

Sustituibilidad de la fuerza laboral: El caso de los jugadores de fútbol

Santiago Velásquez *

Asesor: Arlen Guarín

Resumen

Las ligas y equipos de fútbol constituyen un escenario sumamente idóneo para contrastar hipótesis sobre teorías económicas clásicas como la teoría del productor. Aprovechando el conjunto de eventos inesperados, definido por las lesiones de los jugadores durante la competición, es posible identificar cómo estas lesiones afectan la contribución de un jugador en los siguientes partidos y a su vez cómo la ausencia de este factor de producción afecta posteriormente el desempeño y la producción de la firma. Este estudio presenta evidencia sobre el efecto de estos choques para las 5 principales ligas europeas con información desde la temporada 2014-2015 hasta la temporada 2018-2019. Se encontró evidencia de un efecto negativo y significativo en el desempeño individual sin efecto sobre la productividad agregada del equipo en promedio. El efecto individual es robusto en las combinaciones de jugadores y equipos de alta y baja productividad, siendo los jugadores de mayor productividad en equipos de baja productividad los únicos que tienen efecto sobre las variables agregadas del equipo.

Palabras Clave: Event Study, Deportes, Productividad

JEL Codes: D24, L83

*svelas45@eafit.edu.co

1. Introducción

En el estudio de las ciencias económicas se está frecuentemente interesado en elementos tales como la oferta y la demanda de bienes y servicios, en particular, por parte de las firmas, se está interesado en la forma en que estas producen, y cómo interactúan los diferentes factores productivos en esta producción. De acuerdo con Rosen & Sanderson (2001) muchos elementos básicos de la oferta y demanda, pueden claramente verse en los mercados de trabajo de las actividades deportivas y muchos de estos pueden y han sido empíricamente investigados.

La literatura relacionada con la economía del deporte, aunque ha sido estudiada desde hace décadas (Rottenberg 1956, Neale 1964, Sloane 1971), se encuentra aún en desarrollo debido tanto al avance de nuevas técnicas estadísticas, el aumento en la capacidad computacional y el acceso a la información; como a la variedad de teorías aplicadas e hipótesis interesantes que pueden ser validadas en los contextos deportivos, pues, las ligas y equipos, en nuestro caso de fútbol, constituyen un escenario sumamente idóneo para contrastar tales hipótesis sobre teorías económicas clásicas como la teoría del productor, dado que, a diferencia de otros ámbitos en donde no siempre se puede medir directamente la producción y la contribución de los factores a esta, y tampoco es fácilmente diferenciable ni cuantificable, en este contexto sí es posible obtener medidas de desempeño que puedan representar tanto la producción de la firma, medida esta como los resultados de los equipos en sus respectivas ligas, y la contribución de los factores productivos, específicamente la mano de obra, pudiendo ser esto cuantificado con variables de desempeño de los jugadores dentro de cada encuentro. Adicionalmente, el hecho de que dicha producción por parte de la firma y productividad de la mano de obra pueda medirse de forma constante a través de intervalos relativamente cortos de tiempo, permite crear bases de datos de tipo panel lo cual representa una gran ventaja para el uso de las técnicas econométricas.

A grandes rasgos, muchos de los estudios previos explotando datos deportivos se han concentrado en temas como la demanda de trabajo, sustitución de factores, salarios (Ca-

ruso et al. 2017), etc, tal como lo documenta Rosen & Sanderson (2001) en su estudio. Mientras que algunos de los trabajos más recientes han centrado su atención en contrastar hipótesis relacionadas con otras áreas de la economía como lo es la economía conductual Gauriot & Page (2019), Bryson et al. (2020) nuestro trabajo por su parte se enfoca en identificar las contribuciones de los factores productivos a la producción haciendo uso de eventos exógenos que permiten cuantificar el efecto de forma causal.

En el caso particular de esta investigación, se plantea ahondar en la medición de la contribución de los factores productivos, específicamente la mano de obra, a la función de producción de las firmas, mediante datos de partidos de fútbol para las 5 principales ligas europeas: Bundesliga (Alemania), La Liga (España), Ligue One (Francia), Premier League (Inglaterra) y Serie A (Italia), con información partido a partido para cada jugador, desde la temporada 2014-2015 hasta 2018-2019.

Aprovechando el conjunto de eventos inesperados, definido por las lesiones de los jugadores durante la competición, es posible identificar como estas lesiones afectan la contribución de un jugador en los siguientes partidos y como la ausencia de este factor de producción afecta posteriormente el desempeño y la producción de la firma, permitiendo dar respuesta a preguntas como, ¿cuál es el efecto de primer orden de esas lesiones en la participación de los jugadores en el juego? y ¿es posible por parte de la firma compensar/suplir por completo el efecto generado por la ausencia de un jugador?

Las estimaciones de este estudio permiten identificar parámetros de la función de producción que son difíciles de capturar en otros contextos. Entender la sensibilidad del sistema de producción de una firma a los cambios en los factores de producción y contemplar también los posibles cambios en la función de producción según el nivel de productividad del individuo afectado o su equipo entendiéndose como las elasticidades cruzadas entre estos factores. Esto constituye una contribución de interés general, y en particular para los campos de la microeconomía y la teoría organizacional, estando también enfocado a personas de diversas áreas del conocimiento que están interesadas en conocer un poco

más acerca de los hallazgos relacionados con los deportes.

El resto de este trabajo está dividido de la siguiente forma: La sección 2 incluye una revisión del marco teórico de la economía del fútbol y la contribución de la mano de obra a la producción en contextos de equipo, especialmente en el ámbito deportivo. La sección 3 presenta estadísticas descriptivas de los datos utilizados en este estudio, mientras que la sección 4 discute la metodología utilizada. Por último, la sección 5 presenta los resultados y sus efectos heterogéneos para finalmente concluir y describir los siguientes pasos.

2. Marco Teórico

El fútbol es el deporte más jugado alrededor del mundo (Dvorak et al. 2004). Según Efe (2020) para el 2018, las 5 principales ligas europeas tuvieron ingresos entre 1.700 y 5.400 millones de euros representando así alrededor del 75 % del total de ingresos de todas las ligas de los 55 países pertenecientes a la Unión de Asociaciones de Fútbol Europeas (UEFA). A su vez, el desarrollo de esta actividad tiene impactos en la economía de los países, por ejemplo, de acuerdo con Ernst&Young-LLP (2019) durante el 2018 la Premier League generó 3.3 billones en impuestos, contribuyó con 7.6 billones al PIB del Reino Unido y mantuvo alrededor de 12.000 empleos directos y 100.000 indirectos. Lo anterior es una muestra de la importancia tanto social como económica de este deporte en particular alrededor del mundo y a su vez, es una motivación adicional para centrar la atención en este para estudios académicos.

Como en la gran mayoría de deportes de alta competencia, los problemas físicos y las lesiones juegan un papel importante en el rendimiento de los atletas. El efecto de las lesiones sobre el desempeño de los jugadores de fútbol ha sido evaluado previamente por varios estudios académicos centrados principalmente en el área de medicina (Stubbe et al. 2015, Zuke et al. 2018), dichos estudios han encontrado efectos significativos sobre la cantidad de minutos jugados y diferentes variables de desempeño y vuelta a la competición,

a pesar de que esta última tiene una definición muy amplia Zuke et al. (2018). A pesar de estas aproximaciones desde la medicina del deporte, aún se carece de evidencia empírica que estudie este tipo de acontecimientos y sus implicaciones desde un punto de vista económico.

Dada la naturaleza del estudio, es importante dividir las posibles aproximaciones al tema en dos grandes áreas de estudio y su intersección. En primer lugar, el estudio de Rosen & Sanderson (2001) provee un estado del arte al momento de su publicación de los trabajos previos en el área de economía de los deportes y analiza con detalles los problemas económicos que pueden verse evidenciados en este contexto. Tomando en cuenta que “los deportes son un foro excepcional para el análisis económico empírico” (Rosen & Sanderson 2001); se logran documentar estudios sobre la demanda de trabajo, sustitución de factores, salarios, producto marginal del trabajo, discriminación de salarios, oferta de trabajo y calidad de esta, así como diversos temas de organización industrial, negociación y competencia. Asimismo, Scully (1995) señala que pueden asociarse los resultados a nivel de equipo a medidas de desempeño de estos, y a su vez, las contribuciones a estos resultados hacen parte de la función de la producción y pueden verse estimadas como el producto marginal de dicha función. Por otro lado, algunos otros estudios adoptan un contexto empírico más desde el punto de vista de las asociaciones deportivas (ligas) y el negocio del espectáculo que el de los equipos como firmas maximizadoras de beneficio, lo que representa una diferencia significativa respecto a nuestro trabajo.

Un poco más relacionado con nuestra investigación, el estudio de Bonhomme (2020) propone una metodología para estimar la contribución individual en contextos de equipo donde únicamente se observa el resultado agregado, más allá del hecho de que en nuestro caso podamos medir también el desempeño individual, este estudio plantea un detallado marco conceptual del problema teórico con el que estamos tratando.

