

,

Master's Thesis

Identificación empírica del cambio del poder de negociación de agentes libres de la MLB.

Maestría en Economía
Centro de Investigación y Docencias Económicas.

Antonio Huerta Montellano

CDMX, Mayo, 2023.



CIDE

Supervisada por Dr. Sonia DiGiannatale

Resumen

En la presente tesis se lleva a cabo un análisis empírico con el propósito de identificar las medidas de desempeño que explican las variaciones en el poder de negociación de los agentes libres de *Major League Baseball (MLB)*, quienes son jugadores que tienen la prerrogativa de elegir el equipo para el que van a jugar durante el periodo de contrato. Además, se realiza un estudio del impacto de diversas variables como la edad o el estatus de agente libre en este cambio de poder de negociación. En el caso de la edad, se examinan los diferentes efectos de esta variable a lo largo del tiempo. Finalmente, se examinan casos adicionales de interés, como la robustez de los resultados y los cambios estructurales en distintas especificaciones con el fin de obtener conclusiones más sólidas sobre las hipótesis planteadas en este trabajo.

Índice

1. Introducción	1
1.1. Dinámica de un partido de béisbol	5
1.2. Problema de Riesgo Moral	7
1.3. Modelo de Riesgo Moral	8
1.4. Modelo de Bellman	11
2. Bases de datos	13
2.1. Variables agregadas	16
2.2. Imputación de datos	17
3. Estadísticas descriptivas	19
3.1. Compensaciones salariales	21
3.2. Estadísticas deportivas	23
3.2.1. Bateadores	23
3.2.2. Fildeadores	25
3.3. Resumen de la sección	28
4. Modelo empírico	29
4.1. Implementación del modelo empírico	30
4.2. Casos de interés	32
5. Análisis econométrico	33
5.0.1. Identificación	33
5.1. Estimaciones conjuntas	37
5.2. Efectos de la edad en el poder de negociación	42
6. Efectos de convertirse en agente libre	45
7. Otros resultados	46
7.1. Efectos del COVID-19	46
7.2. Análisis por componentes principales	47
7.2.1. Diferencias-en-Diferencias	48
8. Conclusión	49
Apéndice	VI
A. Equipos	VI
B. Posiciones de jugadores	VII
B.1. Fildeadores	VII
B.2. Bateadores	IX

C. Medidas de desempeño	X
C.1. Medidas de desempeño deportivas	X
C.1.1. Bateadores	XII
C.1.2. Fildeadores	XIV
C.1.3. Definiciones misceláneas	XVI
D. Variables salariales	XVII
E. Estadísticas descriptivas	XIX
E.1. Diagramas de caja	XIX
E.1.1. Estadísticas deportivas acumuladas - Bateadores	XIX
E.1.2. Estadísticas deportivas acumuladas - Filderos	XX
E.1.3. Datos contractuales - Bateadores	XX
E.1.4. Datos contractuales - Filderos	XXI
E.1.5. Datos generales - Bateadores	XXII
E.1.6. Datos generales - Filderos	XXII
E.2. Diagramas de densidad	XXIII
E.2.1. Datos contactuales - Bateadores	XXIII
E.2.2. Datos contactuales - Filderos	XXIV
E.2.3. Datos generales - Bateadores	XXV
E.2.4. Datos generales - Filderos	XXVI
E.3. Series de tiempo	XXVII
E.3.1. Estadísticas deportivas acumuladas - Bateadores	XXVII
E.3.2. Estadísticas deportivas acumuladas - Filderos	XXVIII
E.3.3. Estadísticas deportivas por juego - Filderos	XXIX
E.3.4. Estadísticas deportivas por cada 10 juegos - Bateadores	XXX
E.3.5. Estadísticas deportivas por cada 10 juegos - Filderos	XXXI
E.3.6. Salarios - Bateadores	XXXII
E.3.7. Salarios - Filderos	XXXII
F. Problema de identificación	XXXIII
F.1. Estimaciones individuales	XXXIV
F.1.1. Bateadores	XXXIV
F.1.2. Lanzadores iniciales	XXXVIII
G. Estimaciones conjuntas	XLIII
H. Efectos de la edad a través del tiempo	XLIV
I. Estimaciones del periodo de transición a agente libre	XLIV
I.1. Bateadores	XLIV
I.2. Lanzadores iniciales	XLIV
J. Análisis por PCA	XLV
J.1. Bateadores	XLV
J.2. Lanzadores iniciales	XLVI

K. COVID-19	XLVII
K.1. Bateadores	XLVII
K.2. Lanzadores iniciales	XLVII
L. Diferencias-en-Diferencias	XLVIII
L.1. Bateadores	XLVIII
L.2. Lanzadores iniciales	XLIX

Índice de figuras

1.	Audiencia por juego entre el 2019 y el 2021 de distintas ligas mayores profesionales. Fuente: https://www.statista.com/statistics y elaboración propia.	1
2.	Valores promedio para los contratos ofrecidos en 2023 para distintas ligas profesionales en distintos deportes medidos en millones de dolares [M\$]. Se muestran los valores del punto medio de estos rangos. Fuente: https://www.spotrtrac.com/ y elaboración propia.	2
3.	Configuración usual de las posiciones de los jugadores en un campo de beisbol (MLB (2023)).	6
4.	Representación secuencial de la interacción agente-principal sin supervisión (información asimétrica) con respecto al esfuerzo, e , que implementa el agente.	8
5.	Configuración usual de las posiciones de los jugadores en un campo de beisbol. MLB (2023)	VIII

Índice de cuadros

1.	Cantidad de bases de datos por categoría.	16
2.	En este cuadro se muestran los efectos de la pandemia en los promedios anuales de las estadísticas deportivas de los bateadores y fildeadores de la <i>MLB</i> . Las siglas <i>FA</i> denota la agencia libre, mientras <i>NFA</i> indica a los jugadores que no pertenecen a dicha agencia.	19
3.	Diagramas de cada de los promedios de los logaritmos del sueldo ajustado para agentes libres y no-agentes.	22
4.	Diagramas de caja de algunas de las medidas de desempeño de los bateadores.	24
5.	Diagramas de caja de algunas de las medidas de desempeño de los lanzadores iniciales.	26
6.	Resultados de las estimaciones individuales del problema de indentificación para los bateadores. En caso de la estadística deportiva asociada a X_{var} haya sido significativa para el modelo, se indica el periodo de la variable que lo fue (pueden ser ambos periodos).	36
7.	Resultados de las estimaciones individuales del problema de indentificación para los lanzadores iniciales. En caso de la estadística deportiva asociada a X_{var} haya sido significativa para el modelo, se indica el periodo de la variable que lo fue (pueden ser ambos periodos).	37
8.	Estimaciones conjuntas con las medidas de desempeño de los bateadores para los estimadores <i>pooling</i> , <i>within</i> , efectos aleatorios (RE) y primeras-diferencias (FD).	39
10.	Estimadores de las regresiones para ambos periodos de los bateadores. Se muestra la estadística deportiva asociada a los estimadores, la dirección del cambio del estimador de la edad y el p-value de aplicar un test de Hausman a dichos periodos.	44
11.	Estimadores de las regresiones para ambos periodos de los lanzadores iniciales. Se muestra la estadística deportiva asociada a los estimadores, la dirección del cambio del estimador de la edad y el p-value de aplicar un test de Hausman a dichos periodos.	45
12.	Medidas de desempeño deportivas en las bases de datos.	XI
30.	Bateadores: Estimaciones para cada una de las estadísticas deportivas transformadas, X_{var} , independientemente del periodo. Se asumió un estimador en primeras diferencias.	XXXIV
31.	Lanzadores iniciales: Estimaciones para cada una de las estadísticas deportivas transformadas, X_{var} , independientemente del periodo. Se asumió un estimador en primeras diferencias.	XXXVIII
32.	Bateadores: Hausman Test Results	XLIII
33.	Lanzadores iniciales: Hausman Test Results	XLIII
34.	Caption	XLIV
35.	Test para cambio estructural entre periodos	XLV
36.	Bateadores regulares: Modelos con PCA	XLVI
37.	Lanzadores Iniciales: Modelos con PCA	XLVII
38.	Hausman Test Results for COVID-19 and MLB Hitters	XLVII

