma^2_{i_{c'}}}{n_{c'}})^{\frac{1}{2}}} \quad (2)$$

donde <x_i> es el valor promedio de la característica X_i para los miembros de la clase *C*. Esta medida nos indica qué tan diferentes son los miembros de dos clases (*C, C'*) en términos de sus correspondientes valores de una característica X_i. ε' puede verse como una medida de "distancia" en cada una de las "coordenadas" del espacio multidimensional.
Finalmente, queremos saber si las características X_i confirman la hipótesis de que un agente es ganador o no. Para esto usamos una medida denominada *score*, determinada mediante el análisis *bayesiano*. Esta medida es la que nos permite clasificar las regiones del espacio multidimensional en donde se localizan los agentes, en términos de qué tan probable es que un agente en esa parte del espacio tenga ganancias típicas de la clase *C*.

$$P(C|\mathbf{X}) = \frac{P(\mathbf{X}|C)P(C)}{P(\mathbf{X})} \quad (3)$$

La regla de Bayes está dada por que relaciona la probabilidad a posteriori $P(C|\mathbf{X})$ con la probabilidad a priori $P(C)$ y la función de probabilidad condicional $P(\mathbf{X}|C)$. Dado que $P(\mathbf{X})$ es la probabilidad de todo un conjunto de características, la regla de Bayes se transforma en

$$P(C|\mathbf{X}) = \frac{P(X_1, X_2, .., X_n|C)P(C)}{P(X_1, X_2, .., X_n)} \quad (4)$$

el problema con la ecuación anterior es que la probabilidad conjunta de todas las características X_i es necesariamente 0 ó 1, lo que imposibilitaría usar la regla de Bayes para distinguir entre comportamientos fenotípicos. Una forma de resolver esto es considerar la regla de Bayes para el complemento de la clase *C*, es decir,

$$P(\bar{C}|\mathbf{X}) = \frac{P(X_1, X_2, .., X_n|\bar{C})P(\bar{C})}{P(X_1, X_2, .., X_n)} \quad (5)$$

y relacionar la regla de Bayes para la clase *C* y su complemento, de la siguiente forma:

$$\frac{P(C|\mathbf{X})}{P(\bar{C}|\mathbf{X})} = \frac{\frac{P(\mathbf{X}|C)P(C)}{P(\mathbf{X})}}{\frac{P(\mathbf{X}|\bar{C})P(\bar{C})}{P(\mathbf{X})}} \quad (6)$$

si los eventos X_i son independientes, tenemos

$$P(\mathbf{X}|C) = \prod P(X_i|C) \quad (7)$$

definiendo

$$S'(X) = \ln \frac{P(C \mid X)}{P(\bar{C} \mid X)} \quad (8)$$

tenemos

$$S'(\mathbf{X}) = ln \frac{\prod P(X_i|C)}{\prod P(X_i|\bar{C})} + ln \frac{P(C)}{P(\bar{C})} \quad (9)$$

dado que $\ln \frac{P(C)}{P(\bar{C})}$ no indica nada sobre la correlación de la clase *C* con las características X_i. despreciamos este valor. Si además

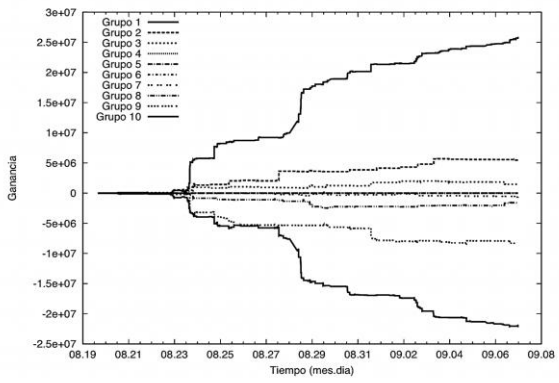


Fig. 3. Ganancias excesivas sistemáticas de cada grupo de agentes durante el primer periodo

aplicamos igualdad de logaritmos, tenemos que el *score* $S(\boldsymbol{X})$ es

$$S(\mathbf{X}) = \sum ln \frac{P(X_i|C)}{P(X_i|\bar{C})} \quad (10)$$

Un *score* alto/bajo indica mayor/menor probabilidad de pertenecer a la clase *C*.