De acuerdo con Bonhomme (2020) se está interesado en ciertas relaciones principales entre los trabajadores y los equipos. En primer lugar, al igual que Arcidiacono et al. (2017),

se toma en cuenta el hecho de que los trabajadores son heterogéneos en productividad lo cual posteriormente afecta el desempeño total del equipo, por tanto, resulta de suma importancia contar con datos con una estructura de tipo red o empleado-empleador con movilidad entre los empleados, pues de esta forma es posible evidenciar el desempeño de los trabajadores en diferentes equipos logrando medir cuales de estos contribuyen más. Luego, las relaciones de ordenamiento de trabajadores entre equipos y la complementariedad de la mano de obra son también tenidas en cuenta en este marco conceptual y por último, un término de otros factores que recaen en el error del modelo como ruido en la estimación.

A grandes rasgos, Bonhomme (2020) utiliza dos tipos de funciones de producción (producción aditiva y no-lineal) haciendo uso de dos supuestos claves transversales como el supuesto de exogeneidad de la red que establece que el proceso de formación de los equipos es independiente de los choques al equipo condicional en los tipos de trabajador y un supuesto de independencia serial que asume independencia en los choques de los equipos que colaboran repetidamente a través del tiempo. Una ventaja del método de identificación utilizado en nuestro trabajo es que no requerimos de la aplicación de este tipo de supuestos que en el contexto del fútbol pueden ser bastante restrictivos.

Teniendo en cuenta lo anterior, es posible realizar una descomposición de la variación del producto y asignar qué parte corresponde a cada una de las relaciones mencionadas anteriormente. Finalmente, conviene reseñar estudios tales como los de Arcidiacono et al. (2017) y Devereux (2018) que intentan también medir las contribuciones individuales a los resultados de equipo pero sin tener en cuenta algunas innovaciones propuestas en nuestro estudio como el uso de choques exógenos inesperados para cuantificar los efectos de forma causal y el hecho de que no se requieren los supuestos de independencia serial y exogeneidad de la red. Ambos estudios se enmarcan en el contexto deportivo, el primero haciendo uso de datos sobre baloncesto y el segundo sobre tenis. Así, por un lado, el estudio de Arcidiacono et al. (2017) se concentra en estimar la importancia de los jugadores en el desempeño de sus compañeros desarrollando un modelo que permite cuantificar este

efecto mediante un algoritmo para computar los parámetros propuestos en el estudio. Por otra parte, Devereux (2018) encuentra que alrededor del 50 % de la variación en los resultados a nivel de equipo esta explicada por las habilidades de equipo más que habilidades individuales, explotando el hecho de que en el tenis, los jugadores en ocasiones compiten tanto de forma individual como en pareja.

Es importante resaltar que los estudios mencionados en esta última parte han sido desarrollados recientemente y de igual forma, hacen uso de nuevas metodologías y modelos matemáticos sofisticados que se encuentran en la frontera del desarrollo científico. De igual forma, cuentan con datos más recientes siendo esto evidencia sobre la vigencia actual de los resultados. Por tanto, estos motivos mencionados le brindan fuerza a la validez empírica de nuestro ejercicio.

3. Datos

Los datos con los que se cuenta contienen información partido a partido para cada jugador de todos los equipos de las cinco principales ligas de futbol de Europa – Premier League (Inglaterra), La Liga (España), Serie A (Italia), Bundesliga (Alemania) y Ligue One (Francia) – con un periodo de análisis de 5 temporadas, desde la temporada 2014-2015 hasta la temporada 2018-2019, habiendo sido estos extraídos de forma automatizada mediante el uso de métodos de web scraping.

De forma general, se cuenta con información para 136 equipos, 6145 jugadores y 9130 partidos, así como con múltiples variables de productividad en distintos niveles de agregación, tal como el desempeño durante toda la temporada o durante un partido en específico tanto del equipo como cada uno de los jugadores. Para abordar nuestra pregunta empírica, es crucial tener información para todos los jugadores incondicional en si estos venían actuando en los partidos o si por el contrario eran suplentes o no convocados, pues estos últimos constituyen las opciones del equipo para reemplazar a los jugadores habituales en caso de que se vean afectados por lesión. El cuadro 1 contiene estadísticas descriptivas

para las principales variables de interés, en ésta, el panel A contiene información para toda la muestra, mientras que el panel B restringe únicamente a los partidos en los cuales los jugadores tuvieron acción.

Cuadro 1: Estadísticas descriptivas variables de desempeño. Elaboración propia.

	Panel A. Muestra Completa				Panel B. Jugadores Participantes			
	Media	Desv. Estandar	P1	P99	Media	Desv. Estandar	P1	P99
	Variables Individuales							
Asistencias	0.0402	0.2126	0	1.000	0.0825	0.2986	0	1.000
Asistencias Esperadas	0.0363	0.1244	0	0.646	0.0743	0.1700	0	0.808
Disparos	0.4629	1.0444	0	5.000	0.9487	1.3321	0	6.000
Goles	0.0506	0.2493	0	1.000	0.1037	0.3492	0	2.000
Goles Esperados	0.0505	0.1725	0	0.883	0.1035	0.2356	0	1.137
Goles No de Penalty	0.0461	0.2352	0	1.000	0.0944	0.3299	0	1.000
Minutos Jugados	35.283	41.165	0	90.000	72.319	28.194	1	90.000
NPGX	0.0460	0.1547	0	0.789	0.0943	0.2110	0	1.012
Pases Claves	0.3501	0.8424	0	4.000	0.7175	1.0913	0	5.000
Tarjetas Amarillas	0.0733	0.2607	0	1.000	0.1503	0.3574	0	1.000
Tarjetas Rojas	0.0018	0.0420	0	0.000	0.0036	0.0601	0	0.000
XGBuildUp	0.0794	0.2015	0	0.973	0.1627	0.2640	0	1.199
XGChain	0.1344	0.2953	0	1.386	0.2756	0.3739	0	1.685
	Variables de Equipo							
Asistencias Esperadas	0.9226	0.6735	0.046	3.118	0.9479	0.6869	0.050	3.165
Diferencia de Goles	0.0586	1.8770	-5.000	5.000	0.1147	1.8897	-4.000	5.000
Disparos	11.980	5.1836	3.000	27.000	12.1674	5.2170	3.000	27.000
Goles Anotados	1.3986	1.2759	0.000	5.000	1.4343	1.2971	0.000	5.000
Goles en Contra	1.3400	1.2466	0.000	5.000	1.3196	1.2388	0.000	5.000
Goles Esperados	1.2986	0.8804	0.093	4.032	1.3319	0.8958	0.098	4.123
NPGX	1.1895	0.8133	0.090	3.793	1.2211	0.8283	0.095	3.857
Pases Claves	8.9640	4.2630	1.000	21.000	9.1130	4.3009	2.000	21.000
Probabilidad de Derrota	0.3636	0.4810	0.000	1.000	0.3530	0.4779	0.000	1.000
Probabilidad de Empate	0.2487	0.4323	0.000	1.000	0.2463	0.4308	0.000	1.000
Probabilidad de Victoria	0.3877	0.4872	0.000	1.000	0.4008	0.4901	0.000	1.000
XGBuildUp	2.0397	2.1487	0.000	9.993	2.1144	2.2142	0.000	10.381
XGChain	3.4531	3.0573	0.116	14.229	3.5699	3.1445	0.119	14.723

Nota: Panel A, N=688.412 Observaciones. Panel B, N=335.864 Observaciones.

Como se puede evidenciar, se cuenta con métricas de desempeño tradicionales como lo son goles, asistencias, disparos o pases claves. Una de las potenciales restricciones asociadas al fútbol es el hecho de que en este deporte el número de anotaciones por encuentro es generalmente bajo comparado con otros como por ejemplo baloncesto (Gaviria (2000), Matano et al. (2018)). Para superar este problema en nuestro estudio también contamos con novedosos indicadores que permiten medir con mayor precisión y variabilidad el desempeño de un jugador del campo de juego como lo son la probabilidad de anotar o realizar jugadas importantes dentro de un partido. Dentro de estas nuevas métricas se encuentran los “Goles Esperados” y las “Asistencias Esperadas” los cuales se obtienen a través de modelos predictivo que indican la probabilidad de éxito de la jugada a partir de variables como distancia del disparo, ángulo, tipo de jugada (Tiro libre, jugada individual,

tiro de esquina) parte del cuerpo utilizada para el disparo (pierna, cabeza, entre otros) sin estar estas medidas condicionadas al resultado final del evento tal como se describe en Bundesliga (2018).

