

DATA MINING Y ANÁLISIS MATEMÁTICO DE LAS CUOTAS DE LAS CASAS DE APUESTAS DEPORTIVAS ONLINE

GONZALO PÉREZ – SEOANE TORRES – CABRERA

gonzaloperezseoane94@gmail.com

Universidad San Pablo CEU / NFQ Advisory Services

C/ Julián Romea, 23, 28003 Madrid

CARLOS QUESADA GONZÁLEZ

carlos.quesadagonzalez@ceu.es

Facultad de Económicas y Empresariales. Universidad San Pablo CEU.

C/ Julián Romea, 23, 28003 Madrid

Recibido (22/09/2018)

Revisado (23/11/2018)

Aceptado (20/12/2018)

RESUMEN: Este artículo aborda el estudio empírico del comportamiento de las cotizaciones de las apuestas 1X2 en las casas de apuestas online en España. En el estudio se presenta un marco teórico en el que el apostante puede obtener una rentabilidad y se estudian posibles estrategias, a través de la modelización matemática, mediante las cuales dicha disquisición teórica se podría llevar a cabo en la práctica. Para la realización del estudio empírico se ha obtenido una base de datos correspondiente a 115 días de cotizaciones sobre apuestas de 175 partidos de fútbol de la Liga Santander cotizados en la plataforma de apuestas online Sportium mediante técnicas de programación y automatización conocidas como *web-scraping*. A lo largo del estudio se analizaron cuestiones tales como el juego justo, los márgenes obtenidos por las casas de apuestas y las variables que intervienen en el proceso de asignación de precios a las cotizaciones.

Palabras claves: Apuestas online, webscraping, análisis de datos.

ABSTRACT: This paper addresses an empirical study on the behavior of the quotes for 1X2 bets in online betting houses in Spain. The study shows a theoretical approach where the bettor can obtain a systematic profitability and studies possible strategies in order to achieve that theoretical result in an actual, real-life framework. Programming and automation techniques, usually known as web-scraping, were used to obtain a dataset consisting on 115 days of bets on 175 football matches in Spanish National League from the Sportium online betting platform. Throughout the study, issues such as fair gaming, the margins obtained by bookmakers and the variables that intervene in the process of assigning prices to the quotes were analyzed.

Keywords: Online gambling, webscraping, data analysis.

1. Introducción

El mundo de las apuestas es un sector ampliamente estudiado, especialmente en los últimos años, con el interés suscitado a causa de la liberalización del mercado¹ y la aparición y proliferación de apuestas online, ver (Wood & Williams, 2007) y (Kealey, 2005). En (Boyle, 2006) se estudia cómo las casas de apuestas ganan dinero y generan situaciones de beneficio centrándose en el funcionamiento de las transacciones y en cómo, desde el punto de vista del apostante, se puede minimizar ese margen favorable a las casas. Diferentes técnicas para apostantes son estudiadas en (Vaughan Williams, 2004) aunque con especial interés en ciertos tipos de apuestas especialmente frecuentes en Inglaterra. Las cotizaciones y estrategias en el ámbito futbolístico han sido objeto de estudio en numerosos trabajos previos como (Christie, 2006) si bien estos estudios tienen un enfoque tradicional. En mundo de las apuestas también ha sido estudiado desde el punto de vista de la economía (Martínez Élices, 2016) donde se usan modelos econométricos para analizar la eficiencia de las apuestas. Otros aspectos de las apuestas, como la regulación de las mismas se han estudiado en (Pérez, 2010).

El enfoque de este artículo es diferente al de la literatura previa, es conocido que las cotizaciones para un mismo evento en dos casas de apuestas diferentes pueden diferir lo suficiente para que haya un arbitraje que el apostante pueda aprovechar obteniendo un beneficio sistemático. Hace años era posible crear *scripts*, pequeños programas automáticos, que detectaban estas situaciones y realizaban apuestas de forma automática. Aunque la frecuencia de tales circunstancias es poco habitual, cuando ocurre es dinero seguro. Apostando cantidades elevadas, pero no tanto como para hacer sospechar a las casas, se conseguían obtener importantes sumas de dinero de forma regular. El grado de eficiencia del mercado de apuestas deportivas es cada vez más elevado (Fama, 1965) y además, en la actualidad, las casas de apuestas rastrean permanentemente este tipo de actividades y comparten entre ellas la información sobre apuestas, con lo que las ventanas temporales en las que el desfase ocurre se reducen a segundos y detectan a cualquier usuario con comportamientos sospechosos (Dirección General de Ordenación del Juego, 2017).

En este artículo, los autores proponen otra forma de llegar a este arbitraje favorable para el apostante. En vez de considerar las diferencias de cotizaciones entre dos casas de apuestas, se consideran las diferencias de cotizaciones en una misma casa de apuestas pero en distintos momentos temporales. Se plantea la hipótesis de si un agente individual, o apostante, podría llegar a obtener ganancias de forma sistemática colocando apuestas a los tres posibles desenlaces de un partido de la muestra seleccionada en aquellos momentos temporales del periodo de cotización de la apuesta en los que las cuotas alcancen su valor máximo. El riesgo implícito de esta estrategia radica en que en el momento de realizar la primera apuesta, existe la posibilidad de que nunca se llegue a una situación en la que haya ganancia segura. Es por ello que es necesario estudiar el comportamiento de las apuestas, para saber si se puede esperar que exista esta situación en circunstancias reales y si se puede diseñar una estrategia ganadora. (Levitt, 2004)

Gracias a las nuevas técnicas de minado de datos, es posible obtener la información de páginas web de forma automática y periódica, lo que se conoce como web-scraping. Estas nuevas herramientas posibilitan la validación del enfoque teórico, mediante la creación una base de datos con información sobre la evolución de las cotizaciones (Aluja, 2001).