3.Resultados

3.1 Tablas de medidas, rangos y *scores* en el primer periodo

Los agentes fueron agrupados en 10 conjuntos, de acuerdo a su clasificación en *Sharpe ratio* (clase *C*). Las ganancias de cada grupo durante el primer periodo del mercado se muestran en la Fig. 3. Los grupos 4, 5 y 6 no tuvieron actividad, mientras que las ganancias de los grupos 2 y 3 son comparables a las pérdidas de los grupos 7, 8 y 9. Notablemente, el grupo 1 tiene ganancias extraordinarias y comparables a las pérdidas del grupo 10. Por el momento nos concentraremos en determinar si estos resultados de ganancias y pérdidas corresponden a estrategias similares y bien definidas.

Los valores de $\varepsilon(C|X_i)$ para las variables más importantes pueden verse en la Tabla 1. Podemos ver que existe una correlación muy alta para los grupos 1 y 10 con el número de operaciones (X_1), el número de compras (X_2) y el número de ventas (X_3), así como con el volumen total (X_6), D (X_4), y el número de contrapartes (X_5). Es decir, esta caracterización nos está indicando que estrategias similares (en términos de las variables señaladas), pueden derivar tanto en muy altas ganancias como en muy altas pérdidas.

Por otro lado, en la Tabla 2 podemos ver los valores de ε', comparando el primer grupo contra cada uno de los demás grupos de la clase *C*. Podemos ver que esta medida es capaz de diferenciar a los agentes del grupo 1 de los agentes de los grupos 2, 3, 7, 8 y 9, pero no así de los agentes del grupo 10.

Tabla 1. Valores de ε en el primer periodo, para los diferentes grupos de la clase *C*

Grp	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	3.87	4.86	3.87	0.88	3.87	3.87	-0.1	-0.1
2	-1.1	-1.1	-1.1	-0.1	-1.1	-1.1	1.87	0.88
3	-0.1	-0.1	0.88	-0.1	0.88	-0.1	-1.1	0.88
7	-1.1	-1.1	-1.1	2.87	-1.1	-1.1	1.88	1.87
8	-1.1	-1.1	-1.1	-1.1	-0.1	-1.1	-1.1	-1.1
9	-0.1	-0.1	-1.1	2.88	-1.1	-1.1	1.88	1.87
10	2.45	1.53	2.45	1.53	3.37	2.45	-0.3	-1.2

Tabla 2. Valores de ε' en el primer periodo, para los diferentes grupos de la clase C, con respecto al primer grupo

1 vs.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
2	3.6	3.7	3.2	1.7	3.7	4.3	-0.2	-0.5
3	3.7	2.6	2.6	3.4	3.1	3.5	1.6	-0.4
7	3.8	3.9	3.5	0.6	4.6	4.0	0.2	-0.5
8	3.2	3.4	2.8	2.1	3.1	3.0	1.5	0.8
9	2.6	2.4	2.6	0.8	2.8	3.0	-0.2	-0.5
10	-0.4	0.5	-0.8	-0.4	-0.4	0.3	1.0	0.5

En la Tabla 3 podemos ver la clasificación en *score* de los agentes de los grupos 1 y 10. Podemos ver que la gran mayoría de agentes en los primeros dos deciles del *score* están repartidos entre estos dos grupos. De hecho, podemos utilizar ε' para cuantificar las diferencias en la clasificación de *score* de la clase *C*, lo cual se muestra en la Tabla 4. Podemos observar que el grupo 1 se diferencia notablemente de los demás grupos, con excepción del 10.