39.	Hausman Test Results for COVID-19 and MLB Starting Pitchers	XLVIII
-----	---	--------

1. Introducción

Los contratos deportivos profesionales son un tema de gran interés tanto en la literatura no académica (ContractsCounsel (2023), Brown (2022), Thorpe (2010)) como en la académica (Hoffman (2014), Trendafilov et al. (2017), Congdon-Hohman and Lanning (2018)) porque dichos contratos son cuantiosos en términos monetarios y en ocasiones surgen polémicas en torno a los mismos. En este contexto, la presente tesina se enfoca en analizar empíricamente la estructura de dichos contratos con el objetivo de comprender la relación causal entre la compensación de los jugadores profesionales de béisbol y diversas medidas de desempeño y otras características relevantes. Específicamente, se utiliza una metodología de econometría aplicada para obtener resultados robustos y estadísticamente significativos. Se seleccionó el béisbol profesional como caso de estudio debido a las características particulares que presentan los contratos de sus jugadores en comparación con los contratos de jugadores de otros deportes, como por ejemplo, los contratos de jugadores profesionales de béisbol suelen estar asegurados por su duración y son independientes de su desempeño posterior de dichos jugadores. La contribución de esta tesina al campo de la econometría aplicada a los deportes es realizar un análisis de tipo causal, la cual contrasta con la mayoría de trabajos afines que se enfocan en propósitos predictivos.

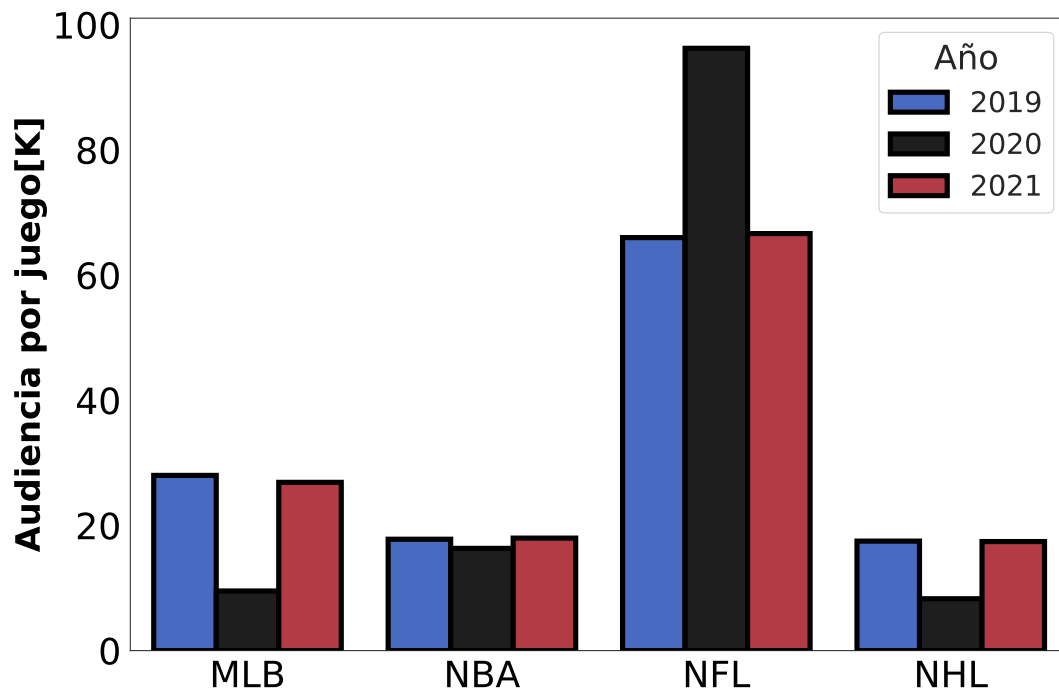


Figura 1: Audiencia por juego entre el 2019 y el 2021 de distintas ligas mayores profesionales. Fuente: <https://www.statista.com/statistics> y elaboración propia.

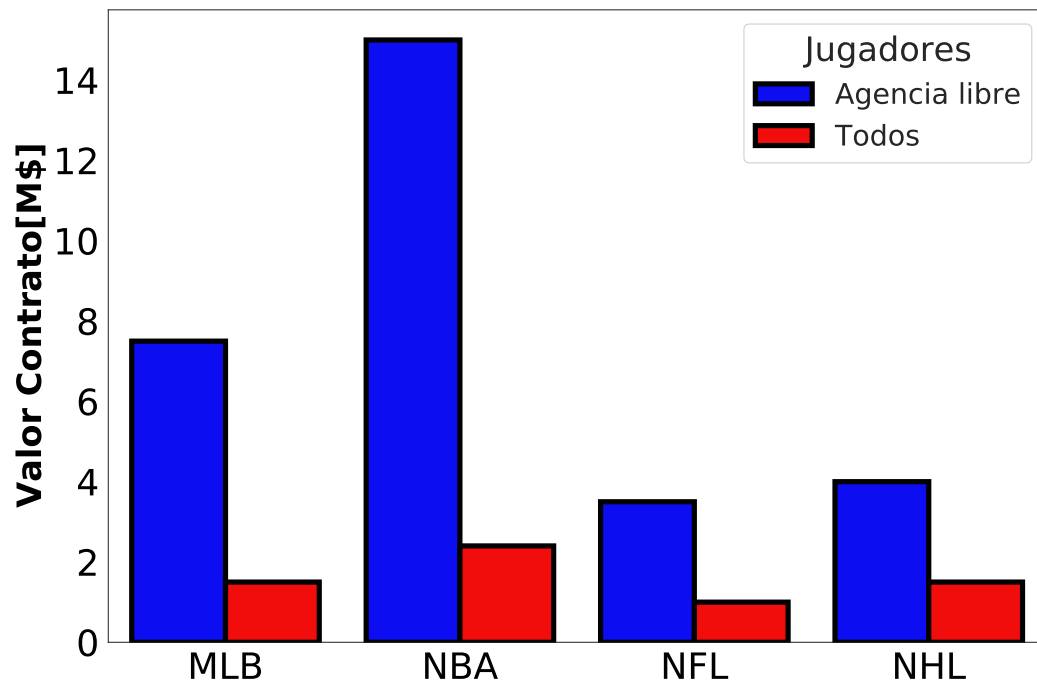


Figura 2: Valores promedio para los contratos ofrecidos en 2023 para distintas ligas profesionales en distintos deportes medidos en millones de dolares [M\$]. Se muestran los valores del punto medio de estos rangos. Fuente: <https://www.spotracc.com/> y elaboración propia.

