

**FACTORES SOCIOECONÓMICOS REGIONALES QUE INFLUYEN EN EL ÉXITO DEPORTIVO DE
LOS EQUIPOS DE FÚTBOL PROFESIONALES**

CRISTIAN CORRALES RAMÍREZ

UNIVERSIDAD DEL VALLE

FACULTAD DE CIENCIAS SOCIALES Y ECONÓMICAS

DEPARTAMENTO DE ECONOMIA

SANTIAGO DE CALI

2015

**FACTORES SOCIOECONÓMICOS REGIONALES QUE INFLUYEN EN EL ÉXITO DEPORTIVO DE
LOS EQUIPOS DE FÚTBOL PROFESIONALES**

CRISTIAN CORRALES RAMÍREZ

Trabajo de grado presentado para optar al título de economista

Director:

JORGE MARIO URIBE

UNIVERSIDAD DEL VALLE

FACULTAD DE CIENCIAS SOCIALES Y ECONÓMICAS

DEPARTAMENTO DE ECONOMIA

SANTIAGO DE CALI

2015

Contenido

Resumen.....	1
I. Introducción	1
II. Marco Teórico	5
III. Literatura Relacionada.	7
IV. Análisis Empírico	12
A. Metodología	12
B. Efectos Dinámicos.	15
C. Datos	20
D. Resultados	21
V. Conclusiones.....	30
VI. REFERENCIAS.....	33
VII. ANEXOS	35

Resumen.

En este trabajo exploramos el enlace entre el desarrollo socioeconómico regional y el desempeño deportivo de los equipos de fútbol. Se identifican dos principales factores detrás de este enlace: el ingreso per cápita y la población de la región. Además, se tienen en cuenta factores socioeconómicos adicionales que pueden interferir en la distribución de talento, tales como los niveles de turismo, de urbanización, entre otros. Se prueba la teoría usando un panel de las regiones europeas que tienen equipos participantes en la UEFA Champions League. Se encuentra consistentemente soporte a lo predicho por la teoría donde regiones con mayor población y niveles de ingreso tienen equipos más competitivos. Se encuentra evidencia de los efectos de otras variables económicas que influyen en dicho desempeño.

I. Introducción

En la actualidad, el fútbol puede ser considerado el deporte más popular en todo el mundo, despertando la pasión de multitudes y atrayendo la mirada de muchos y muchas, desde el más chico hasta el más grande, sin importar el estrato socioeconómico de los individuos, convirtiéndose así en un deporte que está lleno de matices sociales. De esta manera, las personas que están directamente relacionadas con el fútbol, como entrenadores, directores deportivos, agentes de jugadores, comentaristas deportivos, periodistas, entre otros, se ven en la obligación de involucrarse más en temas sociales y económicos y no solamente en los estrictamente deportivos.

Bajo este contexto, los comentarios acerca de “las verdades” del fútbol abundan entre sus seguidores y como dicen Kuper y Szymanski (2009) en su libro titulado “Soccernomics” este deporte está lleno de clichés que nadie analiza, frases comunes como “el equipo perdió porque el entrenador no está a la altura” o “el fútbol es un negocio”, entre otras, que se escuchan entre los aficionados. Poco se tiene en cuenta que las afirmaciones de ese tipo tienen una explicación y un trasfondo socioeconómico, más que las corazonadas de los seguidores de un equipo de fútbol.

Es así como los métodos estadísticos y económicos permiten tener una mayor claridad sobre el porqué a algunos equipos de fútbol les va mejor que a otros, ya que “buena parte de lo que ocurre en el fútbol se puede explicar –e incluso predecir– mediante el estudio de los datos, incluso de aquellos datos ajenos a este deporte” Kuper y Szymanski (2009).

De esta manera, se busca establecer la relación que existe entre el desarrollo socioeconómico regional y el éxito de los clubes de fútbol profesional, siendo relevante por su avance en la investigación de carácter regional, ya que en la literatura se encuentran estudios sobre la relación entre desarrollo socioeconómico y los equipos nacionales de fútbol, pero poco se ha avanzado en esas mismas implicaciones de desarrollo en los equipos de fútbol por regiones.

Asimismo, la investigación cobra aún más relevancia al hacer un esfuerzo por integrar la disciplina económica con el deporte. De esta forma, permite aplicar métodos y técnicas estadísticas y econométricas al fútbol y sacar conclusiones, que por su carácter interdisciplinario, son de interés para varios campos, tales como el académico, cultural,

empresarial y la gerencia deportiva. Del mismo modo es de gran interés para los clubes de fútbol a la hora tomar decisiones de *marketing*.

El fútbol es un deporte que nos ofrece una gran cantidad de datos por partido, ya que semanalmente se juegan muchos encuentros futbolísticos que proporcionan dichos datos. Por pertinencia y por disponibilidad de las variables socioeconómicas regionales, y con el fin de tener unos resultados más consistentes, la investigación está enfocada en los clubes de fútbol de las regiones europeas, dejando abierta la posibilidad de hacer la investigación en los clubes de fútbol latinoamericanos y las selecciones nacionales de fútbol en el futuro.

Finalmente, esta investigación pretende iniciar, como ya se ha mencionado anteriormente, una relación entre la economía y el deporte, ya que poco se ha desarrollado en Colombia (y el mundo), trabajando en pro de forjar el carácter interdisciplinario de la economía con el fin de extender el campo de acción de la profesión a diferentes áreas del conocimiento.

De igual manera, se pretende incentivar a los demás investigadores del área económica y financiera, a que en la medida de disponibilidad de datos, se profundice esta unión entre las dos disciplinas. En ese sentido, motivar a los futuros economistas a extender las investigaciones, como se viene haciendo en otros países, a otros deportes, ya que, como dice Lewis (2004), hoy en día, no sólo el baloncesto y el fútbol americano, sino también el fútbol, el críquet y el rugby, entre otros deportes, fomentan una subcultura de personas inteligentes que los contemplan no sólo como un juego, sino como un problema que hay

que resolver, y en ese sentido, explorar este mercado que aparentemente sigue unas reglas diferentes a los mercados tradicionales.

Bajo este escenario, la investigación va dirigida a determinar por qué existen diferencias en el desempeño de los equipos de fútbol y en ese sentido, responder a una de las preguntas claves: ¿cuáles factores socioeconómicos regionales influyen en el éxito deportivo de los clubes de fútbol profesionales en Europa? Dichos factores socioeconómicos se analizan bajo el contexto regional europeo y el desempeño deportivo se mide como lo sugiere Mourão (2010). El factor diferenciador radica en la metodología, ya que en este trabajo se busca modelar las idiosincrasias y diferencias culturales de las regiones, que se quedan por fuera en la metodología usada en Mourão (2010), y que buscamos rescatar y tener en cuenta, al usar una aproximación empírica de datos panel.

Se podrán ver beneficiados los académicos al desarrollar nuevas aplicaciones metodológicas para resolver dicho problema y tener un nuevo campo de acción con mucho por resolver. Además se ven beneficiados directamente los equipos de fútbol que deberán tener en cuenta factores, tales como el tamaño de mercado al cual se enfrentan y el desarrollo económico de la región en la que se sitúan, para no efectuar operaciones riesgosas como marketing y compra de jugadores. Ignorar este tipo de investigaciones, puede llevar a los equipos a enfrentar problemas financieros al operar bajo operaciones de incertidumbre que bien podrían incrementar los costos en un momento determinado.

Lo que resta de este trabajo está organizado como sigue. En la sección II, se describe, a través de un modelo teórico simple, los diferentes enlaces que existen entre el desarrollo

socioeconómico de una región y los equipos de fútbol de la misma. En la sección III, se hace una revisión rigurosa de la literatura relacionada y se da una idea del estado del arte hasta el momento. En la sección IV se discute el modelo empírico, los problemas existentes y se presentan los resultados. La sección V concluye.

II. Marco Teórico

Hacia Un Modelo Teórico de Desarrollo Regional y Desempeño Deportivo.

Esta sección presenta un modelo teórico simple que integra los mecanismos por los cuales el desarrollo económico de una región explica el éxito deportivo de los equipos de fútbol de la región. En este sentido, seguimos de cerca los planteamientos de Bernard and Busse (2004) quienes desarrollan un modelo para explicar los determinantes del éxito en los juegos olímpicos a nivel de país, y se hace una aplicación al contexto de los equipos de fútbol a nivel regional.