3.2 Adaptabilidad/predictibilidad: tablas de medidas, rangos y *scores* en el segundo periodo.

En la Fig. 4 podemos ver la ineficiencia del mercado en el segundo periodo, considerando la misma partición del mercado que en el primer periodo. Se puede notar que, aunque es menor, la ineficiencia no ha desaparecido. Esto reafirma

Tabla 3. Rangos de *score* en el primer periodo, para los agentes de los grupos 1 y 10

Agentes del grupo 1	**Rangos en *score***
mammutfan	1
Hanni1982	5
famfan	6
saladin	4
gruener	3
zigzag	7
Wahlaal	22
rapper	20
kaufunger	16
Truck676	30
cezanne	23
Agentes del grupo10	**Rangos en *score***
Traurig	26
Geoman	41
Tishimdorf	21
Progno	19
Briutt	18
Tob11	43
Camporesi	14
fischmob	8
Rodrigues	17
Mauritius	11
BAYERNP	2
Angelo	12

Tabla 4. Valores de ε' entre *scores* de los diferentes grupos en el primer periodo.

Grupo 1 vs.	**2**	**3**	**4**	**5**	**6**
	-5.2	-4.3	-19.4	-19.4	-19.4
Grupo 1 vs.	**7**	**8**	**9**	**10**	
	--6.7	-5.0	-3.9	-1.4	

que las ganancias de los agentes más ganadores no se deben a suerte, sino que tienen estrategias ganadoras, aunque aún queda preguntarse por qué es más baja.

En la Fig. 5 vemos las ganancias de cada grupo de la clase *C*. A diferencia del primer periodo, ahora todos los agentes tienen participación en el mercado. Aún las mayores ganancias pertenecen al grupo 1, aunque su magnitud es menor que en el primer periodo. Las pérdidas del grupo 10 también son notoriamente menores, aunque siguen siendo sistemáticas. El grupo 4 obtiene las mayores pérdidas, pero en gran parte debido a una sola operación. Veamos que ha sucedido con las estrategias.

En las Tablas 5 y 6 puede verse los valores de ε y ε' para el segundo periodo. En cuanto a ε, podemos ver que la correlación ya no es tan grande como en el primer periodo, mientras que ε' nos muestra, por un lado, que ahora es un poco más difícil distinguir al grupo 1 de los grupos 2, 3, 4, 5, 6, 7, 9; y por el otro, es más fácil distinguir al grupo 1 del grupo 10.

Si observamos la Tabla 7, vemos que el número de agentes de los grupos 1 y 10 que están en los primeros dos deciles del *score* se ha reducido. Si utilizamos nuevamente la medida ε' sobre el *score* de los diferentes grupos (Tabla 8), podemos ver que ahora es mucho más difícil distinguirlos de lo que era en el primer periodo (Tabla 4). Lo anterior implica que los agentes han cambiado su posición en el espacio multidimensional de las características X_i. A manera de ejemplo, en la Tabla 9 mostramos los valores de 6 coordenadas, para dos agentes de cada grupo (1 y 10). En cada caso, se muestra un agente que se ha mantenido en la misma clase y otro que se ha mudado a una clase diferente.

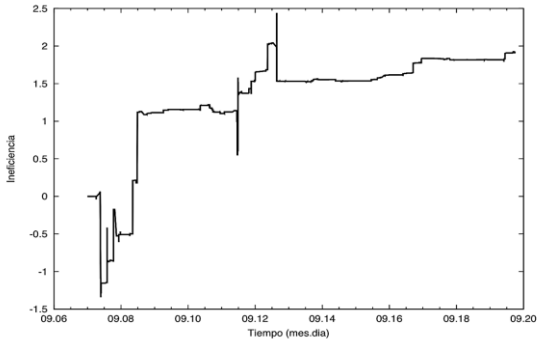


Fig. 4. Ineficiencia grupo 1 vs. el resto. Segundo periodo. (Ver texto arriba para más descripción)