Las *Grandes Ligas de Béisbol*, o mejor conocidas como *MLB* por las siglas en inglés de *Major League Baseball*, es la organización de béisbol más conocida en el mundo a pesar de sólo ser practicada de manera profesional en pocos países.¹ Incluso, no cuenta con un gran número de fanáticos asiduos a nivel mundial en comparación con otros deportes tales como el fútbol soccer de América (EUA), cricket o voleibol (Brown (2022), Status (2021)). Tan solo en Estados Unidos el béisbol de la *MLB* no es el más visto, siendo superado considerablemente por el fútbol americano, en particular por la *National Football League* (*NFL*), y en algunas ocasiones por el basquetbol (*National Basketball Association* con siglas *NBA*), ver Figura 1. En dicha gráfica se muestran los valores del contrato medio de los rangos promedios de contratos ofrecidos en el 2023. El real para cada deporte es 5-10[M\$], 10-20[M\$], 2-5[M\$] y 2-6[M\$], respectivamente.

No obstante, esto no ha sido impedimento para que dicho deporte haya cobrado relevancia en países como EUA donde la *MLB* ha llegado a firmar con televisoras nacionales contratos anuales con un valor de más de 1,500 millones de dolares (Brown (2022)). A pesar de que desde hace más de una década la Liga Nacional de Fútbol (*NFL*) obtiene ganancias y

¹Sólo se practica de manera oficial en EUA, Canadá y Japón.

audiencia anuales mayores a las de *MLB* ([Janvrin \(2022\)](#)), esta última ofrece contratos y sueldos con un valor agregado mayor a los que ofrece la *NFL*, solo en el año 2018 los salarios para los novatos en la industria americana del béisbol tuvieron un valor total de \$545,000, mientras que en la industria profesional de fútbol americano solo se registró un valor agregado de \$365,000 para los novatos. Al observar los gráficos de barras del promedio de audiencias por juego y el valor promedio de los contratos ofrecidos (*Figuras 1 y 2*) resulta interesante que, a pesar de que la audiencia promedio de la *MLB*, *NBA* y *NHL* son relativamente distintas entre sí, el valor promedio de los contratos ofrecidos a todos los jugadores no son muy distintos entre sí.

Más aún, resulta interesante el caso de la *NFL* pues a pesar de que tanto la audiencia como sus ganancias son mayores a los del resto de las otras ligas; incluso, los contratos que ofrecen a los jugadores que son agentes libres² son, en agregado, menores al resto de ligas. Salvo el caso del basquetbol, se aprecia que la *MLB* ofrece mejores compensaciones salariales para su agencia libre. En otras palabras, la cantidad monetaria promedio de los contratos que ofrece la *MLB* son mucho mayores a los que se ofrecen en la mayoría de deportes en EUA, independientemente de la popularidad o ganancias que estos puedan generar ([Treatment \(2022\)](#)).

Sin embargo, esta no es la única razón por la que los contratos de los agentes libres en la *MLB* son de interés cuando se les compara con los hechos observables en otros deportes. Por ejemplo, los contratos están completamente asegurados para los jugadores, esto quiere decir que no importa si en el periodo de contrato el desempeño del jugador baja o si este se lastima, el jugador recibirá su paga completa; esto no ocurre en todos los deportes ya que en la mayoría sólo se les garantiza parcialmente su salario frente a heridas o condiciona el pago del contrato al desempeño que tenga el jugador ([Heyman \(2023\)](#)). Asimismo, los contratos que se ofrecen a los agentes libres tienden a durar más años que en otros deportes; esto permite que tanto los jugadores como los equipos tengan más estabilidad de acuerdo a sus intereses ([Todd \(2017\)](#)).

Relacionado a esto último, el béisbol es uno de esos pocos deportes donde hay una imposición sobre el salario máximo (*salary cap*) que se le puede ofrecer a un jugador en este contrato donde si se supera este salario tope, el equipo tendrá que pagar impuestos extras para poder ofrecer dicho sueldo. Esto tiene como propósito hacer más equitativa la distribución de talento entre los equipos ya que equipos con mucho poder adquisitivo podrán ofrecer salarios más tentadores en comparación con el resto por lo que captarían una mayor cantidad de talento³ ([Blum \(2019\)](#)). Por último, se sabe que el desempeño de un jugador durante el partido tiene un mayor impacto para la victoria de su equipo en comparación con otros deportes que se juegue por equipos, lo cual se refleja en el tamaño del contrato que reciben los jugadores que contribuyan mayormente a la victoria del equipo ([Paine \(2023\)](#)). La razón es que en el béisbol aísla las acciones individuales. Por ejemplo, cuando el bateador intenta conectar un *hit* con la bola lanzada por el lanzador;

²Aquellos que pueden decir con qué equipo negociar su contrato -más adelante se provee una definición exacta.

³Esta medida se implementó a finales de los años 90. Esta medida perjudicó a los *New York Yankees* quienes tenían una enorme ventaja de Series Mundiales ganadas.

para batear o marcar un *strike* no se necesitan más jugadores que los antes mencionados, donde las acciones de estos dos pueden desencadenar una serie de eventos que tengan un gran impacto en el resultado del equipo. Por lo general, los *Most Value Players* (*MVP*) tienden a poseer contratos cuyo valor es mayor al que se le ofrecen a los *mvp* de otros deportes (Paine (2023)).

Entonces, el tema de los contratos en la *MLB* sea bastante relevante en esta industria porque hay una tendencia que entre mejor sean los resultados del equipo tales como las victorias, mayor sea el valor promedio de los contratos que ofrece.⁴ Como consecuencia, se hacen análisis exhaustivos ya sea para mejorar el desempeño del equipo o jugadores, ajustar tácticas de acuerdo al equipo contrincante o realizar apuestas basadas en el desempeño tanto de manera agregada como individual. Por lo tanto, es importante entender qué factores influyen en la cantidad monetaria de los contratos que se le ofrecen a los jugadores. Relacionado con esto, desde hace poco más de una década se ha intensificado la búsqueda de modelos que puedan ofrecer un entendimiento sobre cuáles variables influyen en los sueldos de los jugadores de la *MLB*, así como modelos para predecirlos.

Hay trabajos que han empleado técnicas sofisticadas tales como *Machine Learning* y *Computer Vision* (Connor Heaton (2022), Hoffman (2014)). Sin embargo, estos no tienen como propósito desarrollar ningún análisis causal, sino predecir el resultado de un partido u ofrecer una clasificación más especializada de los jugadores de acuerdo a su desempeño en el juego. Por otro lado, uno de los artículos más famosos sobre aplicaciones de econometría en el béisbol (Congdon-Hohman and Lanning (2018)) ofrece un análisis con un enfoque más centrado a los salarios. No obstante, este se enfoca en mostrar evidencia de que hay una relación positiva entre los incentivos en las compensaciones salariales de los jugadores (incluyendo agentes libres) y su desempeño, es decir, entre mayor sea su compensación salarial, mayor es la probabilidad de que incremente sus estadísticas deportivas promedio tales como los *home runs* o dobles (*doubles*). Posteriormente, se escribió otro trabajo relacionado al béisbol (Hoffman (2014)) donde se trata de hallar qué variables pueden predecir mejor el salario de algún jugador en temporadas posteriores. A pesar de obtener buenos resultados en las predicciones,⁵ este trabajo no puede considerarse como un análisis causal.