Se considera primero que el talento futbolístico está distribuido aleatoriamente entre las regiones europeas, sin embargo, como lo plantea el estudio de Mourão (2010), entre otros autores ya mencionados, diferentes factores pueden interferir con esta distribución. Se considera entonces, que las regiones constituyen una división aleatoria de la población de cada país, de esta manera, se espera que el éxito deportivo de un equipo de futbol sea proporcional a la población de la región donde juega. Además, se examina el papel de los recursos económicos regionales en la distribución de talento y por consiguiente en el éxito deportivo de los equipos de la región. La idea es que para tener jugadores competitivos se requieren gastos considerables y condiciones regionales óptimas para el desarrollo de las

capacidades. Por lo tanto, el talento futbolístico de una región i en el periodo t es el resultado de combinar recursos económicos (Y_{it}), la población local (N_{it}), y una capacidad organizacional endógena (A_{it}):

$$(1) \quad T_{it} = f(N_{it}, Y_{it}, A_{it})$$

Además, se consideran otros factores económicos y sociales que pueden afectar la distribución de talento en una región dada, tales como el nivel de desempleo, el nivel de urbanización, entre otros. Dichos factores están dados por Z_{it} .

$$(2) \quad T_{it} = f(N_{it}, Y_{it}, A_{it}, Z_{it})$$

Ahora, se considera que los puntos logrados por los equipos de una región i en el periodo t , ($Puntos_{it}^*$), están en función del talento que se produce en esa región:

$$(3) \quad Puntos_{it}^* = h(T_{it})$$

Como lo mencionan Bernard and Busse (2002), no hay una guía teórica sobre la forma precisa de $f(\cdot)$ ni de $h(\cdot)$. Aquí, como en la literatura destacada, se utiliza una función de producción Cobb-Douglas para aproximar la producción de talento en una región dada. Por su parte, la relación entre el talento generado por una región y los puntos logrados por los equipos de esa región en un periodo determinado, está dada por una función logarítmica.

$$(4) \quad T_{it} = A_{it} N_{it}^{\alpha} Y_{it}^{\beta} Z_{it}^{\xi}$$

$$(5) \quad Puntos_{it}^* = \ln(T_{it})$$

Combinando las ecuaciones (4) y (5), se obtiene una especificación que describe los puntos logrados por cada región i en el periodo t :

$$(6) \quad Puntos_{it} = \begin{cases} \ln(A_{it}) + \alpha \ln(N_{it}) + \beta \ln(Y_{it}) + \xi \ln(Z_{it}), & \text{si } Puntos_{it}^* \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

La ecuación (6) puede ser probada empíricamente para los equipos que tienen su oficina principal en las regiones europeas y participaron en la UEFA Champions League durante el periodo de análisis.

III. Literatura Relacionada.

Es importante resaltar que en la actualidad existe una gran discusión acerca de la dirección de la causalidad entre el desarrollo económico regional y el éxito deportivo. Hay una gran extensión de literatura expuesta por diferentes autores quienes argumentan que el éxito deportivo y los grandes eventos deportivos en una región generan efectos sobre el desarrollo de la misma. Entre ellos se encuentran Downward y Dawson's (1999) quienes exponen que un equipo deportivo actúa como un bien público en una determinada región, generando efectos positivos tales como lealtad, amor por la región, turismo, prestigio, entre otros; además hacen una reflexión sobre la poca literatura que había hasta el momento y el énfasis que se daba al estudio de los determinantes de la demanda por el deporte. Por otra parte, si se argumenta que el desarrollo económico regional tiene efectos sobre el éxito deportivo de los equipos deportivos, se encuentra una menor cantidad de literatura que en el caso anterior, y se estaría de acuerdo con autores como Mourão (2010) que muestra cómo el éxito de los equipos de fútbol está significativamente

influenciado por factores regionales, tales como el ingreso per cápita regional, la densidad de población en la región y el grado de urbanización de la región donde se encuentra la oficina principal del equipo.

Vale la pena aclarar que el enfoque de esta investigación es el segundo que se expuso antes, es decir, en el que los factores socioeconómicos regionales generan efectos sobre el desempeño deportivo de los equipos deportivos. En ese sentido, como ya se ha mencionado, existen pocos trabajos que aborden el problema desde una perspectiva regional y por lo general se encuentran investigaciones relacionadas con las selecciones nacionales de fútbol y en momentos determinados, tales como previos a un Mundial de Fútbol o a los Juegos Olímpicos.

A nivel regional, sobresale el artículo base en esta investigación, titulado “Regional Determinants of Competitiveness: The Case of European Soccer Teams” Mourão (2010). En este documento, el autor utiliza información de 140 regiones europeas en el periodo comprendido entre 1990 y 2006. Sobre este punto, dada una mayor disponibilidad de datos a la fecha, se busca ampliar el periodo analizado para así tener resultados más robustos. Con respecto a la medida de desempeño deportivo, el autor utiliza las participaciones de los equipos de fútbol en la UEFA Champions League, aspecto que nos parece rescatable dado que es el evento de competencia futbolística con mayor demanda en Europa y en el mundo a nivel de clubes.

El modelo utilizado por Mourão (2010) es un modelo Tobit con efectos aleatorios, argumentando que la variable dependiente que mide el desempeño deportivo esta

censurada por la izquierda al no tomar valores negativos y además hace variaciones en el modelo presentando tanto el modelo Tobit con efectos aleatorios y un modelo Tobit tipo 2, exponiendo, de acuerdo a los resultados, que no existen grandes diferencias entre los dos modelos. En este aspecto nace la diferencia y la motivación para esta investigación, dado que consideramos que la metodología empleada por Mourão (2010) no recoge aspectos idiosincráticos y culturales de cada región, y como propone Andreff, Andreff y Poupaux (2008) esas diferencias culturales son relevantes para explicar las diferencias en el desempeño deportivo de los equipos de fútbol profesionales.

Los resultados obtenidos por Mourão (2010) muestran que un equipo de fútbol tiene mayor probabilidad de éxito deportivo si su oficina principal se encuentra en una región con una gran población, con un alto PIB per cápita y un nivel de urbanización elevado. Dichos resultados, nos resultan muy interesantes, y nos animan a usar una metodología, a nuestro parecer más apropiada, para contrastar estos resultados.

Siguiendo en el ámbito regional el mismo Mourão, realizó en el año 2008 una investigación, probando la misma hipótesis que en el artículo referenciado anteriormente, pero esta vez para los equipos de fútbol profesionales en Portugal. En Mourão (2008), aunque se utilizan metodologías distintas a Mourão (2010), podemos encontrar resultados muy similares. En esta ocasión, el autor concluye que hay tres factores socioeconómicos que aumentan la probabilidad de que un municipio portugués albergue la oficina principal de un equipo que juegue en primera división. Dichos factores son: el PIB per cápita, la densidad de la población, y el nivel de urbanización.

A nivel de Selecciones de fútbol nacionales, se encuentra un mayor número de investigaciones que prueben la hipótesis de la relación entre desarrollo económico y desempeño deportivo. Así entonces, encontramos publicaciones como el libro de Kuper y Szymanski (2009), en donde exponen que las Selecciones Nacionales de fútbol ganan competencias de acuerdo a la población del país y a su nivel de desarrollo económico, ya que, los autores exponen que estos factores socioeconómicos influyen en los aspectos sociales de los jugadores (como nivel de nutrición, estado de salud, educación, entre otros) y además, aspectos sociales de los seguidores del fútbol de cada país.

También en el 2009, Leeds and Leeds hacen una investigación acerca del éxito deportivo de las selecciones de fútbol, contrastando la hipótesis de que el régimen político, la herencia colonial y/o las instituciones políticas del país afectan el desempeño del equipo de fútbol nacional medido por el número de puntos que otorga la FIFA. En Leeds and Leeds (2009) evidencian una vez más que tanto el ingreso del país como su población influyen en el éxito de su selección nacional, además nos muestran que variables como, haber albergado una Copa Mundial de la FIFA y las variables políticas antes mencionadas, afectan dicho desempeño deportivo.

Hoffmann, Lee, y Ramasamy (2002) nos muestran un interesante estudio donde modelan el éxito deportivo de los países en el fútbol, medido por el número de puntos otorgados por la FIFA. Lo innovador de esta investigación es que, además de sumarse a los autores que corroboran la relevancia de factores culturales, demográficos y económicos, introducen un factor climático, argumentando que los países con climas templados tienen más éxito en los deportes en general. Una de las razones que dan los autores es que en

países con climas templados los jóvenes tienden a practicar más deportes al aire libre que en los países donde hace demasiado calor o son muy húmedos. Otro resultado muy interesante que nos presentan los autores es la U invertida entre desempeño en el fútbol y el PIB per cápita; donde inicialmente el aumento en el ingreso per cápita lleva a aumentos en los puntos recibidos por un equipo de futbol. Sin embargo, cuando el ingreso excede cierto nivel, un incremento adicional del ingreso conlleva niveles más bajos del desempeño futbolístico de los países.