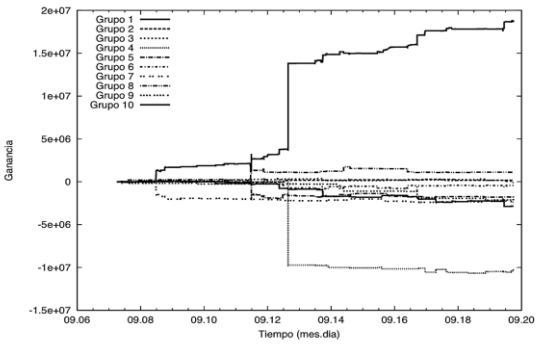


Fig. 5. Ganancias/pérdidas de los diferentes grupos de agentes en el segundo periodo. (Ver texto arriba para más descripción)

Una pregunta interesante es, ¿las regiones en este espacio en donde fueron clasificados los agentes más ganadores en el primer periodo, siguen siendo las mismas en el segundo? En la Tabla 10 se muestra el rango en la clase *C* para el segundo periodo, junto con el rango en *score*. Puede observarse que, de los 11 agentes más ganadores en el segundo periodo, 10 de ellos pertenecen a los dos primeros deciles en *score*, aunque 7 de ellos no pertenecieron al grupo de más ganadores en el primer periodo. Similarmente, 8 de los 11 agentes más perdedores en el segundo periodo pertenecen a los dos primeros deciles en *score*, aunque 9 de ellos no pertenecieron al grupo de los más perdedores en el primer periodo. Estos resultados

Tabla 5. Valores de ε en el segundo periodo, para los diferentes grupos de la clase C.

Grupo	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	1.87	1.87	0.88	-1.1	2.87	1.87	0.88	0.88
2	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1
3	-1.1	-1.1	0.88	-1.1	-0.1	-0.1	-1.1	-1.1
4	0.08	0.08	0.08	1.17	0.08	0.08	1.17	1.17
5	0.08	0.08	0.08	-1.0	0.08	0.08	1.17	1.17
6	0.08	-1.0	1.17	1.17	0.08	1.17	-1.0	-1.0
7	-0.1	0.88	-1.1	-1.1	-0.1	-0.1	-1.1	-1.1
8	-0.1	-0.1	0.88	0.88	-0.1	-0.1	-1.1	-1.1
9	-1.1	-1.1	-1.1	1.88	-1.1	-1.1	2.87	1.87
10	0.61	0.61	-0.3	0.61	-0.3	-0.3	-0.3	0.61

Tabla 6. Valores de ε' en el segundo periodo para los diferentes grupos de la clase C, comparados con el grupo 1 de la misma clase.

1 vs.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
2	2.36	2.30	1.81	2.91	3.12	2.84	3.04	2.13
3	1.17	1.09	1.02	2.87	1.61	1.49	2.40	1.60
4	1.73	1.09	1.02	2.87	1.61	1.49	2.40	1.60
5	1.23	1.70	0.44	0.98	1.76	1.38	-0.3	-0.04
6	1.33	1.67	0.45	-0.2	1.04	1.44	1.53	0.99
7	1.49	1.33	1.39	1.83	1.86	1.83	2.22	1.71
8	1.43	1.72	0.54	0.85	1.57	1.72	2.36	1.61
9	2.21	2.17	1.66	-1.0	2.55	2.29	0.30	0.11
10	1.10	0.99	0.94	-0.3	1.08	1.53	.135	0.69

Tabla 7. Rango de *score* en la segunda mitad, para los agentes de los grupos 1 y 10

Agentes del grupo 1	Rangos en score
mammutfan	4
Hanni1982	16
Famfan	34
Saladin	1
Gruener	12
Zigzag	10
Wahlaal	75
Rapper	88
Kaufunger	40
Truck676	18
Cezanne	56
Agentes del grupo 10	**Rangos en score**
Traurig	37
Geoman	107
Tishimdorf	19
Progno	38
Briutt	45
Tob11	56
Camporesi	65
fischmob	9
Rodrigues	24
Mauritius	15
BAYERNP	11

Tabla 8. ε' en *score* del segundo periodo mostrando que los agentes del grupo 1 son más similares a los demás en el segundo periodo comparado con el primero