En resumen, se han escrito artículos relacionados con los salarios y la agencia libre. Sin embargo, no son trabajos que hagan un análisis causal sobre las diferencias salariales entre jugadores que son agentes libres y quienes no lo son. Más aún, no consideran el problema de riesgo moral (*moral hazard*). El único trabajo empírico que se ha realizado para estimar el poder de negociación es aplicado al hockey, aunque no analiza la dinámica de esta variable ni obtiene resultados robustos (Navin and Sullivan (2002)). Esta es la motivación principal por la que se realiza el presente trabajo; es decir, hacer un análisis causal con el objetivo de explicar la diferencia entre los salarios de los agentes libres en la *MLB* de manera indirecta a través de la dinámica del cambio en el poder de negociación. En particular, nuestra pregunta de investigación concierne a la identificación del cambio del poder de negociación (*bargaining power*) de los agentes libres. Dicha variable de estado

⁴<https://www.sportrac.com/mlb/>

⁵Las predicciones del modelo tenían un *score* de $R^2 = 0,7$.

se estimará de manera indirecta a través de variables relacionadas con el desempeño deportivo, además de las compensaciones. No obstante, cabe señalar que este trabajo también sirve como un marco a partir del cual determinar que variables del desempeño impactan más en la cantidad de dinero que el agente libre percibe en su contrato. Más aún, el presente análisis se puede aplicar a otros deportes que no sean béisbol tales como el fútbol americano, basquetbol o hockey.⁶ Sin embargo, este trabajo no debe entenderse como un análisis contrafactual entre jugadores que son agentes libres y quienes no lo son, sino como análisis aplicado a un subconjunto de jugadores que pertenecen a la agencia libre de la *MLB* a partir de los resultados teóricos y numéricos obtenidos al resolver un problema de riesgo moral.

1.1. Dinámica de un partido de béisbol

Para poder obtener un mayor entendimiento e intuición sobre las medidas de desempeño que se emplearán en el modelo empírico es necesario entender las reglas que rigen un partido béisbol. Por otro lado, en el apéndice se tiene una sección especializada para entender las funciones que se desempeñan en cada una de las posiciones que algún jugador puede ocupar en un partido tanto en la ofensiva como en la defensiva. Cabe señalar que muchos de los tecnicismos del béisbol nombrados en esta y demás secciones se encuentran explicados de forma detallada en el apéndice.

Un partido de béisbol se lleva a cabo en un campo de juego como el que se muestra en la *Figura 3*. Cada juego se conforma de nueve entradas (*innings*) con la particularidad que no hay ninguna imposición de tiempo para cada una de estas. Por lo tanto, cada partido puede durar una cantidad indeterminada de tiempo.⁷ Las entradas están conformadas por dos mitades: Alta y baja. En la parte alta juega a la ofensiva el equipo visitante; mientras que en la baja, el local. Para que una mitad concluya, y se pase a la siguiente, el equipo defensivo tiene que conseguir hacer tres *outs*. Cada *out* define un tercio de la entrada. De manera similar a otros deportes, también se cuenta con eliminatorias específicas (*playoffs*) para otorgar a algún equipo el título de campeón mundial (La Serie Mundial conocida en inglés como *World Series*). Los *playoffs* se llevan a cabo una vez la temporada regular haya finalizado. La importancia de la temporada regular radica en que a partir del desempeño y victorias de los equipos se seleccionan los mejores para competir en los *playoffs*. Estas eliminatorias constan de cuatro etapas donde en la última etapa quedan solo dos equipos para disputarse la serie mundial.

Hasta ahora, hemos ofrecido un panorama general del béisbol, pero aún no hemos explicado cómo se juega ni cómo se sabe qué equipo gana o no en un partido. Para determinar cuál equipo es el ganador se cuentan las carreras (*runs*) que cada uno anotó a lo largo del juego. Entonces, el equipo que haya conseguido hacer más carreras a lo largo del partido es declarado ganador; en caso de que en los nueve *innings* se haya llegado a un empate, se realiza un *inning* extra hasta que se logre el desempate ([Official Playing Rules Committee \(2021\)](#), [Baseball Rules \(2023\)](#)).

⁶Actualmente se cuenta con páginas como [Spotrac](#) o [Sportsdata](#) para conseguir las bases de datos correspondientes.

⁷En promedio, los partidos tienden a durar alrededor de tres horas.

Entonces, lo que sigue es explicar cómo se anota una carrera (*run*).⁸ Para que se pueda considerar que un jugador ofensivo anotó una carrera, este tiene que pasar por las cuatro bases; es decir, tiene que recorrer el cuadro (diamante) que se encuentra dentro del *infield* en la *Figura 3* (el cuadro limitado por los números 2, 3, 4 y 5 -estos número indican la posición del jugador). Un ofensivo llega a la primera base (donde se encuentra el jugador llamado *primera base* indicado por el número 3) si logra conseguir hacer una bola justa (*fair ball*), antes de que lo saquen por *strike out*, e inmediatamente correr hasta dicha base sin que le marquen un *out*. Después, para que este pueda llegar a la siguiente base (sin que lo saquen por *out*), dependerá de que otro bateador haga otro *hit* en territorio válido, y así consecutivamente hasta llega a la última base (situada enfrente de la caja de bateo).

⁸Nos limitaremos a una explicación general de cómo se anota una carrera (*run*) Sin embargo, los casos específicos que involucran errores del equipo defensivo (*errors*) se abordan en el apéndice, así como las estadísticas deportivas que toman en cuenta estos factores.

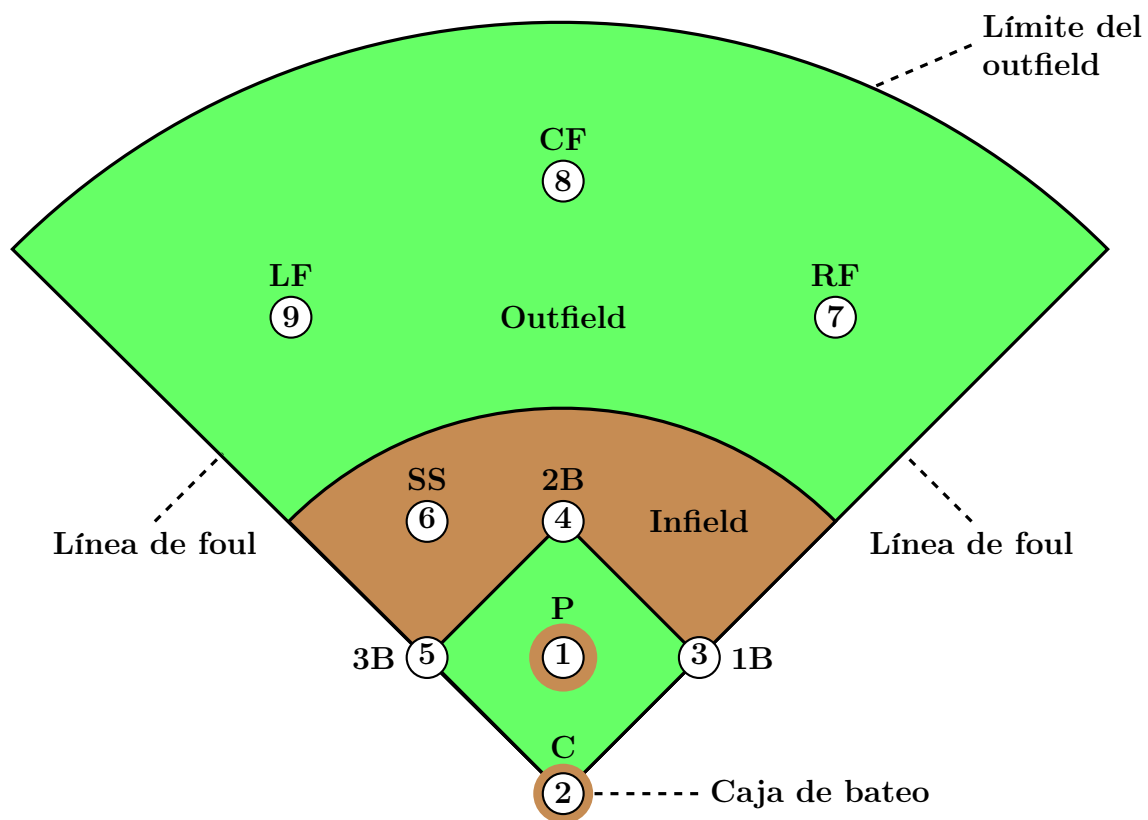


Figura 3: Configuración usual de las posiciones de los jugadores en un campo de béisbol (MLB (2023)).