Como ya se mencionó, existen muchas investigaciones para momentos determinados de los países (estudios de eventos). Es el caso de los momentos previos o posteriores a los Juegos Olímpicos. De esta manera encontramos investigaciones como las de Hawksworth (2008) o Rathke and woitek (2007), que toman como base los Juegos Olímpicos de 2002, celebrados en Salt Lake City, y encuentran que a mayores niveles de población y de PIB per cápita, los países tienen una mayor probabilidad de ganar más medallas olímpicas. En particular Hawksworth (2008) argumenta que el Ingreso per cápita está asociado a mayores niveles de nutrición y salud, lo cual es un factor determinante para obtener medallas en cualquier deporte. Además expone que el régimen político también influye en los triunfos deportivos de un país en los juegos olímpicos, modelando una variable para caracterizar si el país fue o es parte de un estado comunista (incluyendo la Unión Soviética y Cuba).

IV. Análisis Empírico

Se presenta ahora la metodología empírica para analizar los factores socioeconómicos que influyen en el éxito deportivo de los clubes de fútbol. Se describen los datos utilizados, se introduce el modelo econométrico y se presentan los principales resultados en la investigación.

A. Metodología

De manera general se tiene:

$$(7) \quad Y_{it} = \beta_0 + X'_{it}\beta + \varepsilon_{it}$$

Donde Y_{it} corresponde a la variable a explicar, que en este caso es el éxito deportivo de los equipos de la región i en el periodo t ; X'_{it} denota la matriz de variables explicativas, correspondiendo en este caso al conjunto de determinantes socioeconómicos de cada región i en el periodo t , que influyen en el éxito deportivo; β_0 y β corresponden a los parámetros a estimar; y ε_{it} denota el error aleatorio no observable. De esta forma, gran parte del trabajo empírico, se enfocará en la correcta modelación del término de perturbación que en general podría descomponerse como:

$$(8) \quad \varepsilon_{it} = \mu_i + \vartheta_{it}$$

Donde μ_i es un efecto individual no observable y ϑ_{it} es ruido blanco. Diferentes aproximaciones, que parte de supuestos distintos y cuya bondad se verifica en la práctica, permiten la modelación de la variables μ_i a través de técnicas de efectos fijos, efectos aleatorios, o agrupamiento. Incluso existen otras metodologías que también se exploraran

en este trabajo como la de panel dinámico, en el cual se supone que la variable Y_{it} depende de su propio rezago en la ecuación 1, Baltagi (2005).

La pertinencia de utilizar datos panel en esta investigación, corresponde al control que se tiene sobre el efecto de la heterogeneidad no observable de los individuos, en este caso, la heterogeneidad de cada región NUT2. Dicho efecto, puede corresponder, entre otros, a factores idiosincráticos y culturales de cada región. En otras palabras, la idea de usar un panel es que, en este caso, el éxito deportivo de los clubes sea explicado por los mismos factores independientemente de que se trate de regiones distintas, después de controlar por posibles efectos fijos o aleatorios en cada región.

Es importante tener en cuenta que, Las estimaciones a través de OLS sugieren coeficientes estimados consistentes pero ineficientes. Además, los errores estándar no son exactos, lo cual puede llevar a resultados erróneos que dependerán de la forma exacta del proceso de error. Para tratar los problemas de heterocedasticidad y correlación contemporánea presentes en el modelo y de esta manera obtener estimaciones más precisas, se puede utilizar el método de estimación de Mínimos Cuadrados Generalizados Factibles (MCGF) propuesta por Parks (1967) o un método más reciente propuesto por Beck y Katz (1995), donde se aprovechan características propias de los datos para corregir los errores estándar obtenidos por OLS. Esta metodología se denomina Errores Estándar Robustos para Panel (PCSE).

En su trabajo, Beck y Katz (1995), evalúan la exactitud del modelo PCSE y hacen una comparación con la eficiencia de los estimadores de OLS y MCGF. Los autores demuestran

que corregir los errores estándar a través de PCSE, conlleva a estimaciones más precisas en comparación a las demás alternativas. De acuerdo con los autores, la corrección de los errores estándar es posible debido a que se cuenta con T (número de años en la muestra) repeticiones de la información contenida en la correlación contemporánea de los errores y puede ser utilizada para aproximar la matriz de varianzas y covarianzas de los errores.

En este sentido, con el fin de obtener estimaciones más precisas, en este trabajo se utiliza el método propuesto por Beck y Katz (1995), sustituyendo los errores estándar obtenidos en OLS por los errores corregidos que se obtienen de PCSE.

Por otra parte, es ampliamente aceptado que la UEFA Champions League o Liga de Campeones es la competición mayor demandada en Europa a nivel de clubes, reuniendo los equipos más competitivos de cada país europeo, razón por la cual se utilizará como medidor del éxito de un equipo de fútbol. En el periodo analizado en esta investigación, la competición tiene una estructura común en todos sus años, constando de 5 fases hasta llegar a la final del torneo: Fase de grupos (32 equipos), octavos de final, cuartos de final, semifinal y final.

Bajo este contexto, para medir el éxito que un equipo de fútbol tiene en dicha competencia, se construye un ranking como el propuesto por Mourão (2010), en el cual se muestra la competitividad de un equipo de fútbol de acuerdo con los logros obtenidos en cada etapa del certamen de la siguiente manera: si un equipo no llega a la segunda fase, es decir, a octavos de final, su puntuación será 0 para ese año. Si el equipo llega a octavos de final, su puntuación será 0,25. Si el equipo pasa a cuartos de final, el indicador será

0,75 ($0,25 + 0,50$). Si el equipo juega una semifinal, puntuará 1,75 ($0,25 + 0,50 + 1,00$). Si el equipo llega a la final, su puntuación será de 3,75 ($0,25 + 0,50 + 1,00 + 2,00$). Y por último, si el equipo gana la final, su puntuación será 5,75 ($0,25 + 0,50 + 1,00 + 2,00 + 2,00$).

No obstante, cabe resaltar que el interés de esta investigación es hacer un análisis de competitividad regional, y no uno a nivel de clubes de fútbol, por lo cual se construye una variable, para medir dicha competitividad entre regiones, que será nuestra variable dependiente o nuestra variable a explicar en la aproximación de contraste empírico. En este sentido, a cada región NUT2 se le asignará un puntaje que corresponde con la suma de los puntajes obtenidos por los clubes de fútbol que se encuentran en esta región. Por ejemplo, Lombardía, que es una región NUT2 de Italia, en el año 2003 tendría una puntuación de 7,5, producto de la suma de las puntuaciones obtenidas por dos de sus equipos que compitieron ese año en la UEFA Champions League: Internazionale Milano con puntuación de 1,75 por haber sido eliminado en la semifinal y el AC Milan con puntuación de 5,75 por haber ganado la final en ese mismo año.

La construcción de las variables explicativas o variables independientes para cada región NUT2, se hizo de acuerdo a lo sugerido en la literatura relacionada, en cuyas investigaciones se prueban variables para los casos de selecciones de fútbol nacionales y en algunos casos a nivel de clubes.

B. Efectos Dinámicos.

Un modelo como el planteado en (7), supone que el efecto total que generan los factores socioeconómicos sobre el equipo de la región sucede contemporáneamente, lo cual

puede ser válido solo para ciertos factores y bajo ciertas condiciones. Considere primero un escenario donde, en un año dado se presenta un incremento en el ingreso per cápita de una región. Probablemente, los habitantes de dicha región tendrán un mayor ingreso disponible para gastar en actividades de ocio, como el fútbol, en ese mismo año. En este sentido, es muy factible que el efecto total del ingreso per cápita sobre el desempeño del equipo de esa región sea contemporáneo y por lo tanto, sería capturado por las estimaciones vía OLS.

Considere ahora un escenario en donde, en una región se presenta un aumento en la población en un año determinado. Consideramos que posiblemente los nuevos habitantes no van a generar un efecto inmediato sobre el desempeño deportivo de los equipos de la región. De esta manera, un factor socioeconómico de este tipo no genera un efecto inmediato o contemporáneo, sino que su impacto tendrá un efecto dinámico, y se verá a lo largo del tiempo.