En este artículo se estudian apuestas de tipo 1X2 de fútbol, esto es, aquellas en las que los escenarios posibles contemplan que gane el equipo local, el equipo visitante o que se produzca un empate. En concreto, se han recogido minuto a minuto las cotizaciones del operador Sportium para los partidos que tuvieron lugar en la Primera División de Fútbol Profesional de España en el periodo de tiempo comprendido entre el 9 de diciembre de 2017 al 4 de abril de 2018. El estudio se centra en el análisis descriptivo-estadístico de las cuotas en búsqueda de patrones temporales que puedan arrojar luz para desarrollar una estrategia eficaz de trading en el marco de las apuestas deportivas.

El análisis realizado muestra que existe la posibilidad real de encontrar las situaciones en las que se puede apostar sin riesgo. Los autores presentan además una estrategia de selección de partidos que maximiza la posibilidad de encontrar esas situaciones favorables cuya esencia está en evitar apuestas en partidos con un favorito claro. Como línea de investigación futura se debe profundizar en la creación de la

¹ Ley 13/2011, de 27 de mayo, de regulación del juego

estrategia, estimando y creando un modelo efectivo que permita algoritmizar las apuestas. Además, la ampliación de la base de datos será fundamental para validar las posibles estrategias que se diseñen.

El artículo se estructura de la siguiente manera; en la sección 2 se presenta la hipótesis de forma teórica, en la sección 3 se analiza si efectivamente se dan dichas situaciones teóricas en la realidad del mercado de apuestas y cómo encontrarlas, las secciones 4 y 5 se centran en el estudio de la estrategia a seguir y finalmente la sección 6 presenta las conclusiones.

2. Arbitraje entre casas de apuestas.

Las cuotas de salida para cada tipo de apuesta son fijadas por las casas de apuestas, las cuales, partiendo de una situación informacionalmente eficiente en su forma débil (Fama, 1970), utilizan los datos históricos del transcurso y desenlace de cada evento deportivo. Esta fijación de precios de las apuestas, o determinación de cotización inicial, se realiza teniendo en cuenta toda la información disponible en dicho momento inicial. Dichas cotizaciones necesariamente han de reflejar fielmente las probabilidades de ocurrencia de los distintos desenlaces posibles. En este momento inicial M_0 , ningún agente de mercado, o apostante, podrá realizar una asignación de precios más eficiente de la realizada sin información privilegiada, ya que estas cotizaciones se han formado a partir de toda la información pasada disponible en el momento presente (Lucas, 1972)

En el periodo de tiempo comprendido entre el inicio de la cotización de las apuestas y el comienzo del evento deportivo, M_0 y M_F respectivamente, las cotizaciones se irán actualizando según dos parámetros: expectativas de los apostantes y noticias u otras informaciones subyacentes. Más concretamente, los mercados de apuestas de cada evento deportivo se consideran informacionalmente eficientes (Fama, 1970), ya que en la situación inicial M_0 se la que se tuvo en cuenta toda la información pasada disponible, y en el transcurso de este periodo los apostantes trasladan sus expectativas al valor de cotización. La respuesta de los mercados ante noticias deportivas, ruedas de prensa o acontecimientos relacionados con el evento futuro es cuasi inmediata. De esta manera, la cotización de las apuestas recogerá más información a medida que se acerque M_F .

Las expectativas de los apostantes se consideran, a priori, racionales, ya que se presupone que buscan minimizar el riesgo y maximizar sus beneficios (Friedman, 1953). Esta racionalidad en la toma de decisiones es relativa, ya que, individualmente ningún apostante poseerá toda la información futura sobre el desenlace del encuentro, ni se presupone que todos los individuos sean racionales en todo momento. Cada apostante tendrá en cuenta la información histórica (ya tenida en cuenta en la formación del precio de la apuesta), los flujos de información que reciba y las expectativas individuales respecto al posible desenlace del evento deportivo. Tomando esta información, el apostante decidirá si el precio de la apuesta es representativo de valor respecto al diferencial que suponen sus expectativas individuales (Becker, Gary S. 1993). Por lo tanto, si la información histórica disponible es la misma para las casas de apuestas, que fijan los precios de cotización inicial, y para los potenciales apostantes, ¿qué elemento diferencial provoca que las cotizaciones varíen en el periodo comprendido entre M_0 y M_F ?. Cada apostante individual podrá tener en cuenta un número limitado de variables para realizar una estimación de cómo afectarán al desenlace del mismo (o sencillamente, seguirán corazonadas o sentimientos). Estas estimaciones son el elemento diferenciador para cada cada de apuestas, ya que provocan que las cotizaciones de las apuestas oscilen; si muchos apostantes de una casa de apuestas concreta parecen contrarios a las cuotas de partida, la casa de apuestas ajusta dichas cuotas. En su conjunto los precios de las apuestas reflejarán las estimaciones del conjunto de apostantes. Si se toma como válida la “sabiduría popular” que influye directamente en la formación de las cotizaciones en el periodo comprendido entre M_0 y M_F , se estaría alcanzando un grado de eficiencia informacional fuerte en la formación de precios (Fama, 1970), en el que se tienen en cuenta la información histórica disponible, los nuevos flujos de información y noticias, que actualizan los precios de las apuestas, y las expectativas conjuntas del futuro desenlace del evento deportivo. A medida que se acerque M_F , mayor cantidad de información estará disponible y mejor reflejarán las cotizaciones la probabilidad de ocurrencia de cada desenlace.