1.2. Problema de Riesgo Moral

Este apartado tiene como propósito dar las bases teóricas para entender cómo se construyen los modelos de información asimétrica utilizados en el presente análisis, así como la interpretación de las ecuaciones utilizadas en los problemas de optimización que estos modelos implican. Debido a que uno de nuestros objetivos en esta sección es que el lector tenga un entendimiento sólido de los modelos es que partiremos desde el modelo más básico para entender las ideas generales del problema de riesgo moral, para pasar paulatinamente por un modelo más general antes de definir el modelo principal.

Sin embargo, antes de adentrarnos a los modelos empezamos definiendo qué es un agente libre tanto en el ámbito de la *MLB*, como en el contexto de riesgo moral que surge y si dichos conceptos son compatibles. Los requisitos que tiene que cumplir un jugador de la *MLB* para poder ser considerado como un agente libre depende de satisfacer cierta cantidad de años de servicio. Cuando un jugador empieza a trabajar de manera profesional en la *MLB*, él no puede escoger en qué equipo entrar, sino que es asignado a un equipo. Entonces, una vez este jugador haya llegado a los seis años de servicio en la *MLB* o de su organización, el jugador puede negociar con distintos equipos con el objetivo de ser contratado como jugador por este, y si ambas partes están de acuerdo en firmar, entonces se lleva a cabo el contrato como agente libre (*Free Agency Definition* (2023)). Por otro lado, en el contexto de riesgo moral, un agente se define como aquel individuo a quien se le delega, por parte de un principal, el realizar alguna tarea o tomar decisiones, las cuales son en representación de dicho principal (por ejemplo, un dueño de una empresa); esta interacción agente-principal se conoce como una relación de agencia (*Gardner* (1995)) puesto que los beneficios del principal dependen o se ven afectados por el esfuerzo que el agente realiza para llevar a cabo dichas tareas. Por lo tanto, un jugador de la *MLB* se puede ver como aquella persona a quien se delega la tarea de contribuir a la victoria del equipo por el cual es contratado y en representación de este.

A pesar de que se puede observar el desempeño de un jugador a lo largo de un partido, el principal no conoce el verdadero esfuerzo que este realiza para contribuir en la victoria del equipo; por ejemplo, si dicho jugador ha entrenado lo suficiente tanto para entender y ejecutar apropiadamente las jugadas como para estar físicamente apto para las exigencias que requieren la posición que ocupa. En otras palabras, hay una asimetría de información en esta relación agente-principal ya que el principal observa parcialmente el esfuerzo del agente, pero el jugador sí puede conocerlo directamente, es decir, el jugador posee información que el principal no. Como la asimetría de la información en la relación agente-principal se presenta después de haber concretado el contrato, se está ante un problema de riesgo moral (*moral hazard*). A pesar de que no se puede cambiar el hecho de que es imposible supervisar totalmente el esfuerzo del jugador una vez contratado, esta relación agente-principal se acentúa aún más cuando el jugador pasa a ser un agente libre puesto que los costos de agencia crecen significativamente. Esto se debe a que los salarios ofrecidos a los agentes son en promedio mucho mayores a los de un jugador que no es agente, se puede apreciar en los diagramas de caja y gráficas de la sección E. Esto implica que el equipo tiene que optar por obtener menores ingresos con tal de pagar los salarios de los agentes libres.

1.3. Modelo de Riesgo Moral

En el caso de los agentes libres de la *MLB*, estos reciben un salario fijo, salvo bonos,⁹ a largo de las temporadas por las que son contratados, estos periodos pueden ser desde una temporada hasta trece temporadas, donde cada temporada empieza a inicios de la primavera y termina en otoño. El hecho de que los agentes libres obtengan un salario fijo (Heyman (2023)) independientemente de su desempeño e incapacitaciones por lesiones, puede generar dos tipos de incentivos sobre el desempeño del jugador, los cuales pueden ser negativos ó positivos.¹⁰ Tampoco se puede ignorar el hecho de que también hay incentivos externos al contrato; por ejemplo, si el jugador no planea continuar con algún contrato posterior al que ya están, no habrá un gran incentivo para que el jugador se esfuerce

⁹En las bases de datos usadas para este trabajo, los bonos no representan más del 10 % del salario fijo del agente libre de su salario en nómina.

¹⁰Una posible consecuencia de este hecho es que conforme pasan las temporadas durante el periodo de contratación, las variables correspondientes al desempeño del jugador podrían perder significancia estadística para explicar los salarios. No obstante, se verificará si este efecto está presente en los periodo de análisis que abarca la base de datos.