Para nuestro análisis, queremos distinguir entre los dos escenarios posibles, ya que estos tienen implicaciones con respecto a la interpretación de nuestros resultados y para la identificación de los variables (es decir, identificar si son parámetros con efectos dinámicos o no).

Claramente, las estimaciones obtenidas mediante OLS, no pueden capturar los efectos dinámicos del sistema. Con este objetivo, se han desarrollado diferentes metodologías para estimar factiblemente un panel dinámico. Por ejemplo, Arellano y Bond (1991) desarrollan un estimador consistente a través del Método Generalizado de los Momentos

(GMM) para modelar los efectos dinámicos de datos panel. De esta manera, se consideran los rezagos de la variable independiente, como variables explicativas, es decir:

$$(9) \quad Y_{it} = \beta_0 + \sum_{j=1}^p \alpha_j Y_{i,t-j} + X'_{it} \beta + \varepsilon_{it}$$

Donde α_j son los p parámetros adicionales a estimar, mientras que el resto de la especificación es igual que en (7). Note que por construcción, los rezagos de la variable dependiente ($Y_{i,t-j}$), incluida en el lado derecho de la ecuación, está correlacionada con los efectos idiosincráticos no observados (ε_{it}), lo que hace que las estimaciones estándar sean inconsistentes. La metodología de Arellano y Bond (1991) utiliza el estimador de GMM, para estimar $\alpha_1, \dots, \alpha_p, \beta_0, \beta$. Aquí, las condiciones de momentos están formadas a partir de las primeras diferencias de los errores en (9). Además, los rezagos de la variable dependiente, de las variables predeterminadas y de las variables endógenas son usadas como instrumentos.

Sin embargo, trabajos como el de Blundell y Bond (1998) y el de Arellano y Bover (1995), muestran que el estimador de Arellano y Bond (1991), puede no desempeñarse bien si el proceso autoregresivo es demasiado persistente o si la relación entre la varianza del efecto individual no observable (μ_i en la ecuación (8)) y la varianza del error idiosincrático (ε_{it}) es muy grande. En este sentido, con base en dichos trabajos, se propone un estimador que utiliza condiciones de momentos adicionales: *Panel Dynamic Data System* (DPDSYS).

En este trabajo se exploran ambas metodologías, sin embargo, consideramos que el estimador más pertinente es el que incluye momentos adicionales (DPDSYS), ya que, este

estimador está diseñado para bases de datos con más paneles que periodos. Además, este método supone que no hay autocorrelación en el término de error idiosincrático, asunto que será probado en la sección siguiente. Asimismo, para que las condiciones de momentos adicionales sean válidas, se requiere la condición inicial de que los efectos individuales no observados no estén correlacionados con la primera diferencia de la primera observación de la variable dependiente, $E(\mu_i \Delta Y_{i2}) = 0$ para todo i ¹.

A continuación se presentan las variables, que consideramos más relevantes, junto a la causa de su relevancia y el signo que se espera obtener. La tabla 1 presenta las estadísticas descriptivas de los datos utilizados para todas las regiones de la muestra completa 2001-2010.

Producto Interno Bruto per cápita (Log): una región con un mayor nivel de ingreso per cápita, tendrá una mayor capacidad de consumo de bienes, entre ellos el fútbol. Por lo cual, los equipos de fútbol, cuyas oficinas principales se encuentren en regiones con un ingreso per cápita elevado, contarán con una mayor demanda de su fútbol, lo que genera mayores ingresos para el club, que se verá reflejado en mejores contrataciones e instalaciones deportivas, elevando de esta manera su competitividad. El signo esperado de esta variable es positivo.

¹ La prueba de no autocorrelación de las primeras diferencias de los errores, en primer y segundo orden, se puede ver en los anexos.

Tabla 1—ESTADÍSTICAS DESCRIPTIVAS DE LAS REGIONES

	N	mean	sd	min	max
Puntos	470	0.329	1.010	0	7.500
Densidad	467	892.1	2,118	10.20	13,871
Tasa Desempleo	468	7.976	3.825	1.800	27.80
Log Ingreso pc	461	9.516	0.740	6.685	10.42
Log Población	470	14.69	0.890	12.75	16.34
Log Turismo	469	6.822	1.196	4.025	9.570
Number of Reg	47	47	47	47	47

Notas: Estadísticas descriptivas de las variables usadas para cada región NUTS2 de la unión europea. De izquierda a derecha los valores por cada variable corresponden a lo siguiente: número de observaciones, media aritmética, desviación estándar, valor mínimo y valor máximo de las variables en todo el periodo de tiempo. Las variables Ingreso per cápita, Población, y Turismo se toman en logaritmo natural.

Fuente: Cálculos propios con base en datos de regiones de Eurostat.

Densidad de la población: una región con alto nivel de densidad de la población, supone un alto grado de urbanización. Como se ha demostrado en la literatura de la economía del desarrollo, el paso de economías rurales a economías urbanas genera externalidades positivas en el desarrollo y el desempeño de los sectores económicos de una región. Por consiguiente, un equipo de fútbol que ejerza de local en una región con un elevado grado de urbanización, tendrá la posibilidad de atraer negocios y empleos, y a su vez incrementar la productividad, junto a todos los beneficios que trae la urbanización en una economía capitalista. Se espera un signo positivo.

Número de establecimientos que ofrecen alojamiento (Log): esta variable es comúnmente utilizada en la literatura para determinar los factores que influyen en el éxito deportivo de las selecciones de fútbol. En nuestro caso, se incluye en el modelo por

tratarse de una competencia internacional, como lo es la UEFA Champions League. En este sentido, una región con un mayor número de establecimientos que ofrezca servicios a los turistas, será más atractiva para los extranjeros, lo cual dinamizará la economía local. Además, será una región llamativa para organizar eventos deportivos internacionales, generando así, un efecto positivo sobre la competitividad de los equipos locales de dicha región. Se espera un signo positivo.

Tasa de desempleo: una alta tasa de desempleo evidentemente lleva a que una gran parte de la población tenga bajos ingresos. De esta manera, los individuos de esta sociedad no podrán destinar ingresos a otros bienes no necesarios, entre los cuales se encuentra el fútbol. Lo anterior, genera una baja demanda por el fútbol de los equipos que alberga una región con estas características, disminuyendo sus ingresos y produciendo una baja competitividad en comparación con equipos cuyas regiones tienen una tasa de desempleo menor. Se espera un signo negativo.

Población a nivel regional (Log): Un equipo de fútbol profesional que tenga su oficina principal en una región más poblada, tendrá un mercado más amplio y, por lo tanto, la posibilidad de tener una mayor cantidad de demandantes de fútbol. El signo esperado es positivo.

C. Datos

Se tienen en cuenta 47 regiones europeas durante el periodo 2001-2010. Dichas regiones se toman del segundo nivel de las demarcaciones territoriales utilizadas por la Unión Europea con fines estadísticos: NUTS2 (Nomenclatura de las Unidades Territoriales

Estadísticas), creadas por EuroStat, la cual es la oficina europea de estadística y nuestra principal fuente de información. En este sentido, las NUT2, y por ende esta investigación, toma en cuenta las regiones cuya población esté entre 0,8 y 3 millones de habitantes.

Se toman 47 de las regiones que albergan la sede principal de algún equipo participante en la UEFA Champions League en el periodo 2001-2010. Dicho periodo se toma por dos motivos: primero, disponibilidad de datos socioeconómicos para todas las regiones NUT2; y segundo, es un periodo en donde se tiene una estructura común en la competición en cuestión, partiendo de una fase de grupos hasta llegar a la final, lo cual facilita el análisis.

Como se puede observar los datos siguen una estructura de panel ya que contienen la información de 47 regiones NUT2, en un periodo de 10 años. Por tal motivo, en este trabajo, para contrastar empíricamente la hipótesis de que los factores socioeconómicos influyen sobre el desempeño deportivo de los equipos de fútbol europeo, se utiliza la metodología de Datos Panel.