Por otra parte, también se cuenta con información sobre el estado de todos los jugadores en cada partido, incluyendo no solo la posición dentro del campo, sino también todas las demás posibles lesiones incluyendo si el jugador estaba lesionado y a su vez, el tipo de lesión por la cual se encontraba ausente. Dada la heterogeneidad en los jugadores y el amplio periodo de tiempo que estamos analizando, es posible encontrar diversos tipos de lesiones, por lo que para efectos prácticos de este análisis se presentan individualmente las lesiones con una frecuencia mayor o igual a 80 y se agrupan en categorías similares todas las lesiones que hubiesen sucedido en menos de 80 ocasiones. El cuadro 2 da una noción de los tipos de lesiones que se encuentran en los datos así como de su frecuencia; para evitar una doble contabilidad de lesiones correlacionadas o simultáneas en nuestras estimaciones principales se incluyen solo lesiones que no hubiesen sido precedidas por otra lesión en los últimos 8 meses, razón por la cual el cuadro 2 presenta las frecuencias tanto para el total de lesiones como para el número de acontecimientos una vez se impone la restricción de los 8 meses ¹.

Para garantizar la homogeneidad de los datos y el tipo de enfrentamientos utilizados, se utilizarán únicamente los partidos correspondientes a las ligas locales, pues estas cuentan con un sistema similar de tabla de posiciones formada con el resultado de los enfrentamientos todos contra todos y una serie de incentivos basados en la posición del equipo al final del torneo, con premios, clasificaciones a torneos internacionales y descensos. Lo anterior genera que algunos jugadores se vean censurados cuando dejan de pertenecer a tales ligas, no obstante, los resultados son robustos a la exclusión de estos jugadores.

¹ Los resultados principales son robustos a la inclusión de todas las lesiones.

Cuadro 2: Numero de lesiones sufridas por tipo. Elaboración propia.

Tipo Lesión	Sin restringir	Sin lesion previa en 8 meses
Abdominal/Intestinal	127	34
Choque	225	79
Cirugía	120	48
Cruzado	141	58
Desgarre Muscular	235	59
Fatiga	105	27
Fractura	295	103
Golpes	198	55
Gripas/Alergias	330	104
Heridas	87	36
Influenza	147	28
Lesión de Isquiotibial (Hamstring)	875	263
Lesión desconocida	448	165
Lesión Hombro	132	143
Lesión Muscular	607	45
Lesión Pantorrilla	214	192
Lesión Pie	118	43
Lesión Rodilla	467	47
Lesión tobillo	380	131
Lesión/Problemas	860	262
Ligamentos	474	155
Muscular	400	124
Ósea (No fractura)	320	110
Otros	258	73
Problema Aductores	362	103
Problema Espalda	138	37
Problema Muscular	544	135
Problema Muslo	435	109
Problema Pantorrilla	130	23
Problemas Rodilla	256	66
Tendón Aquiles	116	34
Tendones	170	50
Total	9714	2941

4. Metodología

Dada la disponibilidad de datos y las diferentes variables de desempeño que se mencionaron anteriormente, es posible utilizar las mediciones de cada partido para los jugadores, con el fin de contrastar de forma correcta el tipo de teorías relacionadas con el desempeño en contextos de equipo y la contribución de la mano de obra a la producción total de tales equipos o unidades económicas. En particular, haciendo uso de choques exógenos que les suceden a los individuos en distintos periodos de tiempo, es posible ajustar los datos con los que se cuenta para realizar una estimación de tipo *event study*, la cual permite computar el efecto de dichos choques en las variables de interés y ha sido ampliamente estudiada previamente (Freyaldenhoven et al. 2019, Roth 2018).

En nuestro caso, los choques exógenos se encuentran definidos por el conjunto de lesiones que les ocurren a los jugadores en el transcurso de la temporada, queriendo identificar el efecto de estas sobre variables de desempeño tanto a nivel individual como resultados agregados del equipo al que pertenece el jugador afectado por el choque.

De acuerdo con Abraham & Sun (2018), debido a la existencia de bases de datos tipo panel, es común en investigación estimar efectos de tratamientos haciendo uso de efectos fijos temporales y de individuo (“two-way”). De igual forma, se está interesado en estimar la dinámica subyacente a estos efectos sobre el tratamiento, lo cual frecuentemente se realiza estimando una ecuación como la siguiente:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\tau} \mu_i 1[t - E_i = \tau] + v_{i,t} \quad (1)$$

“Donde $Y_{i,t}$ es una variable de interés para una unidad i en el periodo t , E_i es el periodo donde la unidad i fue inicialmente tratada y α_i y λ_t son los efectos fijos de individuo y tiempo. . . El tiempo relativo $\tau = t - E_i$ incluido en (1) cubre la mayor parte de los periodos relativos, pero puede excluir otros.” (Abraham & Sun 2018). A su vez, esta metodología cuenta con tres supuestos de identificación claves que es necesario comprender para realizar una correcta estimación:

- Supuesto 1: Tendencias paralelas en las variables para los periodos previos al choque.
- Supuesto 2: No hay conducta anticipatoria previa al tratamiento.
- Supuesto 3: Homogeneidad en el efecto del tratamiento.²

A grandes rasgos, los primeros dos supuestos se relacionan de forma similar al tradicional supuesto de tendencias paralelas del modelo de diferencias en diferencias, en el cual se asume que en la ausencia del evento las diferencias entre las tendencias del grupo tratado y del grupo control hubiesen permanecido constantes; mientras que el supuesto tres se refiere que cada cohorte de tratados experimenta de la misma manera los efectos del tratamiento.

²Estudios recientes como Abraham & Sun (2018), Roth (2018), Freyaldenhoven et al. (2019) han desarrollado nuevas metodologías para afrontar violaciones a estos supuestos y lograr una mejor identificación. Para efectos de este estudio, la discusión sobre la validez de este supuesto será abordada mas adelante en el texto.

Teniendo en cuenta lo anterior, esta metodología presenta características muy apropiadas para su uso en estudios empíricos³, por tanto, en nuestro estudio se estimó una ecuación de la siguiente forma:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{k=-6}^6 \beta_k D_{i,t}^k + \phi_{t<-6} + \phi_{t>6} + v_{i,t} \quad (2)$$

Donde α_i y λ_t son los efectos fijos de individuo y tiempo y los coeficientes de ϕ están asociados a variables que indican los periodos antes de -6 y después de 6 con respecto al periodo donde ocurre el evento, definido este como el partido donde ocurre la lesión o el siguiente partido del equipo en el caso de que la lesión ocurra por fuera de un partido. Con el fin de cumplir con los supuestos anteriormente mencionados, se tomaron en cuenta las lesiones que no estuvieran precedidas por otra lesión en los anteriores 8 meses. Asimismo, dada la posibilidad de que un jugador pudiese experimentar múltiples lesiones que cumplieren con la restricción anterior, se realizaron estimaciones usando datos a nivel jugador-lesión e incluyendo efectos fijos de jugador, como se menciona en la ecuación, al igual que efectos fijos de semana del torneo y temporada para identificar la variación temporal importante en este contexto.

Adicionalmente, dada la heterogeneidad de tipos de lesiones que pueden enfrentar los individuos en el contexto deportivo, es difícil argumentar empíricamente que todas estas cumplan con los supuestos mencionados (principalmente el supuesto 2), ya que algunas de estas se ven afectadas de forma dinámica por realizaciones de los periodos anteriores, es decir, un aumento en el número de minutos jugados o una sobrecarga muscular pueden generar efectos en el desempeño de los jugadores mucho antes del periodo en que se evidencia el tratamiento en los datos. Para enfrentar este problema de la forma más limpia posible se diseñó un algoritmo basado en datos que identifica mediante la extracción de una muestra aleatoria, las lesiones que son más probablemente predecibles en los datos para de esta forma excluirlas del análisis.

³Vease Jacobson et al. (1993)

Es decir, para cada tipo de lesión existente, se estima un event study en una muestra aleatoria del 20 por ciento de los individuos con esta lesión (muestra de entrenamiento), posteriormente se realiza una prueba de hipótesis sobre la probabilidad de que los coeficientes asociados al periodo previo al tratamiento sean conjuntamente diferentes de cero y se guarda el p-valor asociado a la prueba. Finalmente, para el 80 por ciento restante, se estima de nuevo la ecuación sin tener en cuenta como tratamiento los tipos de lesiones con una p-valor inferior a 0.2 en los ejercicios de la muestra de entrenamiento.

De esta forma nos aseguramos que las estimaciones de nuestro modelo no se vean afectadas ya que nos quedamos únicamente con los tipos de lesiones que son poco previsibles de acuerdo con los datos. Además, debido al uso de muestreo aleatorio y al diseño del algoritmo completamente basado en el uso de datos, nos aseguramos de estar seleccionando al mínimo la muestra de forma deliberada evitando así sesgos, dada la ortogonalidad de las bases aleatoriamente seleccionadas.