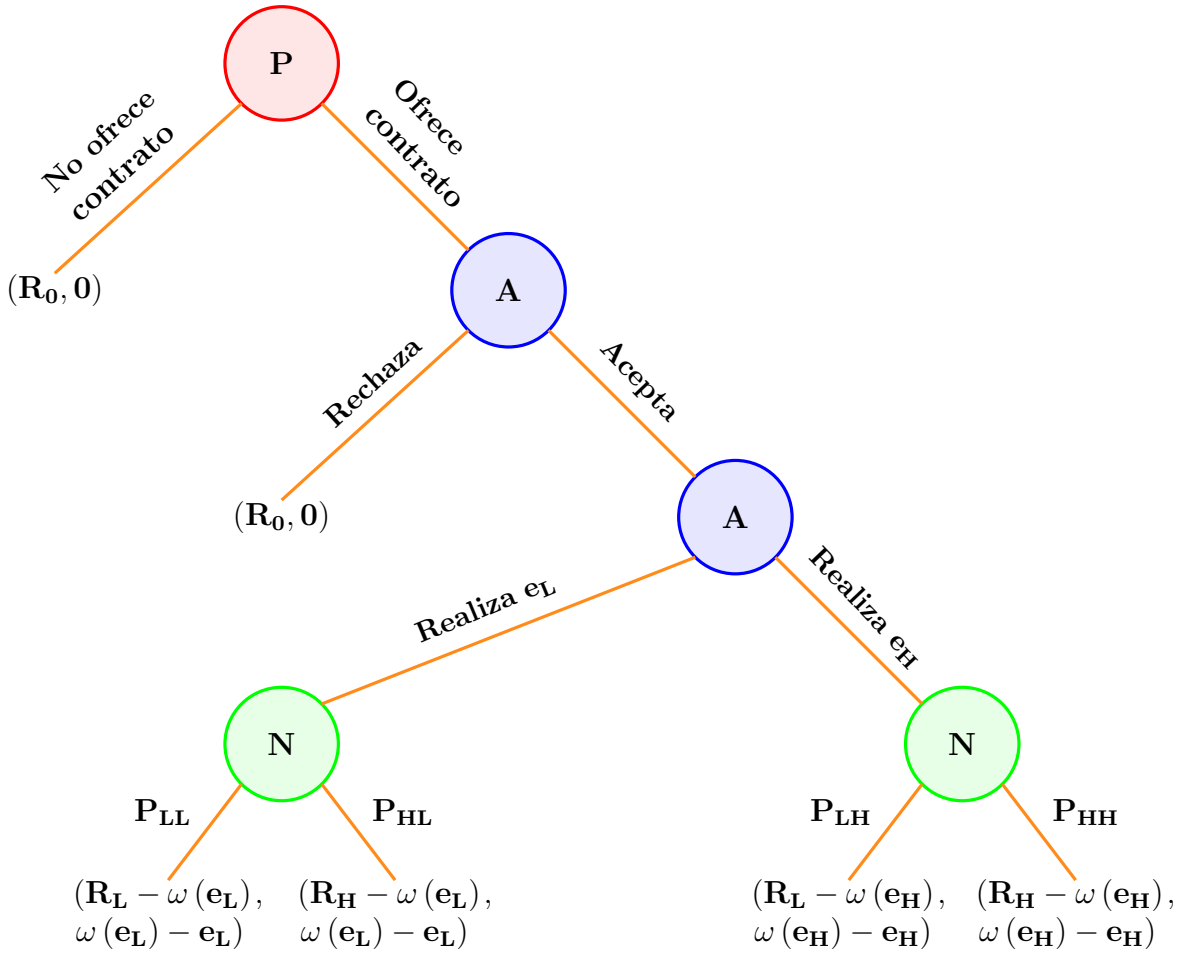


Figura 4: Representación secuencial de la interacción agente-principal sin supervisión (información asimétrica) con respecto al esfuerzo, e , que implementa el agente.

bajo el contrato actual. Por lo tanto, los principales tienen que diseñar un contrato que incentive la conducta deseada en el agente a través de esquemas de compensación salarial donde se busca condicionar el sueldo del agente en función de su desempeño, pero también se toma en cuenta que pueden haber factores externos al agente que afectan sus resultados (*output*) tales como el desempeño del resto del equipo o el clima que está presente durante el partido.

Con el propósito de obtener una intuición general sobre el problema de riesgo moral, consideremos uno de los casos más sencillos de la interacción entre un principal y un agente libre donde se asume que no-supervisión, es decir, el principal no puede monitorear al agente. Esta interacción agente-principal está representada por el juego secuencial en el diagrama (1.2) (Gardner (1995)). Consideremos a un agente que posee una función de utilidad, $u(\cdot)$, que depende tanto de su salario, $\omega(\cdot)$, como del esfuerzo, e , de la siguiente forma: $u(\omega, e) = \omega(e) - e$. Adicionalmente, impondremos los siguientes supuestos sobre las derivadas parciales para que haya mayor realismo: $u_\omega > 0$ y $u_{e_L} - u_{e_H} < 0$ ($u_{e_H} > u_{e_L}$), donde $e_H > e_L$. Estos supuestos nos dicen que entre mayor sea la paga del agente, mayor será su utilidad; mientras que el implementar esfuerzos más grandes implica una disminución no monótona en utilidad. De manera análoga, el principal tiene asociada una utilidad que está en función de las ganancias, $R(\cdot)$ (*output*), y el sueldo del agente, ω . La función de utilidad del principal (beneficios monetarios) se define como $g(R(e), \omega(e)) = R(e) - \omega(e)$ con $g_R > 0$ y $g_\omega < 0$. Similar al caso del agente, la función de utilidad del principal crece conforme mayores sean las ganancias generadas para la empresa, pero menor será la utilidad entre mayor sea la compensación salarial que le tenga que dar al agente.¹¹ Ahora, podemos proceder a describir cómo se lleva a cabo esta interacción descrita mediante un juego secuencial, el cual se resuelve mediante inducción hacia atrás.

La interacción agente-principal descrita por la *Figura 1.2* es una situación de información asimétrica debido a que el principal no conoce el esfuerzo, e , que implementa el agente, pero este último si conoce su esfuerzo. Como consecuencia, el principal tiene que pensar en un esquema de incentivos para que el agente implemente el esfuerzo óptimo que le produzca la mayor ganancia o beneficio que la empresa pueda obtener. Entonces, una solución es que el principal condicione el salario del agente de acuerdo al *output* que este genere. No obstante, estamos suponiendo implícitamente un proceso estocástico en esta interacción puesto que el hecho de que el agente haga un esfuerzo alto, no implica que el *output* será grande (Π_H), sino que existe de una probabilidad P_{LH} de que el *output* sea bajo, y una probabilidad P_{HH} de que sea alto.¹²

Por otro lado, cabe aclarar que el esquema de compensación salarial es independiente del proceso estocástico subyacente, es decir, el principal le pagará al agente de acuerdo al *output* generado y no con base en su esfuerzo. Basta considerar el caso donde el esfuerzo e_i produce un *output* R_i con probabilidad 1, pero sin supervisión; entonces persiste el problema de que el principal no puede observar el esfuerzo del agente y por lo tanto tendrá que condicionar su salario de acuerdo al *output* para que sus preferencias estén

¹¹Tenemos un conjunto finito de esfuerzos donde solo hay dos niveles de esfuerzo: Bajo (e_L) y alto (e_H)

¹²Se usará la notación P_{ij} , con i no necesariamente igual a j . El subíndice i denota si el *output* fue bajo o alto ($\{R_L, R_H\}$); mientras que el subíndice j denota si el esfuerzo realizado fue bajo o alto ($\{e_L, e_H\}$).

alineadas. Sin embargo, si se impondrá una condición para que dado un esfuerzo e_i la probabilidad de que se genere un *output* R_i sea mayor a la de obtener un *output* R_j con $j \neq i$ para asegurar que los intereses del principal y del agente estén en la misma dirección.

Hasta aquí se ha descrito toda la maquinaria matemática, más no la intuición para entender la relación principal-agente. Esto se debe a que se prefirió dar las condiciones matemáticas bajo las cuáles se trata de ofrecer un modelo más simple y aproximado al que se observa usualmente en la realidad. En lugar de describir cómo se soluciona este modelo, describiremos cómo se lleva a cabo esta relación (*Figura 1.2*) para obtener una intuición general sobre cómo se resuelve el modelo de riesgo moral; más adelante se generalizará este ejemplo, pero es importante entender el caso discreto para poder generalizar y entender el modelo continuo.

De manera general, esta relación empieza cuando el principal decide si ofrecer o no un contrato a un agente puesto que esto le representaría beneficios positivos al principal (o al menos no negativos); en caso de no ofrecer el contrato, el agente no recibe ninguna paga (salario $\omega = 0$) y el principal se queda con sus ganancias iniciales R_0 . En caso de que ofrezca el contrato, saltamos a la siguiente etapa del juego donde el agente decide si acepta o rechaza el contrato cuyas consecuencias monetarias en caso de no aceptar son las mismas que en la primera etapa. Como el contratar al agente puede representar un aumento en sus beneficios, entonces el principal le sería conveniente que el agente acepte el contrato (*Restricción Participación del Agente*). Luego, en caso de aceptar el contrato, ahora lo que el agente decide en esta etapa es decir qué esfuerzo, e , implementar, en este caso, el principal tiene que pagarle de acuerdo al *output* que este produzca de tal manera que induzca al agente a que realice el esfuerzo óptimo para la empresa (*Compatibilidad de Incentivos*)