D. Resultados

Para obtener las correlaciones básicas que se quieren estudiar, iniciamos estimando la ecuación (7) por OLS, agrupando todos los años de datos en la muestra. Los resultados se muestran en la Tabla 2. En la columna 1, se hace la estimación sin controlar por ningún

Tabla 2— DESARROLLO REGIONAL Y DESEMPEÑO DEPORTIVO
DE EQUIPOS DE FÚTBOL EN OLS

	(1)	(2)	(3)
Log Ingreso pc	-0.00924 (0.0770)	0.0239 (0.131)	0.221 (0.281)
Tasa Desempleo	-0.0221* (0.0121)	-0.00340 (0.0157)	0.0143 (0.0203)
Log Población	0.252*** (0.0835)	0.280* (0.148)	-0.247 (0.957)
Densidad	0.000109*** (2.41e-05)	9.89e-05** (4.52e-05)	0.000648 (0.000511)
Log Turismo	0.0821 (0.0678)	0.0312 (0.120)	-0.470 (0.306)
Constant	-3.776*** (1.240)	-4.294** (2.167)	4.388 (13.59)
Observations	455	455	455
R-squared	0.139		0.010
Number of Reg		47	47
Reg FE			Y

Notas: Columna 1—3: Regresiones OLS, OLS-RE y OLS-FE, respectivamente. Variable dependiente: puntos logrados por cada región ($Puntos_{it}$). Errores estándar en paréntesis.

*** Significante al 1 por ciento.

** Significante al 5 por ciento.

* Significante al 10 por ciento.

efecto idiosincrático, y se encuentra que el ingreso per cápita de las regiones europeas no afecta el desempeño de los equipos de fútbol de la región. Por su parte, las demás variables tienen un efecto significativo y con un signo esperado sobre el nivel de éxito deportivo. Se encuentra que una región con una mayor tasa de desempleo generará un efecto negativo sobre el éxito de los equipos de la región, mientras que un mayor número de habitantes, en una región más urbanizada y con mayor número de establecimientos que ofrecen alojamiento generan un efecto positivo sobre los puntos alcanzados en la UEFA Champions League por los equipos de esa región.

Siendo conscientes de las limitaciones de estimar por medio de OLS agrupados, y con el fin de obtener un mayor acercamiento a la realidad, en donde se controle por los efectos idiosincráticos de cada región europea, se hace la estimación suponiendo que cada unidad transversal tiene un intercepto diferente. En este sentido, se considera primero un escenario, en donde dichos efectos idiosincráticos son aleatorios entre unidades. Los resultados de esta especificación, la cual se estima a través de OLS-RE, se muestran en la columna 2². Una vez más se encuentra evidencia del efecto positivo (aunque esta vez más débil) entre el tamaño de la población, el nivel de urbanización y el nivel de turismo de la región con el desempeño deportivo de los equipos. Por su parte, la tasa de desempleo sigue teniendo un efecto negativo sobre los puntos obtenidos por los equipos, aunque el efecto es no significativo. Bajo esta especificación el ingreso per cápita tiene un efecto positivo pero nuevamente no significativo.

Aunque nuestra especificación preferida es donde se tienen en cuenta efectos aleatorios entre regiones, también se explora un escenario en donde el carácter individual de las regiones no es aleatorio, sino constante, a través de un modelo con efectos fijos. Los resultados se muestran en la columna 3³. Se encuentra que si se restringe el modelo, a un escenario en donde todas las unidades transversales tienen el mismo intercepto, los

² La prueba del Multiplicador de Lagrange para Efectos Aleatorios formulada por Breusch y Pagan, indica que existen diferencias relevantes entre la estimación a través de OLS agrupados y con efectos fijos, por lo que la varianza del término individual es diferente de cero lo que muestra la pertinencia de usar Efectos Aleatorios. Prueba disponible en los anexos.

³ Los Cuadros 2 y 3 en los anexos muestran que: primero, se confirma nuevamente que hay diferencias relevantes al incluir el término de efectos individuales. Y segundo, el test de Hausman muestra que los estimadores RE y FE no difieren sistemáticamente, por lo que es preferible usar Efectos Aleatorios.

efectos de los factores socioeconómicos sobre el desempeño deportivo de los equipos son nulos.

Los resultados hallados anteriormente podrían sugerir que asumir un modelo en el cual el carácter individual de las regiones es constante, así como suponer que la varianza de las perturbaciones es constante entre los individuos y los años resulta ser un supuesto bastante restrictivo. Además, se puede esperar que en un modelo con datos panel se viole el supuesto de independencia de los errores, ya que diferentes unidades transversales o sus errores puede estar correlacionados temporalmente. En este sentido, es necesario realizar las pruebas pertinentes para comprobar la presencia de perturbaciones no esféricas en el modelo⁴.

Una vez se realizan las pruebas correspondientes, se puede observar que el modelo presenta problemas de heterocedasticidad y de correlación contemporánea, ya que se rechazan las hipótesis nulas de la prueba modificada de Wald y de la prueba de Breusch y Pagan, respectivamente. Sin embargo, el test propuesto por Wooldridge, para detectar problemas de autocorrelación de primer orden, sugiere que el modelo planteado no presenta problemas de este tipo.

Como se discutió en la metodología, nuestros resultados a través de OLS sugieren coeficientes consistentes pero ineficientes, por lo que se utiliza el método propuesto por Beck y Katz (1995). Los resultados encontrados mediante la corrección PCSE se muestran en la tabla 3. Estos hallazgos son consistentes con los resultados de la estimación OLS. Las

⁴ Las pruebas de heterocedasticidad, correlación contemporánea y autocorrelación pueden verse en los anexos.

Tabla 3—DESARROLLO REGIONAL Y DESEMPEÑO DEPORTIVO
DE EQUIPOS DE FÚTBOL EN PCSE

	(1)
Log Ingreso pc	-0.00924 (0.0334)
Tasa Desempleo	-0.0221* (0.0113)
Log Población	0.252*** (0.0329)
Densidad	0.000109*** (3.19e-05)
Log Turismo	0.0821** (0.0416)
Constant	-3.776*** (0.582)
Observations	455
Number of Reg	47
R-squared	0.139

Notas: Variable dependiente: puntos logrados por cada región ($Puntos_{it}$). Estimación mediante Panel Corrected Standard Error (PCSE). Errores estándar corregidos en paréntesis.

*** Significante al 1 por ciento.

** Significante al 5 por ciento.

* Significante al 10 por ciento.

diferencias fundamentales se encuentran, como era de esperarse, en los errores estándar, los cuales son ahora más pequeños que los obtenidos mediante OLS. De esta manera, unos errores estándar más pequeños hacen, por ejemplo, que el efecto del nivel de turismo (medido por el número de establecimientos que ofrecen alojamiento) sea más significativo para explicar el desempeño deportivo de los equipos de la región. Por su parte, el efecto negativo de la tasa de desempleo se mantiene, así como los efectos del tamaño de la población, el nivel de urbanización y el nivel de turismo, los cuales generan un efecto positivo y significativo en los puntos logrados por una región. Sin embargo, el

modelo indica que el ingreso per cápita no genera un efecto significativo sobre el desempeño de los equipos de fútbol. Tanto las estimaciones obtenidas por OLS en la tabla 2, como las obtenidas mediante PCSE en la tabla 3, son consistentes con los signos esperados y la significancia de las variables explicativas consideradas en el modelo, lo cual consideramos un resultado valioso para el propósito de explicar los puntos logrados por los equipos de una región.

Sin embargo, con el objetivo de lograr un mayor grado de profundización en la explicación de este fenómeno socioeconómico, se exploran metodologías alternativas, con el fin de encontrar resultados que se acerquen cada vez más a la realidad. En ese sentido, se emplea la metodología de panel dinámico expuesta en la sección anterior, con el objetivo de distinguir los escenarios posibles e identificar los posibles efectos dinámicos de las variables.

La tabla 4 muestra los resultados del modelo planteado en (9) estimado mediante la metodología DPDSYS⁵. Como punto de partida se inicia con la columna 1, en donde se incluye, además de los factores socioeconómicos anteriores, los rezagos de 2 periodos de la variable dependiente, es decir, los rezagos de los puntos logrados por cada región. Bajo esta especificación, se observa que de todas las variables explicativas, únicamente los puntos logrados en el periodo inmediatamente anterior, es estadísticamente significativa.

Este resultado puede indicar dos cosas: primero, que existe un efecto persistente de los puntos, es decir, en alguna medida el desempeño deportivo del periodo anterior ayuda a

⁵ También se realiza la estimación mediante el estimado desarrollado por Arellano y Bond. Resultados disponibles en los anexos.