Grupo 1 vs.	**2**	**3**	**4**	**5**	**6**
	-4.4	-1.9	-2.1	-1.6	-0.4
Grupo 1 vs.	**7**	**8**	**9**	**10**	
	-1.2	-1.3	-2.9	-0.9	

evidencian el desempeño del modelo mostrando que si es posible predecir cuales estrategias serán exitosas. Debe notarse que esta predicción no es en tiempo, más bien, se está observando

Tabla 9. Valores de las coordenadas en el espacio multidimensional, para dos agentes del grupo 1 y dos del grupo 10, en ambos periodos. Se puede observar que algunos agentes mantienen sus estrategias mientras otros la han cambiado

Agente	G	P	X_1	X_2	X_3	X_5	X_6	C
mammutfam	1	1	2	2	2	5	7	1
		2	4	8	2	9	2	1
cezanne	1	1	20	18	27	34	25	11
		2	86	60	85	81	82	37
Angelo	10	1	1	16	1	7	1	108
		2	67	80	49	74	90	69
BAYERNP	10	1	4	3	5	7	2	107
		2	11	9	22	16	24	102

que, con los datos asociados a las variables X_i, al final del periodo podemos predecir, sin ver las ganancias finales, que estrategias deberían ser las más exitosas.

4 Conclusiones

En el mercado escogido, la metodología propuesta fue útil para identificar y diferenciar las "huellas" de estrategias de los agentes más ganadores con respecto de otros grupos de agentes, salvo de los agentes más perdedores. De alguna forma, las estrategias que generan mayores ganancias y mayores pérdidas son similares en la primera mitad.

Las medidas y *scores* utilizados mostraron un cambio en las estrategias de los participantes del mercado. Pudo observarse que algunos agentes de los grupos 1 y 10 cambiaron su comportamiento, lo cual condujo a una reducción de las ganancias/pérdidas de estos grupos y, como consecuencia, a una reducción en la medida de ineficiencia del mercado.

La medida de *score* también fue útil para identificar rápidamente a los agentes de otros grupos que en el segundo periodo cambiaron su comportamiento, y que empezaron a tener más ganancias o pérdidas. Estos agentes adquirieron

Tabla 10. Valores de rangos en *Sharpe ratio* y *score* en la segunda mitad, así como grupo original, para los 11 agentes con mayor y menor *Sharpe ratio* en el segundo periodo. Esta tabla muestra la eficacia del modelo predictivo (ver texto arriba para más detalle).

Agente	Rango en *Sharpe ratio*	Rango en *score*	Grupo original
mammutfan	1	4	1
Vokert	2	6	6
Armin139	3	2	4
Zigzag	4	10	1
saladin	5	1	1
fischmob	6	9	10
mc0050	7	2	3
eisbaeli	8	14	8
andrewo	9	7	8
truck676	10	18	1
mahlow	11	26	9
Makler1	98	5	5
NZ1968	99	84	9
socke	100	17	6
Wiego	101	50	9
BAYERNP	102	11	10
Vollsepp	103	13	6
REBELL	104	8	7
Bruder	105	20	7
Napoleon	106	21	7
Schwuchtel	107	29	8
tishimdorf	108	19	10

comportamientos similares a los de los grupos 1 y 10 en el primer periodo.

Finalmente, cabe mencionar que hemos construido un modelo predictivo relacionando ciertos parámetros "fenotípicos" con las ganancias. Por supuesto, ésto no demuestra que las ganancias o pérdidas son un resultado causal de una estrategia descrita por los valores correspondientes de estos parámetros. Más bien, muestra que hay una correlación entre las ganancias y los valores de los parámetros que si o no puede representar un efecto causal. Si pensamos en algunas de las variables, es lógico que el número de transacciones está correlacionado con las ganancias/pérdidas – entre más se apuesta, más se puede perder o ganar.