El grado de eficiencia del mercado de oferta de apuestas deportivas se ha incrementado mucho en los últimos años gracias a las nuevas tecnologías de integración, que permite a las casas de apuestas interconectar sus sistemas para que las diferencias entre las cotizaciones que ofrece una casa y otra, nunca

sea excesiva. Hasta hace poco tiempo, la forma más común de aprovechar estas ineficiencias informacionales era utilizando una estrategias de arbitraje, las cuales son estrategias oportunistas que consisten en aprovechar las diferencia de precios de apuestas cotizadas en distintos mercados para obtener una inversión global, sobre dicha apuesta, libre de riesgo y con una rentabilidad equivalente a la divergencia de precios entre los distintos mercados (Economipedia, 2018). Un claro ejemplo de arbitraje aplicado a las apuestas deportivas, en concreto a las *apuestas 1X2*:

Tabla 1. Arbitraje de *apuestas 1X2*

	Cot. local	Cot. empate	Cot. visitante	Margen
Casa 1	1,8	3,2	5,1	6,41%
Casa 2	1,7	3,8	5,0	5,13%
Casa 3	1,7	3,5	5,5	5,57%
Casa 4	1,85	3,0	5,0	5,69%
Arbitraje	1,85	3,8	5,5	-1,44%

La Tabla 1 muestra las cotizaciones de las apuestas 1X2 que ofertan distintas casas de apuestas. Existen diferencias tales entre las cotizaciones de las casas, que si se colocara una apuesta a la cotización más alta de cada desenlace posible (gana local, gana visitante y empate), se estarían cubriendo el 100% de los escenarios posibles al final del evento deportivo, y se habría conseguido una rentabilidad potencial > 0 . Es decir, aprovechando un desequilibrio del mercado, nos habríamos posicionado con una apuesta global, libre de riesgo y con una rentabilidad media de 1,44% sobre el capital medio invertido.

Por obvio que parezca, cabe destacar que las casas de apuestas siempre ganan, independientemente de quién apueste a qué. Naturalmente pierden muchas apuestas, pero la Ley de los grandes números (Poisson, 1837) garantiza que, dado el volumen tan alto de transacciones, las pérdidas se compensan con las ganancias. Es más, dado que las casas de apuestas son generadoras de mercado, se permiten asignar un margen sobre cada tipo de apuesta. Si se calcula la probabilidad de éxito de cada uno de los desenlaces posibles para una apuesta dada, la relación entre las cuotas y la probabilidad de ocurrencia de dicho suceso, es inversamente proporcional.

$$Probabilidad\ ocurrencia = \frac{1}{Valor\ de\ cotización} = \frac{Inversión}{Rentabilidad} \quad (1)$$

Si ampliamos la Ecuación 1 al conjunto total de escenarios posibles, o desenlaces, que pueden tener lugar en un evento deportivo, para un tipo de apuesta dada, se debería de cumplir:

$$1 = \sum \frac{1}{Valor\ cotización\ escenarios\ posibles} \quad (2)$$

Si la Ecuación 2 fuera cierta, estaríamos hablando de “juego justo” o “juego equitativo” (Fama, 1970), que se podría definir como aquel sistema en el que la rentabilidad potencial de una inversión se ha de corresponder de forma proporcional al riesgo asumido por el inversor. Sin embargo, como se ha comentado, este no es el caso, ya que las casas de apuestas diluyen sus márgenes directamente en las cuotas ofertadas al público.

$$1 + Margen = \sum \frac{1}{Valor\ cotización\ escenarios\ posibles} \quad (3)$$

Por lo tanto, en M_0 se parte de una situación de desequilibrio que favorece a la casa de apuestas y perjudica al apostante, que asume un coste (margen), por colocar su apuesta. Este margen es aplicable al conjunto de desenlaces posibles de un evento deportivo, y será variable en función de la casa de apuestas y el tipo de apuesta.

Con estas premisas, y sabiendo que las casas de apuestas son generadoras de mercado y solo intervienen en la asignación de precios a las apuestas en M_0 , se puede afirmar que el comportamiento de las cotizaciones se debe a los efectos de la demanda. En la medida que aumenta la demanda por una apuesta determinada, se reduce su cuota y viceversa. Los márgenes de las casas de apuestas para un tipo de apuesta determinado tiende a ser un valor prefijado por la casa en base a la estimación del número de

apuestas que se vaya a recoger. Si la demanda de una apuesta determinada sobrepasase las estimaciones, se reduciría el margen de la casa que obtiene la casa, y viceversa.

3. Existencia de arbitraje intertemporal

El primer propósito del estudio es verificar las hipótesis planteadas en la introducción. Para corroborar que existe posibilidad de arbitraje mediante la realización de apuestas en distintos momentos, en situaciones reales de cotizaciones ofrecidas por las casas, se ha realizado un estudio sobre las cotizaciones de las *apuestas 1X2* ofrecidas por Sportium de todos los partidos de la Liga de Fútbol Profesional de Primera División, o Liga Santander, en el periodo de tiempo comprendido entre el 9 de diciembre de 2017 y el 4 de abril de 2018 (Sportium.es).

Para realizar la obtención de los datos se utilizó lo que se conoce como *Web Scraping* (Marble, W. 2016). En la actualidad existen muchas páginas webs con mucha información que, sin embargo, no ofrecen una herramienta de acceso a dicha información, tal como una API, ni comparten la misma en un archivo tabulado como un .csv, .xlsx, .txt, etc. El *scraping* es una técnica informática mediante la cual un programa navega de forma autónoma por una web, escaneando determinados contenidos que recoge y almacena en forma de tabla para ser posteriormente procesados. Para este estudio el *scraping* se realizó mediante el lenguaje de programación *R*, que fue programado para entrar en la citada web de sportium y reconocer y almacenar los datos sobre apuestas que la propia web comparte. Más concretamente, se utilizó el paquete *rvest*, que permite acceder a determinadas partes de una página web, siempre que se conozca el identificador de dicha parte de la web. Mirando el código fuente de la web de sportium, los autores pudieron determinar que los conjuntos de datos de interés se hallaban etiquetados como “#main-area .seln-name” y “#main-area .dec”. Así, con el código:

```
library(rvest)

web <- 'http://sports.sportium.es/es/t/45211/La-Liga'

equipos <- read_html(web) %>% html_nodes("#main-area .seln-name") %>% html_text()

ratio <- read_html(web) %>% html_nodes("#main-area .dec") %>% html_text()
```

se obtuvieron cada una de las siguiente variables para todos los eventos deportivos de la muestra:

- Fecha y hora en la que tendría lugar el evento deportivo
- Nombre de los equipos local y visitante
- Valor de cotización de las apuestas 1X2 en la plataforma online
- Momento de captura de cada registro

Este proceso se automatizó para que se ejecutase cada minuto, extrayendo así una captura de toda la información sobre apuestas para cada minuto durante el intervalo de estudio. Se almacenaron los datos de 175 partidos durante 17 jornadas de La Liga, reuniendo un total de 2.146.681 registros de cada variable capturada. En media, se cuenta con 12.266 registros de cada evento deportivo, lo que supone un tiempo de cotización medio en la plataforma online de 8,5 días.

En primer lugar se calculó el margen que obtenía la casa de apuestas en cada uno de los eventos deportivos registrados. El margen medio de los 175 eventos deportivos obtenido por la casa de apuestas, en este tipo de apuesta, fue un 5,79%. Esto quiere decir que, independientemente de los desenlaces de esos 175 eventos deportivos, la casa de apuestas habría ingresado como beneficio, por término medio, del 5,79% de todas las cantidades jugadas a este tipo de apuestas, en ese periodo. Esto corrobora lo que ya sabíamos de forma teórica: “la casa siempre gana”, pero esto no se opone a que un jugador pueda hacer una selección de tiempos adecuada y conseguir ganar en un subconjunto de apuestas.

Para ello se calcula en primer lugar el margen que obtendría la casa de apuestas, por término medio, en el caso que fuéramos capaces de apostar en cada partido, en los momentos en los que las cotizaciones alcanzan su valor máximo. Se pueden dar tres situaciones:

- Margen de la casa > 1
- Margen de la casa ≈ 1

- Margen de la casa < 1

El primer caso corresponde con la situación habitual en la que, si en un único momento aleatorio en el tiempo de cotización de la apuesta, se colocan apuestas a los tres escenarios posibles, la casa obtendrá el margen correspondiente. El segundo caso, en el que el margen medio de la casa se aproximase mucho a la unidad, la casa de apuestas no obtendría margen (exceso de la unidad) y el apostante no podría utilizar la estrategia para beneficiarse, ya que el margen medio no sería inferior a la unidad. El tercer caso es aquel en el que, con la utilización de las cotizaciones máximas de los tres escenarios, el apostante podría llegar a obtener margen de beneficio reduciendo al máximo el riesgo.

Una vez analizados los datos de la muestra completa se obtendría un margen medio de 0,99754, lo que significa que, tomando las cotizaciones máximas de todos los partidos analizados, sería posible que un inversor obtuviese un beneficio sistemático medio, libre de riesgo, del 0,2456%. Queda por tanto probado que la estrategia es viable, siempre y cuando se conozcan los mejores momentos para apostar. El resto del artículo se centra, precisamente en tratar de especificar a priori cómo encontrar estos momentos óptimos.

4. Selección de partidos favorables.

Se ha demostrado que es teóricamente posible ganar a la banca (en media). Se analiza ahora la posibilidad de llevar esto a la práctica, seleccionando eventos que sean especialmente favorables para el apostante, y así maximizar la posibilidad de obtener un beneficio sistemático. Durante el resto de la sección subyace permanentemente la idea de eliminar eventos de la muestra siempre y cuando sea con un criterio elegido *a priori*. La razón para ello es que si un cierto tipo de partidos suele ser especialmente favorable para la casa, el apostante puede empezar por descartar completamente dichos partidos de su estrategia. Por ejemplo, si en la base de datos recogida hubiera un partido que sólo estuvo ofertado durante 5 minutos, lo descartaríamos por completo, porque no habrá tiempo suficiente para que dicha cotización varíe tanto como deseamos. Si el objetivo fuera hacer un análisis descriptivo de todo el sector de apuestas por supuesto sería inaceptable eliminar parte de la muestra, sin embargo, para el diseño de una estrategia es válido ya que el interés no es conocer todos los casos sino poder especificar de antemano algunos que sean favorables al apostante.

En primer lugar, no todas las apuestas de los partidos observados han cotizado el mismo periodo de tiempo, por lo que habrá cotizaciones de apuestas de ciertos partidos de las cuales no se podrán observar variaciones suficientemente grandes. Por este motivo se decidió segmentar la muestra y utilizar como objeto de estudio las cotizaciones de aquellos partidos de las que se tiene un número *aceptable* de observaciones.

Se pretende asemejar la muestra lo máximo posible a una situación real, por lo que el primer criterio restrictivo de segmentación es seleccionar aquellos partidos en los que se tiene un número de observaciones igual o superior al número de observaciones medio en la muestra. El número de observaciones medio por partido es 12.267, y son 89 eventos deportivos aquellos en los que se tienen datos con una frecuencia igual o superior a la citada. El margen medio en este caso sería de 0,98703, por lo que, si un inversor siguiera la estrategia propuesta, solo en aquellos partidos en los que el número de observaciones minuto a minuto es igual o superior a 8,5 días, obtendría un beneficio sistemático medio y libre de riesgo del 1,297%. Con esta restricción, se obtiene un beneficio 5 veces superior al de la muestra completa. Es, además, una restricción muy razonable y fácil de aplicar: “apueste solo si faltan al menos 8 días para el evento”, si no, podría ocurrir que no haya suficiente variabilidad. Si se restringe aun más el criterio de segmentación, aumentando la frecuencia observada deseada a 28.080 observaciones, o lo que es lo mismo, 19,5 días de cotización, el margen medio alcanzado en el sesgo formado por 6 partidos, sería de 0,98339. Esto significaría que, un inversor que siguiera la estrategia propuesta obtendría un beneficio sistemático medio y libre de riesgo del 1,661%.