Finalmente, como se verá en la siguiente sección, es posible realizar una estimación de efectos heterogéneos haciendo uso de la posibilidad de crear mediante los datos categorías de mayor/menor productividad tanto a nivel de equipo como a nivel individual viendo en qué lugar de la distribución se ubicaban justo antes del periodo de la lesión y haciendo una partición por el percentil 75. Con esto, podemos evidenciar empíricamente en qué clase de jugadores, equipos o combinación jugador equipo se encuentran mayormente concentrados los efectos, al estimar una ecuación como la siguiente:

$$Y_{i,t} = \alpha_i + \lambda_t + X'_{i,t}\phi + \sum_{\pi} [\beta_{pre} D_{i,t}^{pre} I_{\pi} + \beta_{post} D_{i,t}^{post} I_{\pi}] + \phi_{t<-6} + \phi_{t>6} + v_{i,t} \quad (3)$$

Donde todas las variables se definen al igual que en 2 y además π representa el número de grupos entre los cuales se quiere ver el efecto y I_{π} es una variable indicadora que toma el valor de uno si el individuo i pertenece al grupo π o cero en otro caso. A diferencia de la regresión tipo *event study* donde se estudia la dinámica del coeficiente, en esta ocasión agrupamos los coeficientes de los periodos previos y posteriores al choque para

asegurarnos de contar con el suficiente poder en los datos para estimar el efecto de cada grupo al no tener que realizar una estimación para un gran número de parámetros.

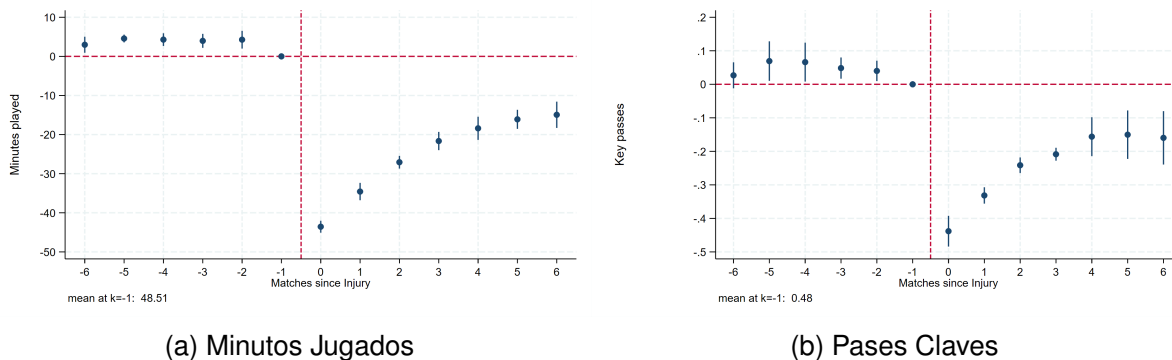
5. Resultados

La figura 1a muestra los coeficientes asociados al ejercicio previo a la aplicación del algoritmo basado en datos para la selección de los tipos de lesiones menos previsibles utilizando como outcomes las variables minutos jugados y pases claves como medidas de la participación de los jugadores en el partido. Como podía esperarse, no todas las lesiones son repentinas o poco anticipables, por el contrario, al tomar en cuenta todos los tipos de lesiones existentes en la base de datos, vemos como existen coeficientes significativos asociados a los parámetros de los periodos previos al evento. Estas “pre-trends” indican que, comparado con el periodo inmediatamente anterior al choque, los periodos anteriores a este los jugadores venían teniendo más minutos por partido como se evidencia en el gráfico 1a. Lo cual puede ser evidencia de un comportamiento anticipatorio.

A su vez, el gráfico 1b nos muestra cómo, además de existir coeficientes significativos en los periodos previos al evento, también la variable de pases claves evidencia una tendencia, es decir, en los periodos previos al periodo de referencia, el número de pases claves realizados por un jugador en cada partido era significativamente mayor, sin embargo, venían disminuyendo en cantidad, lo cual da otro tipo de evidencia de que algunos de los tipos de lesiones en la base de datos no son completamente impredecibles sino que afectan al jugador de forma parcial en periodos anteriores hasta que finalmente este debe ausentarse debido a la lesión.

De acuerdo con lo anterior, es importante asegurarse que el algoritmo desarrollado logre clasificar los tipos de lesiones de tal forma que queden separadas las lesiones menos previsibles de las que tienen influencia en el desempeño de los jugadores en periodos previos para de esta forma cumplir con los supuestos del método y realizar estimaciones causales válidas. La figura 2 muestra el efecto de las lesiones sobre el número de minu-

Figura 1: Estimación sin aplicar el algoritmo



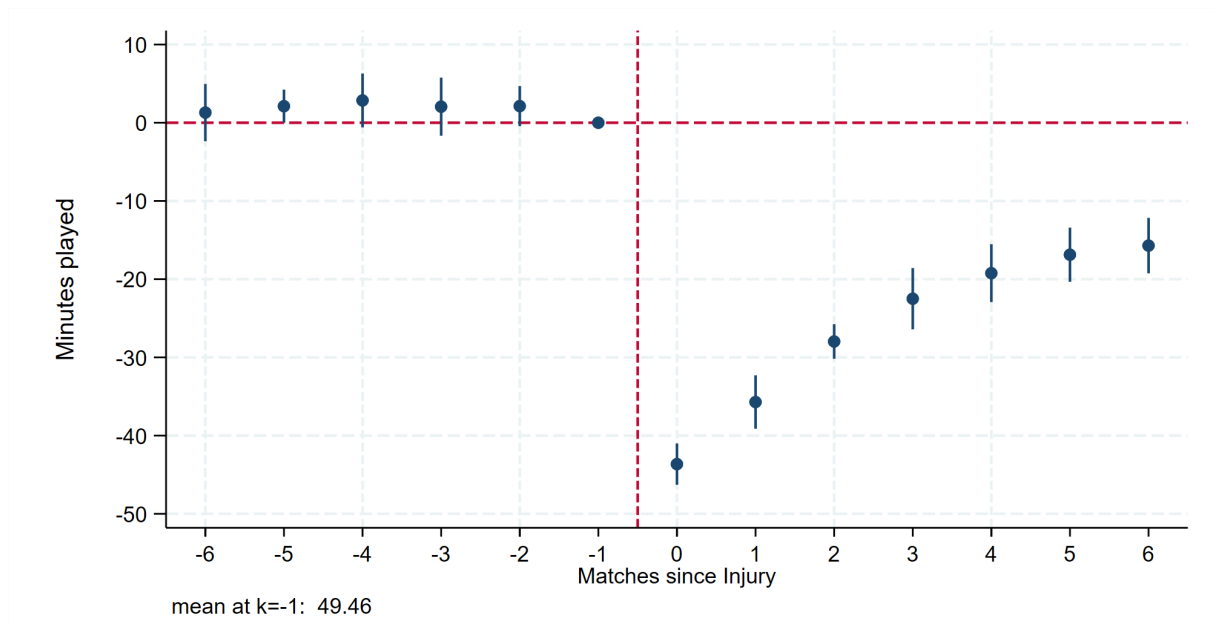
Nota: Esta Figura muestra los coeficientes del estudio de eventos asociados a la estimación de la ecuación (2) usando todas las lesiones. Los resultados exhiben la potencial presencia de comportamientos anticipatorios. Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de individuo, semana del partido-temporada.

tos jugados al aplicar el algoritmo de clasificación basado en datos. En esta podemos ver como, a comparación con la figura 1a los coeficientes asociados a los periodos previos al choque no son estadísticamente significativos al 5% ni existe una tendencia evidente en los estimadores puntuales de tales parámetros.

No obstante, los periodos posteriores al evento sí resultan ser sumamente significativos, por ejemplo, en el periodo 0 los jugadores están presentes en promedio 42 minutos menos, lo cual, comparado con la media del periodo base, se convierte en una reducción relativa del 85% de minutos jugados, siendo este efecto persistente durante los siguientes 6 partidos aunque el estimador puntual disminuye conforme se aleja el periodo del choque. Es decir, los jugadores siguen jugando menos minutos, aunque cada vez esta disminución es menor pues conforme pasan los partidos es probable que los jugadores se vayan recuperando de la lesión.

Así como el número de minutos jugados se ve afectado por las lesiones, los desempeños y participaciones de estos jugdores en los encuentros también sufren un efecto. La figura 3 contiene los resultados del efecto de las lesiones sobre diferentes tipos de variables de desempeño individual. A grandes rasgos, podemos ver cómo, sin importar la

Figura 2: Efecto de la lesión en el número de minutos jugados

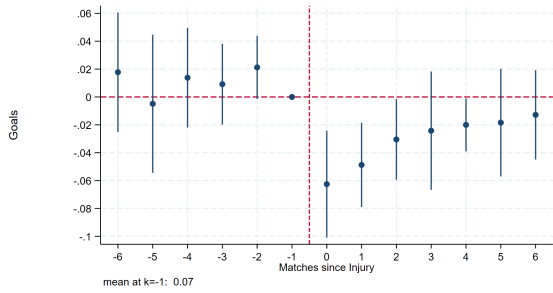


Nota: Esta Figura muestra los coeficientes del estudio de eventos asociados a la estimación de la ecuación (2) usando las lesiones seleccionadas con el algoritmo de clasificación. Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de individuo, semana del partido-temporada.