Tabla 4—DESARROLLO REGIONAL Y DESEMPEÑO DEPORTIVO DE EQUIPOS DE FÚTBOL EN XTPDPSYS

	(1)	(2)	(3)
Puntos _{t-1}	-0.313** (0.137)	-0.362*** (0.0555)	-0.362** (0.143)
Puntos _{t-2}	-0.127 (0.0820)	-0.157*** (0.0505)	-0.157** (0.0783)
Log Ingreso pc	1.686 (1.320)	-0.00777 (0.783)	-0.00777 (0.375)
Log Ingreso pc _{t-1}		1.889** (0.958)	1.889* (1.137)
Log Ingreso pc _{t-2}		-0.525 (0.759)	-0.525 (0.650)
Tasa Desempleo	0.0678 (0.0579)	0.0349 (0.0301)	0.0349 (0.0436)
Tasa Desempleo _{t-1}		0.0289 (0.0401)	0.0289 (0.0276)
Tasa Desempleo _{t-2}		-0.0621 (0.0431)	-0.0621* (0.0334)
Log Población	1.116 (2.057)	2.338 (15.07)	2.338 (10.86)
Log Población _{t-1}		2.594 (15.34)	2.594 (11.63)
Log Población _{t-2}		-1.094 (1.231)	-1.094 (1.097)
Densidad	0.000288 (0.000413)	0.00657 (0.00668)	0.00657 (0.00708)
Densidad _{t-1}		-0.0220 (0.0139)	-0.0220** (0.0111)
Densidad _{t-2}		0.0158* (0.00830)	0.0158*** (0.00433)
Log turismo	-0.576 (0.596)	-0.243 (0.492)	-0.243 (0.608)
Log Turismo _{t-1}		0.723* (0.410)	0.723* (0.376)
Log Turismo _{t-2}		-2.338*** (0.497)	-2.338** (0.908)
Constant	-28.91 (29.19)	-56.65*** (12.22)	-56.65** (27.54)
Observations	370	358	358
Number of Reg	47	47	47

Notas: Variable dependiente: puntos logrados por cada región (Puntos_{it}). Estimaciones mediante *Dynamic Panel Data System* (DPDPSYS). Columna 1: sólo variable dependiente rezagada. Columna 2: variable Dependiente e independientes rezagadas. Columna 3: variable dependiente e independientes rezagadas con errores estándar robustos. Errores estándar en Paréntesis.

*** Significante al 1 por ciento.

** Significante al 5 por ciento.

* Significante al 10 por ciento.

explicar el desempeño deportivo que tiene la región en el periodo siguiente. Segundo, que existe un problema de endogeneidad en el modelo por variables omitidas. Se busca probar empíricamente estos dos primeros resultados en la columna 2. Ahora, en el lado derecho de la ecuación se incluye, además de lo anterior, los rezagos de los factores socioeconómicos en dos periodos. Se puede observar ahora, a diferencia de los modelos anteriores, que el ingreso per cápita tiene un efecto significativo sobre el desempeño deportivo de los equipos de una región. Dicho efecto, lo genera el nivel del ingreso per cápita del año anterior sobre los puntos logrados por la región en el periodo t . Es decir, un aumento en los niveles de ingreso de las personas este año, conllevaría a mejores resultados deportivos el año siguiente. Así mismo, mayores niveles de urbanización en una región, tendrá efectos sobre los equipos de fútbol con un rezago de dos años. Por su parte, el número de establecimientos que ofrecen alojamiento nocturno, genera efectos significativos tanto en el primer rezago como en el segundo (aunque se observa que estos efectos son contrarios) pero no lo hace contemporáneamente. El resto de las variables bajo esta especificación no resultan estadísticamente significativas, salvo los dos rezagos de la variable dependiente.

Como se discutió en la sección anterior, el modelo presenta problemas de heterocedasticidad, lo que hace que los estimadores de la varianza y los errores estándar no sean consistentes ni insesgados. Por lo tanto, se considera una especificación con errores estándar robustos. Los resultados se muestran en la columna 3. De esta manera, consideramos esta especificación como la que en mayor grado explica el comportamiento

de los puntos obtenidos por cada región, en cuanto tiene en cuenta los posibles efectos dinámicos y considera errores estándar robustos.

Se encuentra que el efecto del ingreso per cápita sobre el desempeño deportivo de los equipos de la región en el periodo siguiente se mantiene. Por su parte, la tasa de desempleo de en el año t , generará un efecto negativo sobre los puntos obtenidos con un rezago de dos años. Esto indicaría, por ejemplo, que según el ranking de puntos propuesto en este trabajo, si la tasa de desempleo aumenta en 5% este año en una región determinada, y todo lo demás se mantiene constante, podría significar a los equipos de esa región no pasar de la fase de grupos de UEFA Champions League en ediciones posteriores. Se observa además que los efectos del nivel de urbanización y de oferta de alojamiento se mantienen, siendo ambos estadísticamente significativos en su primer y segundo rezago.

En este punto, se puede notar que para algunas variables los efectos en cada periodo pueden ser ambiguos y presentar signos contrarios en sus rezagos. Lo anterior no permite identificar un efecto total de un factor sobre el desempeño futbolístico de la región. Para tratar con este problema, se propone establecer una combinación lineal entre los 3 coeficientes obtenidos para cada variable. Es decir, una combinación lineal del efecto contemporáneo y sus dos rezagos.

Los resultados de la combinación lineal se presentan en la tabla 5. Se observa que los factores principales del modelo teórico, el ingreso y la población, tienen efectos totales significativos sobre el desempeño deportivo de los equipos de fútbol de las regiones euro-

Tabla 5—DESARROLLO REGIONAL Y DESEMPEÑO DEPORTIVO DE EQUIPOS DE FÚTBOL-
COMBINACIÓN LINEAL DE LOS PARÁMETROS

Puntos	Coef.	Std. Err.	Z	P> z	[95% Conf. Interval]
Log Ingreso pc	1.356898	.7253259	1.87	0.061	-.0647151 2.77851
Tasa Desempleo	.0017267	.0601633	0.03	0.977	-.1161912 .1196446
Log Población	3.837546	1.845651	2.08	0.038	.2201366 7.454955
Densidad	.0003857	.0005096	0.76	0.449	-.0006131 .0013845
Log Turismo	-1.858138	.8267459	-2.25	0.025	-3.47853 -.2377453

Notas: Combinación lineal de los coeficientes obtenidos, teniendo en cuenta el coeficiente contemporáneo y los dos rezagos de cada variable.

peas. Esto es, una región con un mayor ingreso per cápita acompañado de un mayor número de habitantes, tiene equipos de fútbol con mayor éxito en las competencias internacionales, tales como la Liga de Campeones de la UEFA. Además, la variable que mide el nivel de turismo de una región resulta significativa, pero con un efecto total contrario al esperado. Por su parte, aunque algunos rezagos de la densidad y la tasa de desempleo tienen un efecto significativo, como se encontró antes, cuando se mira el efecto total, resultan ser estadísticamente no significativas.

V. Conclusiones

En este trabajo, se elaboró y se contrastó empíricamente el enlace que existe entre los factores socioeconómicos y el desempeño deportivo de los equipos de fútbol en las regiones europeas. Se exploran diferentes especificaciones del problema con el fin de capturar los efectos idiosincráticos de cada región y además capturar los posibles efectos dinámicos que puedan existir en el sistema. En el modelo, las regiones con mayor población y mayor ingreso tienen equipos de fútbol más competitivos por diferentes razones. Primero, generar deportistas más competitivos requiere gastos considerables y

condiciones regionales óptimas; además sostener estos deportistas requiere pagar salarios más altos, por lo que los recursos económicos son importantes para explicar el éxito deportivo de los equipos. Segundo, una mayor población aumenta la probabilidad de tener deportistas talentosos, ya que, se considera que el talento se distribuye aleatoriamente entre las regiones. Finalmente, otros factores socioeconómicos, tales como los niveles de desempleo, de urbanización, de turismo, entre otros, pueden también afectar los puntos que un equipo de determinada región obtenga en competencias internacionales, tales como la UEFA Champions League.

La evidencia empírica proviene de un panel de regiones europeas con información sobre el desempeño deportivo de los equipos de fútbol pertenecientes a cada región, medido por los puntos obtenidos en las ediciones de la Liga de Campeones organizada por la UEFA, así como información socioeconómica a nivel de región. Los datos considerados abarcan el periodo 2001-2010, por tratarse de un periodo con una estructura común de la edición de la competencia europea en cuestión.

Se exploraron distintas alternativas, en búsqueda de un modelo donde se tengan en cuenta los problemas económicos y econométricos que un fenómeno socioeconómico de este tipo presenta. En primer lugar se busca un modelo para controlar las heterogeneidades entre las diferentes regiones, al tiempo que se busca corregir el problema de heterocedasticidad presente en el modelo. Luego, se introducen los efectos dinámicos del sistema, y finalmente se hallan los efectos totales de las variables consideradas sobre el éxito deportivo de los equipos de cada región.