A pesar de que, por término medio, el resultado sea positivo para el inversor, eso no quiere decir que en todos los partidos se den situaciones de arbitraje aprovechables por esta estrategia. De hecho, solo el 50% de la muestra presenta situaciones de arbitraje. Si seleccionamos los 87 partidos de la muestra en los que es posible aplicar la estrategia de forma positiva, el margen medio sería de 0,97692, lo que significaría que el apostante podría obtener un beneficio sistemático medio y libre de riesgo del 2,308%.

A pesar de que actualmente no se han llegado a definir de manera completa y exhaustiva qué aspectos previos a M_0 son determinantes para poder encontrar partidos en los que se podrían dar situaciones de arbitraje durante el tiempo de cotización, sí que se tienen ciertas nociones al respecto.

Entendemos que las variaciones en las cotizaciones se producen como efecto de las apuestas realizadas durante el periodo de cotización. Estas apuestas son fruto de los flujos de información en el mercado y de las expectativas racionales de los apostantes. Estas expectativas se han de considerar aleatorias ya que no todos tenemos el mismo perfil de riesgo, ni somos hinchas de los mismos equipos, ni asimilamos la información de la misma manera. Por lo tanto, si el origen de las variaciones de las cotizaciones es aleatorio, el efecto en las cotizaciones deberá ser igualmente aleatorio. Los movimientos en las cotizaciones son aleatorios, pero son, al fin y al cabo, movimientos. Esto favorece la estrategia propuesta ya que, se pretende buscar el arbitraje en los máximos individuales de cada uno de los escenarios posibles. A mayores movimientos, mayores máximos se generarán y más rentable podría salirle la operación al apostante. Pero, ¿de qué manera se comportan las curvas de cotización una respecto a la otra?

Partimos de la ecuación (3) de “juego justo” o “juego equitativo” real, es decir, aquella en la que se tiene en cuenta el margen de la casa de apuestas para el cálculo de las cotizaciones. Dicha ecuación, desarrollada específicamente para el caso de nuestro estudio es:

$$1 + \text{Margen} = \frac{1}{\text{Cot.Local}} + \frac{1}{\text{Cot.Empate}} + \frac{1}{\text{Cot.Visitante}} \quad (4)$$

Si suponemos que el margen que la casa desea obtener en un partido determinado es constante, entonces los movimientos de las cotizaciones de un escenario deberán ser explicados por los movimientos de los otros dos escenarios restantes. Se realizó una regresión lineal multidimensional con las inversas de las tres cotizaciones, para todos los datos de la muestra. Los resultados arrojaron que la varianza de cada una de las variables es explicada por las varianzas de las otras dos en un 99,7%. Las casas de apuestas publican cotizaciones con hasta dos decimales, eso explicaría que la correlación no sea perfecta. Sin embargo, se puede afirmar, con un alto grado de certeza, que las casas mantienen un margen constante para cada partido y que las variaciones en cada una de las cotizaciones afectan a las variaciones de las otras dos cotizaciones. Esto significa que estamos ante un sistema cerrado, en el que, si únicamente se apostase por uno de los escenarios posibles, su cotización bajaría debido al aumento de demanda, y las cotizaciones de los otros dos escenarios aumentaría, en unas proporciones indeterminadas por el momento, generando máximos y posibilitando la aparición de arbitraje.

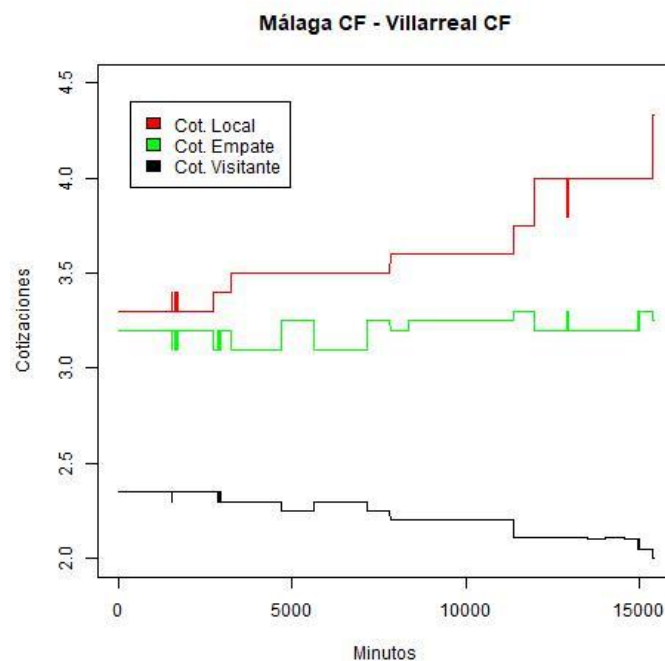


Figura 1. Evolución de las cotizaciones a lo largo del tiempo previo al evento

Para la Figura 1 se ha escogido el partido Málaga-Villareal, y en concreto la Figura de sus cotizaciones, porque dentro del 50% de partidos de la muestra que presentaban características favorables para el estudio, se ha considerado que esta muestra, de forma muy ilustrativa, como se producen las variaciones en las cotizaciones. La Figura muestra las cotizaciones de los tres escenarios de apuesta, para el partido del Málaga CF contra el Villarreal CF que tuvo lugar el 1 de Abril de 2018. Como se puede observar en la Figura, la cotización del equipo local y la cotización del equipo visitante siguen movimientos inversos respecto a un eje ficticio. A medida que se acerca el inicio del partido, el favorito se convierte en más favorito. La cotización del empate se ve afectada por las variaciones de las otras dos apuestas de forma conjunta y no sigue un patrón claro de comportamiento. Las apuestas cotizaron aproximadamente 11 días naturales, en los cuales se recogieron datos que arrojaron la siguiente información:

Tabla 2. Estadísticos descriptivos

	Cotización Media	Variación	%	Máximo	Mínimo
Local	3,61	0,72	19,94	4,33	3,30
Empate	3,20	0,10	3,12	3,30	3,10
Visitante	2,23	0,23	10,31	2,35	2,00

Las variaciones porcentuales presentadas en la Tabla 2 son mayores cuanto mayor es la cotización de la apuesta, sin embargo, cuanto mayor es la cotización de una apuesta, menos afectan sus variaciones al incremento o decremento del margen global. Por lo tanto, para favorecer la estrategia propuesta es necesario que la cotización del equipo favorito, o aquel cuya cotización es la menor, varíe lo suficiente de forma porcentual como para que se de una situación de arbitraje en la que el apostante obtenga un margen positivo respecto a la casa de apuestas (margen < 1). El margen medio de la casa (calculado como en la Ecuación 4) durante el partido de la Figura 1 fue de 4,03%, pero gracias a que la variabilidad de las tres apuestas es elevada, un inversor llegaría a obtener un margen positivo sobre la casa del 4,05%, es decir, las variaciones en las cotizaciones fueron lo suficientemente grandes como para que apareciesen máximos tales que, apostando en esos momentos temporales, se conseguiría reducir el margen estático de la casa en un 8,08%.

Se ha conseguido demostrar que las variaciones de las cotizaciones de cada uno de los escenarios depende, exclusivamente, de las variaciones de los otros dos escenarios. Partiendo de esa premisa, y con el objetivo de sentar las bases para desarrollar una estrategia que retorne beneficios sistemáticos, se comprueba que sí existen dichas situaciones en las que se puede realizar arbitraje apalancándonos en el tiempo de cotización y en los momentos de aparición de máximos.

5. Optimización temporal de las apuestas

Una vez que se ha probado que existe la posibilidad real de “ganar a la banca” aprovechando los movimientos de las apuestas durante su periodo de cotización, se tratará de encontrar los momentos temporales en los cuales las cotizaciones de los tres escenarios alcanzan sus máximos. Si se pudiera predecir esto, se tendría una estrategia infalible de apuesta. La caracterización de tales momentos es difícil y, por supuesto, es siempre una predicción probabilística. Si la caracterización fuera lo suficientemente buena, la incertidumbre generada por dicha probabilidad podría ser compensada con unas ganancias lo suficientemente habituales como para poder asegurar un beneficio regular a largo plazo. Esta sección desarrolla ideas que se encuentran aún en construcción y los autores no pueden ofrecer aún una estrategia ganadora, aunque sí que se exponen algunas ideas y nociones en esta dirección.

La primera aproximación consiste en determinar en qué días es más frecuente encontrar máximos para las cotizaciones, esto es, estudiar la frecuencia con la que aparecen máximos para las cotizaciones de los tres escenarios, en los días previos al inicio del partido. Para este análisis se estudian los momentos óptimos con respecto al inicio de los partidos. Así, el “momento 1” se refiere a las 24h anteriores al partido, el “momento 2” a las 24 anteriores y así sucesivamente. La Tabla 3 muestra el número de veces que las cotizaciones alcanzaron sus máximos, ordenado por el número de días que faltaban para el inicio del partido:

Tabla 3. Frecuencia de aparición de máximos en las cotizaciones

	1	2	3	4	5	6	7	8	9
Cot. Local	12545	15189	13631	14437	11287	8699	8054	6833	7571
Cot. Empate	12071	27031	24580	25293	25077	24453	23747	19788	19298
Cot. Visitante	19441	25873	21795	20977	20201	15413	13346	11430	15656

Siguiendo los hallazgos de la sección 3, se utilizan para desarrollar la estrategia aquellos partidos con un número de observaciones igual o superior al número de observaciones medio de la muestra. Esto nos garantiza que los partidos incluidos en este análisis cotizaron durante todo el periodo objeto de estudio.

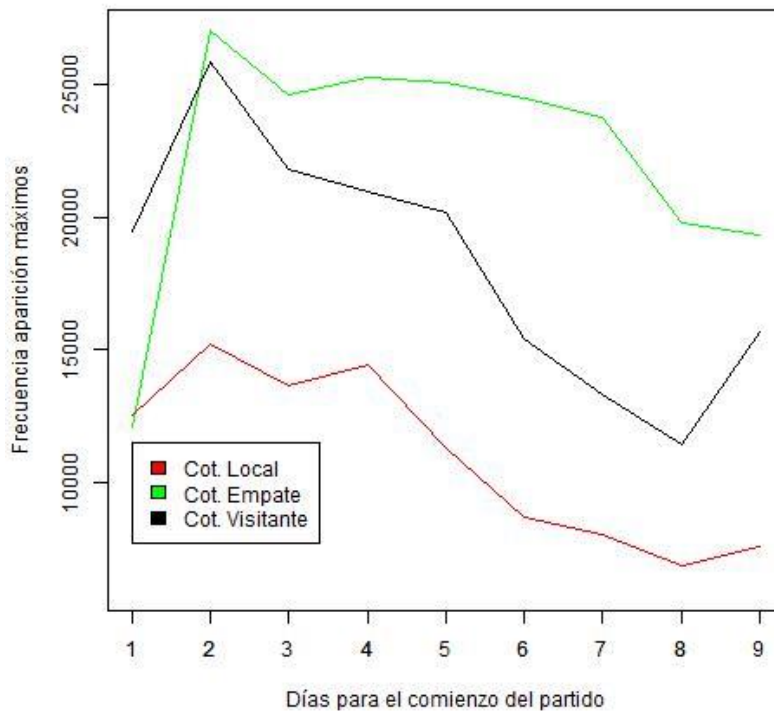


Figura 2. Frecuencia de la aparición máximos con respecto al número de días restantes hasta el comienzo del partido.