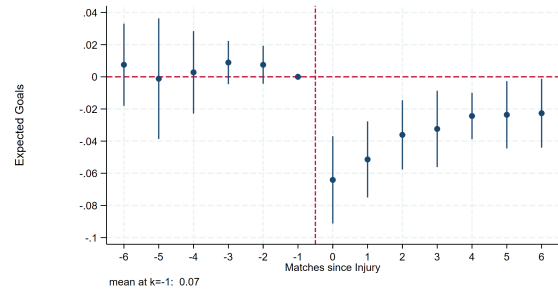
medida que se escoja, existe evidencia de que la lesión genera un choque negativo y significativo en la productividad de los jugadores, medida en este caso como el desempeño dentro de los partidos. Es decir, el factor productivo de mano de obra, efectivamente sufre un choque negativo que se evidencia en su productividad.

Por ejemplo, vemos como tanto el número de goles reales (Figura 3a), como esperados (Figura 3b), tienen una disminución inmediata relativa de aproximadamente del 90%. A su vez, vemos como la estimación para goles esperados es más precisa que su contraparte en términos de error estándar, pues, como se mencionó anteriormente, las variables como goles tienen poca variabilidad en deportes de pocas anotaciones como el fútbol, así, el efecto en goles esperados es persistente durante más periodos, aunque de igual forma, luego de aproximadamente 6 partidos, el efecto deja de ser estadísticamente distinto de cero.

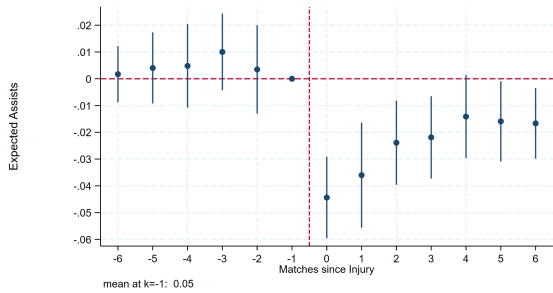
Figura 3: Efecto de la lesion en variables individuales



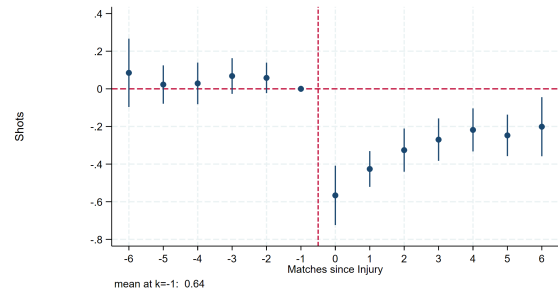
(a) Goles



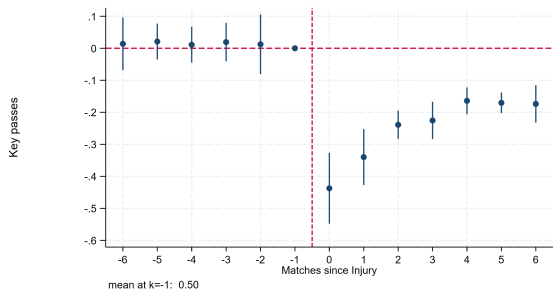
(b) Goles Esperados



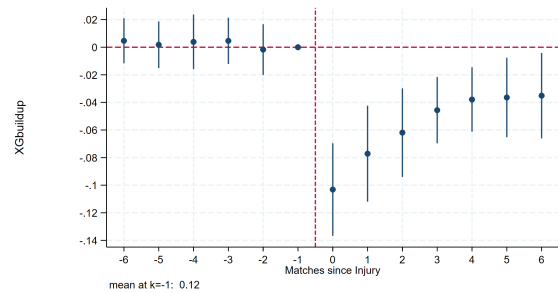
(c) Asistencias Esperadas



(d) Disparos



(e) Pases Claves



(f) XGBuildUp

Nota: Esta Figura muestra los coeficientes del estudio de eventos asociados a la estimación de la ecuación (2) usando únicamente las lesiones menos predecibles de acuerdo con el algoritmo basado en datos. Se evidencia una disminución en la productividad de los jugadores medida usando diferentes variables de desempeño. Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de individuo, semana del partido-temporada.

También, el efecto sobre otras variables de desempeño individual presenta una dinámica similar, en la cual, el periodo del choque presenta el efecto de mayor magnitud y este va disminuyendo conforme avanzan los partidos. Cabe resaltar que, a diferencia de la figura 1b, en este caso, ninguna de las variables presenta tendencia en los periodos previos al choque. Otro punto importante a mencionar es el efecto en la variable XGBuildUp, esta es

una variable innovadora en las métricas de fútbol pues cuantifica el total de goles esperados de todas las posesiones en las cuales el jugador estuvo involucrado en el partido, descontando pases claves y disparos durante los 90 minutos, es decir, esta variable nos da una medida agregada del desempeño del jugador, siendo los *compliers* los jugadores ofensivos al ser la mayoría de estas variables, medidas de desempeño ofensivo.

Teniendo en cuenta la dinámica presentada por los resultados anteriores, siendo esta libre de “pre-trends” y tendencias en los periodos previos al choque, continuaremos analizando los resultados de estimaciones similares al modelo de diferencias en diferencias, donde el coeficiente presentado se interpreta como el promedio del efecto en los 6 periodos siguientes al evento⁴.

Cuadro 3: Efecto de la lesion en variables individuales

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Minutos Jugados	Goles	Goles Esperados	Asistencias Esperadas	Disparos	Pases Claves	XGbuildup
Pre	2.087* (0.971)	0.0114 (0.00829)	0.00509 (0.00579)	0.00480 (0.00463)	0.0526 (0.0317)	0.0155 (0.0157)	0.00264 (0.00463)
Post	-25.95*** (0.816)	-0.0310** (0.00997)	-0.0364*** (0.00618)	-0.0247*** (0.00438)	-0.322*** (0.0346)	-0.250*** (0.0120)	-0.0568*** (0.00973)
Constante	43.29*** (1.019)	0.0548*** (0.00820)	0.0600*** (0.00469)	0.0421*** (0.00417)	0.549*** (0.0293)	0.424*** (0.0183)	0.0959*** (0.00766)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.339	0.124	0.225	0.142	0.297	0.235	0.159
Media en periodo t-1	49.46	0.0700	0.0700	0.0500	0.640	0.500	0.120

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de individuo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10 %, 5 % y 1 %, respectivamente.

El cuadro 3 resume la evidencia del efecto del choque en los resultados individuales, similar a las gráficas, vemos como existe un efecto negativo y significativo en todas las variables. Por ejemplo, hay una reducción relativa del 52.4 % en el número de minutos jugados promedio en los siguientes seis partidos, una disminución del 44 % en la probabilidad de marcar gol y del 52 % en los goles esperados, y de igual forma, el resto de variables presenta disminuciones de alrededor del 50 %. Estos resultados muestran como el conjunto de lesiones seleccionadas presentan un comportamiento que nos permite identificar de

⁴Los resultados para el resto de variables de interés se encuentran en el apéndice A

forma causal el efecto de esos eventos sobre el desempeño individual. Ahora bien, una pregunta relevante en el contexto de un sistema producción es cómo estas caídas en la productividad de los jugadores, pueden terminar afectando la productividad total de la firma. Para responder a esta pregunta realizamos regresiones donde la variable de pendiente es el desempeño del equipo. El cuadro 4 muestra el efecto del choque sobre las variables a nivel de equipo, contrario a lo anterior, vemos como los coeficientes asociados a las variables de desempeño grupal, a pesar de ser también negativas, son sólo marginalmente o no significativos, es decir, en promedio las lesiones individuales pareciesen no tener un efecto estadísticamente significativo o de gran magnitud sobre las variables de creación de juego agregadas medidas en número de pases claves, *NPXG*, numero de disparos, goles y asistencias esperadas; o medidas más complejas como *XGBuildUp* o *XGChain*.

Cuadro 4: Efecto de la lesion en variables de equipo

	(1) Pases Clave	(2) XGChain	(3) NPXG	(4) Disparos	(5) Asistencias Esperadas	(6) Goles Esperados	(7) XGBuildUp
Pre	-0.172 (0.117)	-0.0417 (0.0756)	-0.00596 (0.00719)	-0.243 (0.152)	-0.00832 (0.00859)	0.00160 (0.00916)	-0.0339 (0.0666)
Post	-0.223 (0.115)	-0.103 (0.0571)	-0.0308* (0.0115)	-0.331* (0.147)	-0.0182 (0.0129)	-0.0207 (0.0153)	-0.0772 (0.0500)
Constante	9.052*** (0.0854)	3.472*** (0.0773)	1.192*** (0.0170)	12.13*** (0.115)	0.920*** (0.0125)	1.293*** (0.0190)	2.060*** (0.0604)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.175	0.218	0.183	0.171	0.178	0.177	0.209
Media en periodo t-1	9.270	3.690	1.240	12.41	0.960	1.350	2.200

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de equipo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10 %, 5 % y 1 %, respectivamente.