A lo largo del trabajo, los modelos empíricos sugieren consistentemente que los equipos de fútbol pertenecientes a regiones con mayor ingreso per cápita y mayor número de habitantes obtendrán mejores resultados deportivos. Por su parte, aunque los niveles de urbanización, de desempleo y de turismo generan efectos significativos con algunos periodos de rezago, cuando se mira el efecto total de cada variable, se encuentra que no tienen un efecto total estadísticamente significativo.

La contribución de este trabajo radica en avanzar hacia la construcción e identificación (empírica y teóricamente) de los mecanismos que explican por qué a algunos equipos de fútbol les va mejor que a otros. Los resultados ayudan a comprender mejor la naturaleza de este fenómeno, clarificando la idea que el desempeño de los equipos tiene fundamentos socioeconómicos que pueden ser estudiados más cuidadosamente y no son producto del juego nada más.

Por último y no menos importante, este trabajo contribuye a la exploración de métodos y herramientas que ayuden a enlazar la teoría económica con el fútbol y el deporte en general, para avanzar en un campo en el que hay mucho por aportar y analizar.

VI. REFERENCIAS

- Andreff, M., Andreff, W. y Poupaux, S.** 2008. "Determinantes Económicos de Actuación Olímpica: La Predicción de Medallas Que Se Ganó En Los Juegos de Beijing." *Journal of Political Economy* 2/2008 (Vol 118), p. 135-169 URL: -www.cairn.info/revue-de-política-economía-2008-2-page-135.htm.
- Arellano, M., and S. Bond.** 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *Review of Economic Studies* 58: 277–297.
- Arellano, M., and O. Bover.** 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics* 68: 29–51.
- Baltagi, B.** 2005. "Econometric Analysis of Panel Data." *Wiley, Chichester*.
- Beck, N., y Katz, J.** 1995. "What to Do (and not do) With Time-Seriescross-Section Data." *American Political Science Review*, 89(3), 634-647
- Bernard, A., y Busse, M.** 2004. "Who Wins The Olympic Games: Economic Resources and Medal Totals." *The Review of Economics and Statistics*, 86(1): 413–417.
- Blundell, R., and S. Bond.** 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87: 115–143
- Downward, P., y Dawson, A.** 1999. "The Demand For Professional Team Sports: Traditional Finings and New Developments." *Business School Working Paper*, 997
- Hawksworth, J.** 2008. "Economic Briefing Paper: Modelling Olympic Performance." *PriceWaterHouse Coopers Report*.
- Hoffmann, R., Lee, C. y Ramasamy, B.** 2002. "The Socio-Economic Determinants of International Soccer Performance." *Journal of Applied Economics*, 2, 253-272
- Kuper, S. y Szymanski, S.** 2009. "Soccernomics." *Boston, Massachusetts, USA, Nation Books*.
- Leeds, M. y Leeds, E.** 2009. "International Soccer Success and National Institutions." *Journal of Sports Economics*, 10, 369-390.
- Lewis, M.** 2004. "Moneyball." *Nueva York, W. W. Norton & Company Ltd., 2004*.
- Montero, R.** 2010. "Panel dinámico" *Documentos de Trabajo en Economía Aplicada*, Universidad de Granada. España
- Mourão, P.** 2008. "Desarrollo Local y Competitividad Lugar Los Equipos de Fútbol. El Caso Portugués." *Investigaciones Regionales, mes Pecado*, 135-143

- Mourão, P.** 2010. "Regional Determinants of Competitiveness: The Case of European Soccer Teams." *International Journal of Sport Finance*, 2010, 5, 222-234. 2010.
- Parks, R.** 1967. "Efficient Estimation of a System of Regression Equations When Disturbances are Both Serially and Contemporaneously Correlated." *Journal of the American Statistical Association*, 62(38), 500-509.
- Rathke, A. y Woitek, U.** 2007. "Economics and Olympics: An Efficiency Analysis." *IEW–Working Papers iewwp313*.

VII. ANEXOS

ANEXO 1. Comparación entre modelos.

Datos Agrupados vs Efectos Aleatorios. Si se observa que la varianza de los términos que capturan los efectos individuales son iguales a cero, no habrían diferencias relevantes entre OLS agrupados y Efectos Fijos o Aleatorios. A continuación se realiza la prueba del Multiplicador de Lagrange para Efectos Aleatorios formulada por Breusch y Pagan, para establecer si existen diferencias entre los modelos de datos agrupados y el de efectos fijos. La hipótesis nula de esta prueba establece que la varianza de los términos de efectos fijos es igual a cero, $Var(u) = 0$.

CUADRO 1. Multiplicador de Lagrange para Efectos Aleatorios		
$H_0 : Var(u) = 0$	chibar2(01) = 147.25	Prob > chibar2 = 0.0000

De acuerdo con los resultados de la prueba, se rechaza la hipótesis nula, lo cual indica que los efectos aleatorios u_i son relevantes y es preferible utilizar un modelo donde se tengan en cuenta.

Datos Agrupados vs Efectos Fijos. El modelo de datos agrupados asume un intercepto común para todas las unidades de corte transversas, es decir, no incluye variables dicotómicas para cada región. Por su parte, el modelo de efectos fijos, si tiene en cuenta un término individual para capturar efectos idiosincráticos. De esta manera, para probar la pertinencia del uno o del otro, se hace una prueba F de significancia para todos los términos de efectos fijos. En este sentido, la hipótesis nula establece que todos los u_i son iguales a cero. Por lo tanto, si la hipótesis nula es rechazada, significaría que, por lo menos, algunas variables de efectos fijos pertenecen al modelo.

CUADRO 2. Prueba F de significancia.		
$H_0 : all u_i = 0$	F(46, 403) = 4.78	Prob > F = 0.0000

Se observa que se rechaza la hipótesis nula, indicando que existen efectos fijos relevantes en el modelo, por lo que es pertinente emplear efectos fijos en comparación al modelo de datos agrupados.

Efectos Fijos vs Efectos Aleatorios. Las pruebas anteriores establecen que en el modelo existen efectos de carácter individual significativos que son pertinentes modelar. Se observa que tanto el modelo de FE y RE caracterizan mejor los datos en comparación con OLS-Pooled. Determinar cuál modelo se desempeña mejor con los datos, depende de la correlación entre el término de carácter individual u_i , y las variables explicativas X'_{it} . OLS-RE asume que no existe tal correlación, mientras que OLS-FE si la tiene en cuenta. En este sentido, Hausman desarrolla un test, donde muestra que la diferencia entre los coeficientes de ambos modelos, puede ser usada para probar la correlación en cuestión. El test de Hausman establece la hipótesis nula de que no hay una diferencia sistemática en los coeficientes FE y RE. Si se rechaza, indica que los estimadores si difieren, lo que indica que existe correlación entre u_i y X'_{it} , por lo que Efectos Fijos sería más conveniente.

CUADRO 3. Test de Hausman		
H_0 : Diferencia no Sistemática en Coeficientes	chi2(5) = 5.81	Prob>chi2 = 0.3249

Los resultados muestran que no es posible rechazar la hipótesis nula del test, por lo que $E(u_i X'_{it}) = 0$. De esta manera, es preferible utilizar OLS-RE sobre OLS-FE.

ANEXO 2. *Perturbaciones Esféricas.*

Autocorrelación. Es importante comprobar la independencia de los términos de error del modelo, debido a que utilizar OLS bajo términos de error no independientes, conlleva a no

obtener estimadores MELI (Mejores Estimadores Lineales Insesgados). A menudo, en datos panel, el supuesto de independencia de los errores es violado, ya sea por la presencia de correlación contemporánea, correlación serial o ambas. Aquí, se busca comprobar si los errores de cada unidad transversal están correlacionados serialmente. Por flexibilidad de la prueba, se utiliza la desarrollada por Wooldridge, cuya hipótesis nula plantea que no hay correlación de primer orden, es decir, no existe autocorrelación.

CUADRO 4. Test de Wooldridge para Autocorrelación		
H_0 : no first-order autocorrelation	F(1, 46) = 1.110	Prob > F = 0.2976

La prueba indica que el modelo no presenta problemas de autocorrelación de primer orden en los errores, ya que no se rechaza la hipótesis nula.