Para las tres cotizaciones se observa una clara tendencia alcista a medida que se acerca el inicio del evento deportivo, siendo lo más frecuente que el máximo se alcance en el período a falta de 24h a 48h. Para poder analizar más detenidamente esta tendencia, se ha obtenido una tabla de frecuencias normalizadas a la frecuencia máxima. De esta manera se puede observar como desde el primer día de cotización observado en la muestra, el mercado se mueve cada vez más, generando la aparición de mayor número de máximos a medida que se acerca el momento de inicio del partido M_F .

Se trató de analizar más en detalle el “momento 2”, es decir, el período de 24h a 48h previas al comienzo del encuentro, ya que para las tres cotizaciones es el intervalo temporal en el que más máximos se alcanzan (ver Tabla 3). Los resultados no fueron concluyentes ya que, la aparición de máximos se producía casi igualmente en todos los submomentos. Este resultado podría ser atribuible al enfoque del estudio. Al analizar un conjunto no homogéneo de partidos, en el que, la heterogeneidad viene dada por ciertas características de los partidos, tales como si el equipo favorito juega como local o visitante. Esto condicionará la forma de cotización de las apuestas, y en consecuencia, el momento y cantidad de aparición de máximos.

En un segundo análisis se trató de contrastar si las distribuciones de frecuencias observadas de aparición de máximos se asemejaban a una distribución teórica de probabilidad. Las similitudes de simetría y curtosis nos indican que la curva de la cotización del empate se asemeja a una distribución de frecuencias uniforme, mientras que las curvas de cotización del local y el visitante se asemejan a una distribución beta.

Se realizaron contrastes de semejanza de distribuciones mediante las pruebas Chi cuadrado y Kolmogorov-Smirnoff simuladas mediante Monte Carlo, obteniendo p-valores de 0,706 y 0,698 no siendo posible, por tanto, rechazar la hipótesis de que las distribuciones observadas se ajustasen a las distribuciones mencionadas. A pesar de que las distribuciones observadas no se asemejen a una distribución teórica que permita explicar de forma más concreta su comportamiento, el análisis sigue arrojando luz sobre la realidad subyacente en línea con lo comentado anteriormente. No obstante, la naturaleza no homogénea del conjunto de partidos dificulta el análisis y quedan variedad de líneas de investigación abiertas para el futuro.

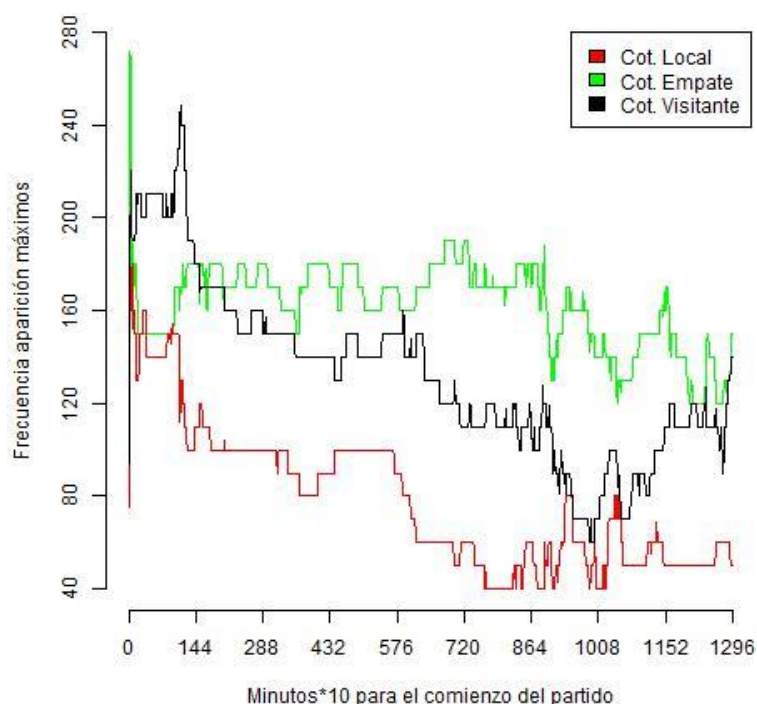


Figura 3. Frecuencia de aparición de máximos en intervalos de 10 minutos previos al comienzo del evento.

Adicionalmente, se calculó la frecuencia de aparición de máximos en intervalos de 10 minutos, en el periodo comprendido entre los 9 días naturales de cotización de la muestra seleccionada anteriormente, y el inicio del partido. El objetivo perseguido en este caso fue inferir a la muestra los resultados obtenidos de los momentos óptimos para encontrar un máximo, partido a partido. En el caso de la cotización del local, el momento en el que frecuentemente se espera encontrar mayor número de máximos es el momento 7, el cual corresponde con 70 minutos antes del comienzo de cada partidos. En el caso de la cotización del empate, el momento 111 es el óptimo, y en el caso de la cotización del visitante, se alcanzaría el óptimo 20 minutos antes del inicio de cada partido.