Dada la evidencia de las figuras y tablas presentadas hasta el momento no se puede rechazar la hipótesis de que, a pesar de que un choque exógeno que genera la ausencia de un factor productivo dentro la mano de obra tiene efecto sobre la productividad de este factor, la unidad productiva o equipo es capaz de responder de forma óptima a este evento para suplir de alguna forma tal disminución y que la producción agregada no se vea afectada, es decir, los equipos logran reemplazar la contribución de los jugadores lesionados de tal forma que el producto o desempeño del equipo, en promedio, no se altere. Sin embargo nuestra falta de poder para rechazar esta hipótesis puede deberse a la posibilidad de estar

mezclando efectos heterogéneos que estudiados de forma individual podrían dar mayor luz sobre el impacto de estos choques exógenos a la productividad individual, heterogeneidad que pretendemos explotar en la siguiente sección.

5.1. Efectos Heterogéneos

Esta sección comprende los resultados de efectos heterogéneos usando la doble interacción entre equipos de productividad alta y productividad baja y jugadores de productividad alta y productividad baja, para realizar tal clasificación se tomó en cuenta la distribución de la variable *XGChain* acumulada hasta la fecha de la lesión tanto a nivel individual como de equipo y se clasificó como productividad alta aquellos por encima del percentil 75.⁵ La importancia de esta variable, que suma los goles esperados de todas las jugadas en las que el jugador se vio involucrado, es que cuenta con características preferibles en nuestro contexto, como el hecho de que cuenta con una mayor variabilidad para los jugadores en todas las posiciones del campo, lo cual permite una mejor identificación del nivel de productividad de estos.

En cuanto a los resultados de productividad individual, el cuadro 5 muestra como el efecto de la lesión se mantiene en la mayoría de los casos sobre los cuatro grupos de análisis, evidenciando magnitudes de los coeficientes asociados al efecto relativamente similares, a pesar de que para el grupo que comprende los jugadores de productividad baja en equipos de alta productividad, el efecto en goles y disparos deja de ser significativo. Y que el efecto en asistencias esperadas y pases claves a pesar de que continúan siendo significativo, ya solo lo es marginalmente debido probablemente a una menor precisión en la estimación.

El cuadro 6 por su lado presenta los resultados a nivel de equipo, donde podemos observar los efectos en las variables de desempeño agregadas, la importancia de esta tabla

⁵Dado la ausencia de indicadores de corte netamente defensivo, la mayoría de variables analizadas en el estudio consideran principalmente el aspecto ofensivo del desempeño de los jugadores.

Cuadro 5: Efectos heterogéneos de la lesión en variables individuales

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Minutos Jugados	Goles	Goles Esperados	Asistencias Esperadas	Disparos	Pases Claves	XGbuildup
(a) J. Baja - E. Baja	-26.49*** (1.391)	-0.0167* (0.00782)	-0.0243*** (0.00430)	-0.0187** (0.00497)	-0.241*** (0.0311)	-0.192*** (0.0168)	-0.0436*** (0.00697)
(b) J. Baja - E. Alta	-19.13*** (1.207)	-0.0200 (0.0107)	-0.0225* (0.00935)	-0.0109 (0.00514)	-0.180* (0.0696)	-0.123*** (0.0256)	-0.0439*** (0.00864)
(c) J. Alta - E. Baja	-28.27*** (1.135)	-0.0487** (0.0129)	-0.0535*** (0.00726)	-0.0288*** (0.00447)	-0.450*** (0.0394)	-0.323*** (0.0226)	-0.0611*** (0.0108)
(d) J. Alta - E. Alta	-26.12*** (1.037)	-0.0448** (0.0141)	-0.0481*** (0.00938)	-0.0392*** (0.00513)	-0.414*** (0.0376)	-0.348*** (0.0272)	-0.0842*** (0.0129)
Constante	42.89*** (1.066)	0.0536*** (0.00852)	0.0590*** (0.00490)	0.0413*** (0.00407)	0.538*** (0.0312)	0.416*** (0.0183)	0.0944*** (0.00776)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.340	0.124	0.225	0.142	0.298	0.236	0.159
Media en periodo t-1	49.46	0.0700	0.0700	0.0500	0.640	0.500	0.120
Media (a)	44.39	0.0200	0.0400	0.0200	0.370	0.270	0.0700
Media (b)	33.81	0.0500	0.0400	0.0200	0.270	0.140	0.0700
Media (c)	56.14	0.0700	0.0900	0.0700	0.840	0.730	0.130
Media (d)	60.36	0.160	0.130	0.100	1.130	0.860	0.220

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de individuo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10 %, 5 % y 1 %, respectivamente.

recae en que empíricamente, es visto que algunos equipos si se ven afectados por la ausencia de cierta clase de jugadores, pudiendo ser los resultados de las estimaciones de variables agregadas, previamente presentados, estadísticamente no significativos debido principalmente a la no identificación de un grupo relevante o, lo que es similar, la inclusión de cierto ruido en las estimaciones que no permite validar correctamente la hipótesis del efecto de un choque individual sobre el desempeño agregado. A su vez, un jugador de cierto nivel de habilidad específico puede tener un efecto diferente dependiendo de del tipo de jugadores que lo rodean, es decir, la calidad del equipo al que pertenece, por tanto es importante controlar al mismo tiempo por los dos márgenes de variación como sucede en el cuadro 6

Los resultados de estas estimaciones indican que, para el grupo de jugadores de alta productividad en equipos de baja productividad, las variables de pases claves, *XGChain*, Goles esperados no de penalty (*NPXG*), disparos, asistencias esperadas, goles esperados y *XGBuildUp* pasan a ser significativas evidenciando una disminución relativa de alrededor de entre 6 % y 8 %. Siendo este el único grupo de los 4 que evidencia efectos significativos

Cuadro 6: Efectos heterogéneos de la lesión en variables a nivel de equipo

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Pases Clave	XGChain	NPXG	Disparos	Asistencias Esperadas	Goles Esperados	XGBuildUp
(a) J. Baja - E. Baja	-0.113 (0.120)	-0.0483 (0.0607)	-0.0160 (0.0154)	-0.206 (0.127)	-0.00453 (0.0121)	-0.000947 (0.0186)	-0.0420 (0.0554)
(b) J. Baja - E. Alta	-0.150 (0.211)	-0.166 (0.102)	-0.0265 (0.0397)	-0.232 (0.310)	-0.0170 (0.0349)	-0.0136 (0.0410)	-0.130 (0.0630)
(c) J. Alta - E. Baja	-0.614*** (0.112)	-0.270** (0.0779)	-0.0747** (0.0172)	-0.754*** (0.112)	-0.0571** (0.0162)	-0.0756** (0.0222)	-0.179** (0.0626)
(d) J. Alta - E. Alta	-0.0382 (0.210)	0.0114 (0.0905)	-0.0125 (0.0306)	-0.152 (0.286)	-0.00159 (0.0264)	-0.00126 (0.0272)	-0.00173 (0.0730)
Constante	9.036*** (0.0788)	3.470*** (0.0755)	1.191*** (0.0162)	12.12*** (0.113)	0.919*** (0.0120)	1.291*** (0.0182)	2.060*** (0.0592)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.175	0.218	0.183	0.172	0.178	0.177	0.209
Media en periodo t-1	9.270	3.690	1.240	12.41	0.960	1.350	2.200
Media (a)	8.190	2.840	1.050	11.21	0.790	1.130	1.600
Media (b)	10.15	4.280	1.360	13.43	1.070	1.510	2.650
Media (c)	8.410	3.120	1.100	11.27	0.860	1.170	1.820
Media (d)	11.79	5.600	1.690	15.39	1.330	1.850	3.520

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de equipo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10 %, 5 % y 1 %, respectivamente.

al 5 %. Además, este resultado resulta ser sumamente intuitivo ya que no es sorprendente que los equipos de más baja productividad sean los que dependen en mayor medida de sus jugadores, más aún si se trata de un jugador que al momento de su lesión tenía un nivel de productividad alto, lo que a su vez, está en línea con las creencias populares de los seguidores de fútbol. ⁶

Por tanto, esta diferenciación de los equipos y jugadores sugiere que existen ciertas características bajo las cuales, no es siempre posible suplir total y óptimamente un choque en la productividad de los factores de producción, más aún, estos choques generan una disminución en la capacidad de juego del equipo. Por último, a pesar de que la evidencia en probabilidad de derrota, empate o victoria no es estadísticamente significativa, vemos como para el grupo en cuestión, los signos asociados a estos coeficientes van de acuerdo a la intuición.