La relación agente-principal que se describió arriba se puede generalizar con el modelo de acción oculta con un continuo de esfuerzos (Mas-Colell et al. (1995)). Este nuevo problema de riesgo moral está descrito por el siguiente problema de minimización

$$\min_{\omega(\pi)} \int \omega(\pi) f(\pi | e) d\pi \quad (1)$$

$$\text{s.t.} \quad \int \nu(\omega(\pi)) f(\pi | e) d\pi - g(e) \geq \hat{u} \quad (2)$$

$$\max_e \int \nu(\omega(\pi)) f(\pi | e) d\pi - g(e) \quad (3)$$

donde $\pi \in [\underline{\pi}, \bar{\pi}]$ y $e \in [e_L, e_H]$. Tenemos que la ecuación (2) es la *Restricción de Participación del Agente* donde \hat{u} es la variable de estado que representa la utilidad de reserva (costo de oportunidad) del agente libre; mientras que la restricción (3) es la *Compatibilidad de Incentivos* a la cual se le impondrá la *Propiedad de Probabilidad Monótona* con el objetivo de asegurar que el salario óptimo del agente sea monótonamente creciente con respecto a los beneficios del principal, π . Para el caso finito, esta última restricción se puede reescribir de la siguiente forma dependiendo del esfuerzo que el principal desee que

se implemente

$$\int (\nu(\omega(\pi)) - g(e_i)) f(\pi | e_i) d\pi > \int (\nu(\omega(\pi)) - g(e_j)) f(\pi | e_j) d\pi \quad \forall i \neq j \quad (4)$$

Notemos que todas las integrales que están escritas en el problema de optimización (1) describen valores esperados, pero para un continuo de estados. Las restricciones de *participación del agente* y de la *compatibilidad de incentivos* también aparecen en la solución del juego secuencial que analizamos al inicio, dado por el diagrama *Figura 1.2*, solo que en nuestro primer ejemplo el conjunto de estados de la naturaleza eran discretos, así como el esfuerzo que podría realizar el agente.

Algo que resalta a primera vista es que a diferencia del primer modelo de acción oculta que vimos, ahora suponemos que el espacio al que pertenece el esfuerzo y el *output* son un continuos, dados por $[e_L, e_H]$ y $\pi \in [\underline{\pi}, \bar{\pi}]$, respectivamente. Como consecuencia, las restricciones del problema de optimización se representan promedio del valor esperado por expresado como una integral y no una suma como en el caso discreto. Así mismo, también cambia la forma de resolver el modelo de riesgo moral, puesto que ahora se define como un problema de optimización donde las restricciones de participación y compatibilidad de incentivos están representadas por las restricciones (2,3). La ventaja del problema de riesgo moral descrito por el problema de optimización dado por (1) es que ya no nos limitamos a espacios discretos, así como que las funciones de costos y las que dependen del *output* no necesariamente tienen que ser lineales, basta con que cumplan con las propiedades de monotonicidad impuestas en la descripción matemática del caso discreto.

Por lo tanto, ahora se aprecia la generalización del problema agente-principal descrito por el juego del diagrama (*Figura 1.2*), así como la relación con este. Lo que nos deja en claro estos problemas es que se pueden mejorar cada vez más los modelos, así como hacerlos más generales o cercanos a la realidad. Sin embargo, tal y como es de suponerse, hoy en día existen más modelos, ya sea especializados a contratos lineales ([Holmstrom and Milgrom \(1987\)](#)) con el propósito de obtener propiedades que se puedan extrapolar; o modelos con compensaciones salariales que no sea vean solo afectadas por el esfuerzo del agente y el *output*, sino también por otras variables de estado.

1.4. Modelo de Bellman

Hasta el momento, solo nos hemos limitado a que la compensación salarial del agente libre dependa del nivel de esfuerzo que este implemente, e con el objetivo de entender las ideas subyacentes en la solución de problemas de riesgo moral y la intuición sobre las restricciones impuestas en el problema de optimización. Sin embargo, dicho modelo de acción oculta presenta otra limitación, aparte de las mencionadas anteriormente, y es que la variable de estado del agente libre se limita a su utilidad de reserva, \hat{u} , pero esto se puede extender a otras variables tales como el poder de negociación, δ , la cual es nuestro objeto de interés. Por lo tanto, vale la pena mostrar la ecuación de Bellman para entender de dónde deriva el modelo econométrico que se usará para el análisis causal del poder de

negociación, δ . Análogo al modelo de acción oculta (1), supondremos un principal y agente aversos al riesgo con los mismas propiedades sobre sus derivadas parciales con respecto a la compensación al esfuerzo implementad, e , y la compensación salarial que recibe el agente, ω .

Este modelo analiza la dinámica del cambio en el poder de negociación donde considera la interacción entre el agente y el principal para periodos discretos; se asume que esta interacción inicia en t_0 y nunca se termina, es decir, para un periodo de contratación infinito. El cambio con respecto al primer modelo de acción oculta dado por el problema de optimización (1) es que en el modelo de Bellman se considera en el periodo t toda la historia previa de los poderes de negociación ($\delta_{t'}$) y *outputs* previos ($\pi_{t'}$), es decir, para $t' < t$. Además, ya no se toma a la utilidad de reserva del agente como variable de estado, sino a su poder de negociación, δ , (DiGiannatale et al (2023)). Antes de definir el problema de optimización, definamos las utilidades $U(\delta)$ del principal y $V(\delta)$ del agente:

$$\begin{aligned} U(\delta) &= \int_{\Pi} [\pi - \omega(\delta, \pi) + \beta \bar{U}(\delta, \pi)] f(\pi; e^*(\delta)) dy, \\ V(\delta) &= \int_{\Pi} [v(\omega(\delta, \pi), e^*(\delta)) + \beta \bar{V}(\delta, \pi)] f(y; e^*(\delta)) d\pi. \end{aligned}$$

Vemos que estas utilidades ya incorporan el poder de negociación (δ) como variable además del *output* (y), esfuerzo (e) y compensación del agente (e). Estas utilidades representan las utilidades esperadas descontadas del principal y del agente, esto significa que la utilidad esperada de los resultados futuros se reduce por un factor de descuento, que refleja la preferencia de estos individuos por el consumo inmediato sobre el consumo futuro. Así mismo, también se observa que estas utilidades son separables con respecto a algunos de sus argumentos. Por otro lado, las distribuciones de probabilidad se asumen invariantes a lo largo del periodo de contratación. No obstante, el problema anterior se puede generalizar, es decir, equivalente al problema de optimización multi-objetivo con el poder de negociación como variable de estado que se muestra a continuación

$$\Gamma(U, V)(\delta) = \max_{\omega(\delta, \pi), \bar{V}(\delta, \pi), \bar{U}(\delta, \pi)} \{U(\delta), V(\delta)\} \quad (5)$$

$$\begin{aligned} \text{s.a. } e^*(\delta) &\in \operatorname{argmax}_{e(\delta) \in A} \int_{\Pi} [v(\omega(\delta, \pi), e(\delta)) \\ &+ \beta \bar{V}(\delta, \pi)] f(\pi : e(\delta)) d\pi \end{aligned} \quad (6)$$

$$0 \leq \omega(\delta, \pi) \leq \pi \quad \forall \pi \in \Pi, \quad (7)$$

$$\delta' = z(\delta, \pi) \in [0, 1], \quad (8)$$

$$(\bar{U}(\delta, \pi), \bar{V}(\delta, \pi)) \in \mathcal{W}(\delta') \quad \forall \pi \in \Pi \quad (9)$$

Γ se define como un operador que mapea del producto cartesiano de dos espacios de funciones continuas y acotadas¹³ hacia sí mismo bajo la norma del supremo. Mientras que para cada δ' $\mathcal{W}(\delta')$ se define como el conjunto de valores de las utilidades esperadas descontadas del principal y agentes generados por contratos que son factibles. Por otro lado, las restricciones son similares a las que se han mencionado en otros modelos. Análogo al

¹³Las funciones están acotadas puesto que las posibles recompensas monetarias lo están.

primer modelo de acción oculta, la restricción (6) impone la compatibilidad de incentivos; para asegurar que el agente no obtenga un salario negativo o superior a los ingresos del principal es que se impone la restricción (7); (8) restringe al poder de negociación, δ' , a pertenecer en el intervalo $[0, 1]$; por último, la restricción dada por (9) garantiza que tanto la utilidad futura del principal y del agente sean factibles. La solución de este problema es única para cada poder de negociación inicial y Pareto eficiente puesto que no se puede obtener una mejor situación con respecto a los salarios que perciben el principal y el agente. Cabe señalar que se utilizarán los resultados de este modelo para definir el modelo empírico que se utilizará para la identificación del poder de negociación, así como sus cambios entre contratos; se abordará más adelante en la sección destinada al modelo empírico.