Agradecimientos

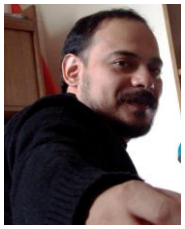
Los autores agradecen el financiamiento de CONACYT a través de los proyectos "Centro de Ciencias de la Complejidad" y "Red Temática de Complejidad, Ciencia y Sociedad", así como al proyecto UNAM-PAPIIT IN120509**.**

Referencias

1. **Aragonés, J.R. & Mascareñas, J. (1994).** La Eficiencia y el Equilibrio en los Mercados de Capital. *Análisis Financiero*, 64, 76–89.
2. **Benink, H.A., Gordillo, J.L., Pardo, J.P., & Stephens, C.R. (2010).** Market efficiency and learning in an artificial stock market: A perspective from Neo-Austrian economics. *Journal of Empirical Finance,* 17(4), 668–688.
3. **Cont, R. (1999).** Modeling Economic Randomness: Statistical Mechanics of Market Phenomena. In M.T Batchelor & L.T. Wille (Eds.), *Statistical Physics in the eve of the 21st Century*. Singapore: World-Scientific. Retrieved from http://www.cmap.polytechnique.fr/~rama/papers/festschrift.ps.
4. **Fama, E.F. (1970).** Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417.
5. **Kovalerchuk, B. & Vityaev, E. (2000).** *Data Mining in Finance: Advances in Relational and Hybrid Methods*. Boston: Kluwer Academic Publishers.
6. **Mansilla, R. (2003).** *Introducción a la Econofísica*. Madrid: Equipo Sirus.
7. **Mantegna, R.N. & Stanley, H.E. (2000).** *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge, UK: Cambridge University Press.
8. **O'Hara, M. (1998).** *Market Microstructure Theory*. Cambridge, Mass: Blackwell Publishers.
9. **Sharpe, W.F. (1994).** The Sharpe ratio. *The Journal of Portfolio Management,* 21(1), 49–58.
10. **Stephens, C.R., Gordillo, J.L., & Hauser, F. (2006).** Testing inefficiency in an Experimental Market using Excess Trading Returns. 1st International Conference on Economic Sciences with

Heterogeneous Interacting Agents (WEHIA 2006). Bolognia, Italy. Retrieved from http://www2.dse.unibo.it/wehia/paper/parallel%20session_5/session_5.4/Stephens_5.4.pdf

11. **Stephens, C.R., Benink H.A., Gordillo J.L., & Pardo-Guerra, J.P. (2007).** A New Measure of Market Inefficiency. Retrieved from http://ssrn.com/abstract=1009669
12. **Stephens, C.R., & Sukumar, R. (2007).** An introduction to Data Mining. The Handbook of Market Research. 455-486. SAGE Publications.



José Luis Gordillo Ruiz Ingeniero en Computación por la Facultad de Ingeniería de la UNAM, y Maestro en Ciencias por el Posgrado en Computación de la UNAM. Actualmente colabora con la Dirección de Cómputo y Tecnologías de la Información y la Comunicación de la UNAM. Sus áreas de interés son cómputo de alto rendimiento, sistemas complejos y econofísica.

Enrique Martínez Miranda Licenciado en Física por la Facultad de Ciencias de la UNAM, y Maestro en Ingeniería por la Facultad de Ingeniería de la UNAM. Ha colaborado en varias empresas de análisis financiero. Actualmente colabora con el Centro de Ciencias de la Complejidad de la UNAM. Sus áreas de interés son mieria de datos, finanzas computacionales, econofísica y sistemas complejos.

Christopher R. Stephens Licenciado en Física por The Queen's College, Oxford y maestria y doctorado por la Universidad de Maryland, E.U. Actualmente es Invetigador Titular C del Instituto de Ciencias Nucleares de la UNAM. Es miembro del Comité Académico del C3 – Centro de Ciencias de la Complejidad de la UNAM. Es investigador de nivel III del SNI. Cuenta con más de 100 publicaciones. Sus intereses son sistemas complejos, finanzas y mineria de datos.

Artículo recibido el 08/12/2010; aceptado el 16/01/2012.