⁶Se estimaron también los efectos en la probabilidad de derrota, empate o victoria donde debido quizás a la poca variabilidad de estas medidas no encontramos efectos significativos (Ver Apéndice A)

La tabla 7 presenta una estimación de tipo *Leave-Out Means*, es decir, estima el efecto del choque sobre las variables agregadas a nivel de equipo descontando la contribución individual de cada jugador. A pesar de que únicamente resulta ser estadísticamente significativo los coeficientes asociados a pases claves y disparos (*XGChain marginalmente significativo*), con una caída en el desempeño total del equipo de alrededor del 3.77% y 2.9% respectivamente, en este caso resulta sumamente interesante evidenciar los signos y magnitudes de los 4 grupos en cuestión. En primer lugar, para los grupos de jugadores con baja productividad, vemos como en comparación con los de productividad alta, los efectos relativos son mucho más pequeños y cercanos a cero. Luego, vemos como para los jugadores más productivos en equipos productivos, el signo de la estimación es positivo en todas las variables; esto sugiere el hecho de que, en este tipo de equipos, las lesiones de jugadores productivos son sustituidas principalmente por jugadores también productivos, lo cual hace que, al descontar su contribución, el resto de jugadores tengan un efecto positivo por la entrada de un jugador sustituto y/o la mayor contribución del resto de jugadores. Por el contrario, el efecto en los jugadores productivos en equipos menos productivos es negativo y significativo en algunos casos, lo cual, en este caso da una intuición de complementariedad de este jugador productivo sobre sus compañeros, es decir, aun descontando la disminución de la productividad individual, el resto de jugadores tienen un peor desempeño, pues el jugador afectado potenciaba el rendimiento de sus compañeros en el campo.

Finalmente, el cuadro 8 presenta una estimación de las elasticidades cruzadas tanto individuales, es decir, como afecta a cada jugador la ausencia por lesión de otro, como las agregadas, cuya interpretación es el efecto agregado sobre el equipo. Vemos como para los grupos de jugadores poco productivos los valores son más pequeños en comparación y pueden cambiar de signo dependiendo de la variable. No obstante, en cuanto a los jugadores buenos en equipos poco productivos, es siempre robusto el signo positivo que refuerza la idea de una complementariedad entre esta clase de jugadores y sus compañeros de equipo, mientras que, para jugadores buenos en equipos productivos, el signo negativo refleja la sustitución que puede existir a esta clase de jugadores en equipos donde la mayor parte de la mano de obra es altamente productiva.

Cuadro 7: Estimación en *Leave-Out Means* a nivel de equipo.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Pases Clave	XGChain	NPXG	Disparos	Asistencias Esperadas	Goles Esperados	XGBuildUp
(a) J. Baja - E. Baja	0.0792 (0.116)	0.0215 (0.0526)	0.00679 (0.0129)	0.0346 (0.117)	0.0142 (0.0108)	0.0233 (0.0160)	0.00162 (0.0492)
(b) J. Baja - E. Alta	-0.0271 (0.198)	-0.0992 (0.102)	-0.00595 (0.0400)	-0.0515 (0.295)	-0.00606 (0.0321)	0.00895 (0.0418)	-0.0866 (0.0619)
(c) J. Alta - E. Baja	-0.290** (0.0955)	-0.153* (0.0701)	-0.0296 (0.0170)	-0.305** (0.0731)	-0.0283 (0.0145)	-0.0222 (0.0204)	-0.118 (0.0564)
(d) J. Alta - E. Alta	0.310 (0.231)	0.153 (0.0873)	0.0352 (0.0320)	0.262 (0.329)	0.0376 (0.0300)	0.0468 (0.0274)	0.0824 (0.0679)
Constante	8.620*** (0.0735)	3.310*** (0.0664)	1.137*** (0.0140)	11.58*** (0.0874)	0.878*** (0.00836)	1.232*** (0.0159)	1.965*** (0.0535)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.171	0.214	0.179	0.170	0.171	0.174	0.205
Media en periodo t-1	8.780	3.490	1.180	11.77	0.910	1.270	2.080
Media (a)	7.920	2.740	1.020	10.84	0.760	1.100	1.530
Media (b)	10.01	4.170	1.320	13.17	1.050	1.470	2.580
Media (c)	7.680	2.890	1.020	10.43	0.800	1.080	1.690
Media (d)	10.93	5.220	1.570	14.25	1.230	1.710	3.300

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de equipo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10 %, 5 % y 1 %, respectivamente.

Cuadro 8: Cálculo de elasticidades cruzadas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Pases Clave	XGChain	NPXG	Disparos	Asistencias Esperadas	Goles Esperados	XGBuildUp
Panel A - Elasticidad individual							
(a) J. Baja - E. Baja	-0.0141	-0.0112	-0.0088	-0.0049	-0.0200	-0.0349	-0.0017
(b) J. Baja - E. Alta	0.0031	0.0425	0.0088	0.0059	0.0106	-0.0108	0.0535
(c) J. Alta - E. Baja	0.0853	0.1050	0.0515	0.0546	0.0860	0.0346	0.1486
(d) J. Alta - E. Alta	-0.0701	-0.0790	-0.0564	-0.0502	-0.0780	-0.0740	-0.0652
Panel B - Elasticidad Agregada							
(a) J. Baja - E. Baja	-0.4125	-0.3080	-0.2978	-0.1436	-0.7594	-0.9588	-0.0372
(b) J. Baja - E. Alta	0.2203	1.4784	0.2888	0.2861	0.5560	-0.3978	1.9727
(c) J. Alta - E. Baja	0.8978	1.3190	0.6563	0.6778	0.9826	0.4150	1.9313
(d) J. Alta - E. Alta	-0.8908	-1.0851	-0.7379	-0.6329	-0.9592	-0.9730	-0.9786

Nota: En el panel A, las elasticidades cruzadas individuales se calculan como la división entre el efecto relativos de las estimación *Leave-Out Means* a nivel de equipo sobre el efecto relativo de la estimación individual, elasticidades mayores a 0.05 en valor absoluto en negrita. En el panel B, las elasticidades cruzadas agregadas se calculan como el coeficiente asociado al efecto de la lesión en la estimación *Leave-Out Means* sobre el coeficiente de la estimación individual, elasticidades mayores a 0.6 en valor absoluto en negrita.

6. Conclusiones

Este documento aporta nueva evidencia sobre el efecto de una disminución en la productividad individual de uno de los factores productivos sobre la producción total en contextos de equipo. Haciendo uso de eventos exógenos determinados por lesiones poco previstas a jugadores de fútbol, se identifica una disminución en el desempeño individual significativa en los periodos posteriores al choque, siendo este medido mediante diferentes variables como goles, asistencias esperadas o medidas modernas de creación de juego; mientras que no se encuentra efecto sobre el desempeño agregado del equipo. No obstante, al dividir la muestra en equipos y jugadores de acuerdo a su nivel de productividad, se encuentra que los jugadores productivos en equipos poco productivos tienen un efecto relativo de aproximadamente 6%-8% en el total de disparos, asistencias esperadas, goles esperados y generación de juego del equipo. Al calcular las elasticidades cruzadas para los 4 tipos de grupos identificados, vemos como para los grupos de jugadores poco productivos no es del todo concluyente debido a valores cercanos a cero y cambios de signos dependiendo de la variable que se analice, no obstante, para jugadores altamente productivos en equipos de baja productividad, se evidencia un signo positivo que indica la complementariedad de este tipo de jugadores sobre el equipo, mientras que los jugadores más productivos en equipos altamente productivos, evidencian por su parte un signo negativo que indica una posible sustituibilidad entre este jugador y sus compañeros al estar por definición en un equipo con mayor proporción de jugadores productivos.

6.1. Trabajo Futuro

Este artículo genera las bases para una agenda de investigación sobre la contribución individual de los factores de producción al producto agregado en contextos de equipos. Como pasos adicionales para el futuro se encuentran la descarga y uso de datos de corte defensivo, que nos permita tener una medida más completa del desempeño de los jugadores y que haga el cálculo de interacciones cruzadas a nivel de roles dentro del campo de juego, analizando por ejemplo el efecto en la productividad ofensiva de la pérdida de

jugadores de corte defensivo. Así mismo, el trabajo futuro debe enfocarse en calcular intervalos de confianza para las elasticidades presentadas en este informe bien sea mediante una estimación tipo *Bootstrap* o una estimación en 2 etapas que calcule el denominador y el numerador de tales elasticidades de forma conjunta. Finalmente, esta en los planes realizar ejercicios de robustez sobre los supuestos intrínsecos del *evento study* en particular el supuesto de homogeneidad del efecto que dado lo reciente de las innovaciones en este campo y las restricciones de tiempo, no pudieron ser realizados para la presentación de este informe, no obstante se proponen como trabajo futuro.

Referencias

Abraham, S. & Sun, L. (2018), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *arXiv preprint arXiv:1804.05785*.

Arcidiacono, P., Kinsler, J. & Price, J. (2017), 'Productivity spillovers in team production: Evidence from professional basketball', *Journal of Labor Economics* **35**(1), 191–225.

Bonhomme, S. (2020), 'Teams: Heterogeneity, sorting and complementarity', *A preliminary draft of the work presented at the 2020 World Congress of the Econometric Society*.

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D. & Singleton, C. (2020), 'Experimental effects of an absent crowd on performances and refereeing decisions during covid-19', *Available at SSRN 3668183*.

Bundesliga (2018), 'xg stats explained: The science behind sportec solutions' expected goals model', <https://www.bundesliga.com/en/bundesliga/news/expected-goals-xg-model-what-is-it-and-why-is-it-useful-sportec-solutions-3177>.

Caruso, R., Di Domizio, M. & Rossignoli, D. (2017), 'Aggregate wages of players and performance in italian serie a', *Economia Politica* **34**(3), 515–531.