Heterocedasticidad. Asumir que la varianza de los términos de error son constantes, resulta ser un supuesto bastante fuerte, incluso, cuando se trabaja con datos panel. Ignorar los problemas de heterocedasticidad en el modelo, llevaría a violar el supuesto de perturbaciones esféricas y conduciría a estimaciones ineficientes. Para verificar este supuesto existen diferentes pruebas, así que por flexibilidad, utilizamos la prueba modificada de Wald para heterocedasticidad, ya que esta no exige normalidad en los errores. En esta prueba la hipótesis nula establece que la varianza de los errores es constante para todo i , $\sigma_i^2 = \sigma^2, \forall i$.

CUADRO 5. Prueba Modificada de Wald para heterocedasticidad.		
$H_0 : \sigma_i^2 = \sigma^2, \forall i.$	chi2 (47) = 1.5e+08	Prob>chi2 = 0.0000

De esta manera, se puede observar que se rechaza la hipótesis nula, así que se determina que el modelo presenta problemas de heterocedasticidad.

Correlación Contemporánea. Al estudiar un fenómeno por regiones, como es el caso de este trabajo, pueden haber choques no observados que afecten a dos o más regiones en el mismo periodo de tiempo. Por ejemplo, pueden existir factores climáticos que afecten el desempeño deportivo de los equipos de fútbol, los cuales son observados por los individuos, pero no por el econometrista. En este caso, se estaría tratando con un problema de correlación contemporánea, el cual es importante tener en cuenta a la hora de estimar. Para detectar este tipo de problema se utiliza la prueba de Breusch y Pagan, donde se testea la hipótesis nula de independencia entre las unidades transversales, es decir, que no hay correlación contemporánea en el modelo.

CUADRO 6. Prueba de Breusch y Pagan para Correlación Contemporánea		
$H_0 : \sigma(i)^2 = \text{Independencia Serial}$	$\chi^2(1081) = 1557.972$	Pr = 0.0000

Los resultados de la prueba sugieren que el modelo presenta problemas de correlación contemporánea, ya que se rechaza la hipótesis nula del test.

Anexo 3. Pruebas de la post estimación a través de DPDSYS.

Autocorrelación de Primer y segundo Orden. Para utilizar un modelo de panel dinámico es pertinente que exista correlación serial de primer orden en las primeras diferencias de los residuales, pero es importante que no exista tal correlación en segundo orden. Para verificar este ítem, existe una prueba post estimación desarrollada por Arellano y Bond,

donde la hipótesis nula es que no existe autocorrelación. Los resultados de la prueba, para el modelo donde solo se incluye el rezago de la variable dependiente (columna 1 de la tabla 4), se muestran en el cuadro 7, en donde no se rechaza la no autocorrelación de primer orden, pero se rechaza la presencia de autocorrelación de segundo orden en el modelo.

**Cuadro 7. Prueba de Arellano y Bond
Para autocorrelación.**

Order	z	Prob > z
1	-2.4544	0.0141
2	-1.6413	0.1007

H0: no autocorrelation

Sobreidentificación. Se hace además, las pruebas de sobreidentificación del modelo a través del Test de Sargan. Es importante resaltar que, para este tipo de modelos, es conveniente que las ecuaciones estén sobreidentificadas. En este sentido, el test de Sargan prueba si la sobreidentificación del modelo es correcta. El cuadro 8, muestra los resultados de la prueba para el modelo que incluye tanto la variable dependiente, como las independientes rezagadas (columna 2 de la tabla 4).

CUADRO 8. Test de Sargan		
H ₀ : overidentifying restrictions are valid	chi2(41) = 223.6433	Prob > chi2 = 0.0000

De acuerdo con los resultados de la prueba, se rechaza la hipótesis nula de que el modelo dinámico esté correctamente sobreidentificado. De acuerdo con Montero (2010), pueden estar ocurriendo dos cosas: Primero, la más evidente, que el modelo no esté suficientemente sobreidentificado, por lo que la posible solución es estimar el modelo en

dos etapas. Segundo, que el modelo tiene problemas de heterocedasticidad, por lo que se desconoce la distribución de del test de Sargan y por lo tanto existe una tendencia a rechazar la hipótesis nula, Arellano y Bond (1991).

Como se mostró anteriormente, las pruebas indican la presencia de heterocedasticidad en el modelo, por lo que la solución es hacer la estimación con errores estándar robustos, (Columna 3 de la tabla 4). Para este último modelo, nuevamente se hace la prueba post estimación de Arellano y Bond para comprobar la no autocorrelación de las primeras diferencias de las perturbaciones. Los resultados se muestran en el cuadro 8.

**Cuadro 8. Prueba de Arellano y Bond
Para Autocorrelación**

Order	z	Prob > z
1	-2.5951	0.0095
2	-.65531	0.5123

H0: no autocorrelation

Nuevamente, se observa que hay evidencia estadística de autocorrelación de primer orden en las primeras diferencias de los errores, lo cual apoya los mecanismos de los efectos dinámicos. Y nuevamente, se comprueba que no existe tal autocorrelación en segundo orden.

Anexo 4. Estimación y Pruebas Post Estimación a través del estimador de Arellano y Bond.

Los resultados del modelo estimado mediante el estimador de Arellano y Bond son mostrados en la tabla 6.

Tabla 6—DESARROLLO REGIONAL Y
DESEMPEÑO DEPORTIVO DE EQUIPOS DE FÚTBOL EN XATABOND

	(1)	(2)	(3)
Puntos _{t-1}	-0.396*** (0.148)	-0.437*** (0.0599)	-0.437*** (0.149)
Puntos _{t-1}	-0.148 (0.100)	-0.180*** (0.0540)	-0.180* (0.0977)
Log Ingreso Pc	0.394 (0.681)	-0.343 (0.758)	-0.343 (0.456)
Log Ingreso Pc _{t-1}		1.342 (0.932)	1.342 (0.856)
Log Ingreso Pc _{t-2}		-0.386 (0.766)	-0.386 (0.506)
Tasa Desempleo	0.0253 (0.0296)	0.00391 (0.0328)	0.00391 (0.0303)
Tasa Desempleo _{t-1}		0.0246 (0.0391)	0.0246 (0.0318)
Tasa Desempleo _{t-2}		-0.0581 (0.0427)	-0.0581 (0.0390)
Log Población	4.513 (4.651)	-6.835 (19.31)	-6.835 (18.55)
Log Población _{t-1}		14.71 (18.63)	14.71 (20.72)
Log Población _{t-2}		-1.064 (1.200)	-1.064 (1.041)
Densidad	-0.000920 (0.000576)	0.0170* (0.00898)	0.0170 (0.0147)
Densidad _{t-1}		-0.0301** (0.0139)	-0.0301* (0.0159)
Densidad _{t-2}		0.0102 (0.00935)	0.0102** (0.00494)
Log Turismo	-0.409 (0.798)	-0.0908 (0.489)	-0.0908 (0.599)
Log Turismo _{t-1}		0.658 (0.402)	0.658 (0.408)
Log Turismo _{t-2}		-2.330*** (0.486)	-2.330*** (0.886)
Constant	-66.26 (69.50)	-90.87*** (33.55)	-90.87 (79.97)
Observations	322	311	311
Number of Reg	47	47	47

Notas: Variable dependiente: puntos logrados por cada región (Puntos_{it}). Estimaciones mediante Estimator de Arellano y Bond (ABOND). Columna 1: sólo variable dependiente rezagada.

Columna 2: variable Dependiente e independientes rezagadas. Columna 3: variable dependiente e independientes rezagadas con errores estándar robustos. Errores estándar en Paréntesis.

*** Significante al 1 por ciento.

** Significante al 5 por ciento.

* Significante al 10 por ciento

Se prueba para esta estimación la autocorrelación de las primeras diferencias de los errores para el modelo con errores estándar (cuadro 9) y para el modelo con errores estándar robustos (cuadro 10)

Cuadro 9. Prueba de Arellano y Bond para Autocorrelación

Order	z	Prob > z
1	-2.3255	0.0200
2	-1.5554	0.1198

H0: no autocorrelation

Cuadro 10. Prueba de Arellano y Bond para Autocorrelación

Order	z	Prob > z
1	-2.5244	0.0116
2	-.76695	0.4431

H0: no autocorrelation

En ambos casos se evidencia autocorrelación de primer orden para las primeras diferencias de los errores, mientras que se rechaza la autocorrelación de segundo orden. Además se hace el test de Sargan (cuadro 11), el cual evidencia una incorrecta sobreidentificación del modelo, por lo que se explora la estimación por errores estándar robustos en la columna 3 de la tabla 6.

CUADRO 11. Test de Sargan		
H ₀ : overidentifying restrictions are valid	chi2(33) = 148.2832	Prob > chi2 = 0.0000