Se tomó el valor de cotización, de cada apuesta, para cada partido, en los momentos 7, 111 y 2 respectivamente y se calculó el margen del inversor en cada uno de los casos. En el 42% casos se conseguía reducir notoriamente el margen estático de la casa de apuestas, y en el 58% restante el margen estático aumentaba. En ningún caso, tomando los valores de cotización de esos momentos, se obtuvo una apuesta global a un partido que beneficiase al inversor. Cabe destacar que los tres momentos óptimos se encuentran en el intervalo 0-24h antes del inicio del partido. Pocos minutos antes del inicio el partido se

observa una gran volatilidad en la aparición de máximos, lo cual dificulta el análisis de los resultados tomando estos momentos.

El resultado obtenido es completamente coherente con los análisis previos y pone de manifiesto la necesidad de caracterizar con más detalle el comportamiento de las cotizaciones. Una de las posibles propuestas de trabajo futuras mencionadas anteriormente sería estudiar minuciosamente cada partido para poder tratar con una cartera de partidos homogénea, ya que, a medida que se identifiquen las variables que afectan positivamente al margen del inversor, se podrá crear una cartera tipificada de partidos en los que se presentarán más habitualmente situaciones de arbitraje aprovechables por el inversor. A medida que se tipifiquen los partidos y se analicen sus periodos de cotización en conjuntos más homogéneos, los autores esperan poder identificar momentos óptimos en los cuales, la probabilidad de encontrarse con máximos supere las pérdidas potenciales, de tal manera que sea posible, en el futuro, desarrollar una estrategia de balance sistemático positivo bajo esta dirección.

La otra línea de investigación abierta y quizás la de mayor relevancia para el desarrollo real de la estrategia teórica propuesta, es la determinación de los momentos óptimos de apuesta. Definidos estos momentos como aquellos en los que cada apuesta, individualmente, presentan mayor probabilidad de llegar a su máximo durante el periodo de cotización. Para poder determinar qué variables cualitativas afectan a que dicho momento varíe de un partido a otro será necesario llevar a cabo otro estudio más exhaustivo. Una vez definidas dichas variables predictivas que expliquen el momento óptimo en un número relevante de ensayos (>50%) se procedería a la carterización de los eventos según sean explicados por dichas variables en mayor o menor medida.

6. Conclusiones

La posibilidad de utilizar la minería de datos es de gran ayuda para estudiar el mercado de apuestas. Si bien era conocida la posibilidad de arbitraje entre casas de apuestas, estas herramientas han permitido cerciorar nuevas situaciones.

Se ha corroborado empíricamente que las cotizaciones de las casas de apuestas se calculan a partir de una situación de juego justo a la que se introduce un margen favorable a la casa de apuestas. A pesar del margen, se demuestra que es posible realizar arbitraje en un mismo operador de juego realizando apuestas en distintos momentos temporales.

Adicionalmente, se muestra una serie de filtros que podemos aplicar a priori sobre las apuestas para mejorar consistentemente el beneficio obtenido por el inversor, siempre y cuando se pudiera prever el momento óptimo de realizar la apuesta. Este resultado es fundamental, ya que permite hacer una elección a priori de en qué apuestas participar y podría ser indispensable como arma para compensar la inevitable incertidumbre que provoca la estimación del momento óptimo de apuesta.

Bibliografía

1. Aluja, T. 2001, “*La Minería de Datos, entre la Estadística y la Inteligencia Artificial*”, Universidad Politécnica de Catalunya, QUESTIÓ, vol. 25, 3, p. 479-498.
2. Becker, Gary S. 1993, “*Nobel Lecture: the economic way of looking at behavior*”, Journal of Political Economy 101, 385-409.
3. Boyle, Malcolm. 2006, “*The Art of Bookmaking*”, High Stakes Publishing, ISBN: 1843440261
4. Christie, Gary. 2006, “*The Football Betting Science*”, Arima Publishing. ISBN: 1845491181
5. Dirección General de Ordenación del Juego. (2017). Memoria actividad anual.
6. Economipedia. (2 de Agosto de 2018). Obtenido de <http://economipedia.com/definiciones/arbitraje-financiero.html>
7. Martínez Élices, Á. (2016). “*Apuestas deportivas: una aproximación desde el análisis económico*”. Universidad de Cantabria.
8. Fama, E.F. (1965), “*The Behavior of Stock-market Prices*”, The journal of Business, vol. 38, no. 1, pp. 34-105.
9. Fama, E. F. (1970). “*Efficient Capital Markets: A Review of Theory and Empirical Work*”. Journal of Finance.

10. Friedman, M. & Savage, L.J. 1948, "*The utility analysis of choices involving risk*", The journal of political economy.
11. Kealy, Paul. 2005, "*The Definitive Guide to Betting Exchanges*", Raceform Ltd. ISBN: 1905153023.
12. Levitt, S.D. 2004, "*Why are gambling markets organized so differently from financial markets?*", The Economic Journal, vol. 114, no. 495, pp. 223-246.
13. Lucas, R.E. 1972, "*Expectations and the neutrality of money*", Journal of Economic Theory 4, 103-124
14. Marble, William, 2016, "*Web Scraping With R*", Stanford University, August 11, 2016.
15. Pérez, L. (2010), "*El mercado de apuestas deportivas*", en Palomar, A. (director) Las apuestas deportivas (pp. 13-33) Thomson Reuters Aranzadi: Navarra.
16. Poisson, S. (1837). Ley de los grandes números.
17. Sportium.es. (9 de Diciembre 2017 al 22 de Febrero 2018). Obtenido de <http://sportium.es>.
18. Vaughan Williams, Leighton. 2004, "*Betting to Win: A Professional Guide to Profitable Betting*", High Stakes Publishing. ISBN: 1843440156.
19. Wood, R.T. & Williams, R.J. 2007, "*Internet gambling: Past, present and future*". Research and Measurement Issues in Gambling Studies (pp. 491-514). Burlington, MA: Elsevier.