Devereux, K. (2018), Identifying the value of teamwork: Application to professional tennis, Technical report, Working Paper Series.

- Dvorak, J., Junge, A., Graf-Baumann, T. & Peterson, L. (2004), 'Football is the most popular sport worldwide', *American journal of sports medicine* **32**(Suppl 1), 3S–4S.
- Efe (2020), 'Las cinco grandes ligas europeas, cada vez más poderosas económicamente', <https://www.mundodeportivo.com/futbol/internacional/20200117/472930369346/las-cinco-grandes-ligas-europeas-cada-vez-mas-poderosas-economicamente.html>.
- Ernst&Young-LLP (2019), Premier league: Economic and social impact, Technical report, Ernst Young LLP.
- Freyaldenhoven, S., Hansen, C. & Shapiro, J. M. (2019), 'Pre-event trends in the panel event-study design', *American Economic Review* **109**(9), 3307–38.
- Gauriot, R. & Page, L. (2019), 'Fooled by performance randomness: Overrewarding luck', *Review of Economics and Statistics* **101**(4), 658–666.
- Gaviria, A. (2000), 'Is soccer dying? a time series approach', *Applied Economics Letters* **7**(4), 275–278.
- Jacobson, L. S., LaLonde, R. J. & Sullivan, D. G. (1993), 'Earnings losses of displaced workers', *The American economic review* pp. 685–709.
- Matano, F., Richardson, L. F., Pospisil, T., Eubanks, C. & Qin, J. (2018), 'Augmenting adjusted plus-minus in soccer with fifa ratings', *arXiv preprint arXiv:1810.08032*.
- Neale, W. C. (1964), 'The peculiar economics of professional sports', *The quarterly journal of economics* **78**(1), 1–14.
- Rosen, S. & Sanderson, A. (2001), 'Labour markets in professional sports', *The economic journal* **111**(469), F47–F68.
- Roth, J. (2018), Pre-test with caution: Event-study estimates after testing for parallel trends, Technical report, Working Paper.
- Rottenberg, S. (1956), 'The baseball players' labor market', *Journal of political economy* **64**(3), 242–258.
- Scully, G. W. (1995), *The market structure of sports*, University of Chicago Press.

- Sloane, P. J. (1971), 'Scottish journal of political economy: the economics of professional football: the football club as a utility maximiser', *Scottish journal of political economy* **18**(2), 121–146.
- Stubbe, J. H., van Beijsterveldt, A.-M. M., van der Knaap, S., Stege, J., Verhagen, E. A., Van Mechelen, W. & Backx, F. J. (2015), 'Injuries in professional male soccer players in the netherlands: a prospective cohort study', *Journal of athletic training* **50**(2), 211–216.
- Zuke, W. A., Agarwalla, A., Go, B., Griffin, J. W., Cole, B. J., Verma, N. N., Bach, B. R. & Forsythe, B. (2018), 'The lack of standardized outcome measures following lower extremity injury in elite soccer: a systematic review', *Knee surgery, sports traumatology, arthroscopy* **26**(10), 3109–3117.

7. Apéndice A

Cuadro 9: Efecto de la lesion en variables individuales

	(1) Asistencias	(2) T. Amarillas	(3) T. Rojas	(4) NPG	(5) NPGX	(6) XGchain
Pre	-0.00194 (0.00499)	0.0217 (0.0119)	0.00174 (0.000992)	0.0140 (0.00838)	0.00719 (0.00608)	0.00986 (0.00781)
Post	-0.0303*** (0.00487)	-0.0474** (0.0128)	0.00102* (0.000384)	-0.0279** (0.0100)	-0.0334*** (0.00650)	-0.0965*** (0.0120)
Constante	0.0521*** (0.00450)	0.0819*** (0.0114)	0.000583 (0.000587)	0.0496*** (0.00803)	0.0550*** (0.00520)	0.163*** (0.00939)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.065	0.062	0.010	0.108	0.215	0.222
Media en periodo t-1	0.0600	0.100	0	0.0600	0.0600	0.190

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de individuo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10%, 5% y 1%, respectivamente.

Cuadro 10: Efecto de la lesion en variables de equipo

	(1) Empate	(2) Victoria	(3) Derrota	(4) Gol Diferencia	(5) Goles Anotados	(6) Goles Concebidos
Pre	-0.00670 (0.0195)	0.0254 (0.0134)	-0.0186 (0.0174)	0.0468 (0.0524)	0.00298 (0.0335)	-0.0438 (0.0387)
Post	-0.00770 (0.0224)	0.0141 (0.0192)	-0.00644 (0.0270)	-0.0175 (0.0586)	-0.0496 (0.0327)	-0.0321 (0.0494)
Constante	0.253*** (0.0221)	0.369*** (0.0197)	0.378*** (0.0231)	0.0281 (0.0620)	1.403*** (0.0423)	1.375*** (0.0422)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.022	0.086	0.071	0.118	0.110	0.060
Media en periodo t-1	0.250	0.390	0.370	0.100	1.470	1.370

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de equipo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10%, 5% y 1%, respectivamente.

Cuadro 11: Efectos heterogéneos de la lesión en variables individuales

	(1) Asistencias	(2) T. Amarillas	(3) T. Rojas	(4) NPG	(5) NPGX	(6) XGchain
(a) J. Baja - E. Baja	-0.0249*** (0.00475)	-0.0513** (0.0160)	0.000537 (0.000748)	-0.0151 (0.00805)	-0.0228** (0.00504)	-0.0698*** (0.00858)
(b) J. Baja - E. Alta	-0.0173* (0.00699)	-0.0213 (0.0189)	0.00118 (0.00112)	-0.0174 (0.0100)	-0.0206* (0.00889)	-0.0671*** (0.0144)
(c) J. Alta - E. Baja	-0.0381*** (0.00686)	-0.0531*** (0.00750)	0.00146 (0.000826)	-0.0390** (0.0117)	-0.0451*** (0.00739)	-0.116*** (0.0124)
(d) J. Alta - E. Alta	-0.0392*** (0.00691)	-0.0483** (0.0138)	0.00134 (0.00122)	-0.0458** (0.0152)	-0.0477*** (0.0101)	-0.141*** (0.0180)
Constante	0.0506*** (0.00444)	0.0812*** (0.0111)	0.000577 (0.000601)	0.0484*** (0.00830)	0.0540*** (0.00535)	0.160*** (0.00955)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.065	0.062	0.010	0.108	0.216	0.223
Media en periodo t-1	0.0600	0.100	0	0.0600	0.0600	0.190
Media (a)	0.0200	0.100	0	0.0200	0.0300	0.100
Media (b)	0.0200	0.0700	0.0100	0.0500	0.0400	0.120
Media (c)	0.0800	0.100	0	0.0600	0.0800	0.230
Media (d)	0.150	0.100	0	0.140	0.120	0.380

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de individuo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10%, 5% y 1%, respectivamente.

Cuadro 12: Efectos heterogéneos de la lesión en variables de equipo

	(1) Empate	(2) Victoria	(3) Derrota	(4) Gol Diferencia	(5) Goles Anotados	(6) Goles Concebidos
(a) J. Baja - E. Baja	-0.00582 (0.0257)	0.0174 (0.0242)	-0.0116 (0.0249)	0.0247 (0.0596)	-0.0268 (0.0435)	-0.0515 (0.0378)
(b) J. Baja - E. Alta	-0.0310 (0.0216)	0.0276 (0.0339)	0.00338 (0.0387)	-0.0137 (0.117)	-0.0308 (0.0651)	-0.0171 (0.0809)
(c) J. Alta - E. Baja	-0.00654 (0.0177)	-0.00119 (0.0239)	0.00773 (0.0292)	-0.0871 (0.0807)	-0.0824 (0.0412)	0.00476 (0.0557)
(d) J. Alta - E. Alta	0.000297 (0.0342)	0.0176 (0.00830)	-0.0179 (0.0310)	-0.0223 (0.0491)	-0.0668 (0.0380)	-0.0445 (0.0652)
Constante	0.253*** (0.0224)	0.369*** (0.0196)	0.378*** (0.0236)	0.0258 (0.0617)	1.401*** (0.0400)	1.375*** (0.0424)
Observaciones	592,617	592,617	592,617	592,617	592,617	592,617
R-Cuadrado	0.022	0.086	0.071	0.119	0.110	0.060
Media en periodo t-1	0.250	0.390	0.370	0.100	1.470	1.370
Media (a)	0.250	0.320	0.430	-0.290	1.230	1.520
Media (b)	0.310	0.430	0.270	0.490	1.670	1.180
Media (c)	0.280	0.270	0.450	-0.250	1.270	1.510
Media (d)	0.180	0.610	0.210	1	2.030	1.030

Nota: Las estimaciones incluyen efectos fijos de individuo, semana del partido-temporada. Errores estándar tipo clúster a nivel de equipo, semana del partido-temporada. *, **, y *** indican nivel de significancia al 10%, 5% y 1%, respectivamente.