2. Bases de datos

Para el análisis y modelos desarrollados se emplearon inicialmente 120 bases de datos de corte transversal, las cuales contiene observaciones que comprenden desde el año 2011 hasta el 2022 (12 años de observaciones). Estas bases de datos se obtuvieron de la versión premium de *Spotrac*.¹⁴ Las bases de datos extraídas se dividen en cuatro categorías: Agentes libres, salarios, bateadores y fildeadores. No importa el año de la base de datos, cada una tiene las mismas variables en su respectiva categoría. Las variables que contiene cada categoría (con la respectiva abreviatura usada en *Spotrac* para referirse a la variable en cuestión) son las que se muestran a continuación:

- Agentes libres

- **YEAR:** Año de la observación.
- **POS:** Posición del jugador.
- **STATUS:** Estado de agente libre.
- **TEAM FROM:** Equipo del que proviene.
- **TEAM FROM TO:** Equipo al que pertenece en el año de la observación.
- **YRS:** Años de duración del contrato.
- **VALUE:** Valor del contrato.
- **AVV:** Valor promedio del contrato.

- Salarios:

- **YEAR:** Año de la observación.
- **POS:** Posición del jugador.
- **TEAM:** Equipo al que pertenece el jugador.
- **BASE SALARY:** Salario base del jugador.
- **SIGNING BONUS:** Bono por firmar en la *MLB*.

¹⁴[Spotrac.com](https://www.spotrac.com)

- **PAYROLL SALARY**: Sueldo de nómina del jugador.
- **ADJ SALARY**: Valor del contrato.
- **SALARY %**: Porcentaje que respresenta el salario del año en curso.
- **CASH**: Dinero en efectivo pagado al jugador en dicho año
- **AVV**: Valor promedio del contrato.
- **CONT YR**: Años de duración del contrato.
- **CONT VALUE**: Valor total del contrato.
- **EARNINGS**: Cantidad total de dinero ganada por el jugador en dicho año.
- **FA YEAR**: Año de agente libre disponible.
- **SIGN AGE**: Edad a la que el jugador firmó como agente libre en dicho año.
- **AGE**: Edad del jugador.
- **WEIGHT**: Peso del jugador medido en libras.
- **HEIGHT**: Estatura del jugador medida en pies.

- **Bateadores**

- **POS**: Posición del jugador.
- **TEAM**: Equipo al que pertenece el jugador.
- **GP**: Partidos totales que jugó el bateador a lo largo de la temporada.
- **GP %**: Porcentaje de partidos que jugó el bateador a lo largo de la temporada.
- **GS**: Juegos iniciados por el jugador(*games started*).
- **GS %**: Porcentaje de juegos iniciados por el jugador.
- **AB**: Turnos *al bate*.
- **H**: Bateos conectados (*hits*).
- **2B**: Dobles.
- **3B**: Triples.
- **HR**: *Home runs*.
- **RBI**: Carreras impulsadas (*runs batted in*).
- **AVG**: Promedio de bateos.
- **OBP**: Porcentaje de embasado (*on-base percentage*).
- **SLG**: Porcentaje de *slugging*.
- **OPS**: Porcentaje de *on-base plus slugging*.
- **WAR**: *Wins above replacement*.
- **TVS**: Valor total estandarizado del contrato del jugador.

- **Fildeadores**

- **POS:** Posición del jugador.
- **TEAM:** Equipo al que pertenece el jugador.
- **GP:** Partidos totales que jugó el bateador a lo largo de la temporada.
- **GS:** Juegos iniciados por el jugador(*innings pitched*).
- **H:** Bateos permitidos/en contra (*hits*).
- **R:** Carreras permitidas/en contra (*runs*).
- **ER:** Carreras limpias.
- **BB:** Bases por bola permitidas/en contra (*walks*).
- **SO:** *Strike outs*.
- **W:** *Wins*.
- **L:** *Losses*.
- **SV:** Salvamentos (*saves*).
- **WHIP:** *Walks and hits per inning pitched*.
- **ERA:** Porcentaje de carreras limpias.
- **WAR:** *Wins above replacement*.
- **TVS:** Valor total estandarizado del contrato del jugador.

Por otro lado, en el caso de las bases de datos de bateadores y fildeadores se descargaron en cuatro formatos distintos en ambos casos:

- **Acumulativa:** Es la cantidad total acumulada de la estadística deportiva a lo largo de toda la temporada. Se puede interpretar como la cantidad nominal de dicha estadística.
- **Promedio anual:** Es la cantidad acumulativa de la estadística promediada a lo largo de la temporada.
- **Por juego:** Es la cantidad acumulativa de la estadística deportiva dividida entre el número de partidos en los que el jugador participó en toda la temporada.
- **Por cada 10 juegos:** Es la cantidad por juego multiplicada por 10.

Nótese que en el *Cuadro* (1) se informa sobre la cantidad de bases de datos a lo largo de 12 años de observaciones para cada uno de los formatos distintos para la estadísticas deportivas de los bateadores y fildeadores.

Adicionalmente, se complementaron las bases de datos anteriores con otras que contienen información sobre cada uno de los equipos de la *MLB*. Sin embargo, estas no se extrajeron de una fuente en particular, sino que se armaron manualmente usando distintas fuentes como *Baseball-reference*¹⁵ y *Baseball-almanac*.¹⁶ Para esta categoría de datos se usaron 12 bases de datos que contienen la siguiente información:

¹⁵<https://www.baseball-reference.com>

¹⁶<https://www.baseball-almanac.com>

Categoría	Cantidad de bases
Agentes libres	12
Salarios	12
Bateadores	48
Fildeadores	48
Total	120

Cuadro 1: Cantidad de bases de datos por categoría.

- Acrónimo del equipo.
- Nombre del equipo.
- Victorias del equipo en el año.
- Juegos totales en los que participó el equipo a lo largo de la temporada.
- *Playoffs* ganados por el equipo a lo largo de toda su historia.
- *Pennants* ganados por el equipo a lo largo de toda su historia.
- Series Mundiales (*World Series*) ganadas por el equipo a lo largo de toda su historia.

Para complementar las bases de datos de los equipos, se hizo manualmente una base de datos auxiliar que contiene datos sobre el acrónimo del equipo, el nombre del equipo, el estado geográfico al que pertenece el equipo y la cantidad de equipos que hay en dicho estado¹⁷. La razón de contar con esta base de datos por separado es porque es más eficiente unirla a las otras bases de datos por medio de algoritmos en lugar de hacerlo manualmente.

2.1. Variables agregadas

En la sección de la implementación del modelo se mencionó que se tienen que considerar las estadísticas deportivas de hasta dos periodos anteriores, así como el cuadrado de las estadísticas deportivas. En ambos casos se usarán las estadísticas descritas en las secciones *D.1.1* y *D.1.2* del apéndice, salvo los partidos jugados y derivados por razones anteriormente explicadas. Para las variables con rezagos también las que se muestran abajo como posibles variables controles o instrumentales:

- **GP**: Partidos en los que participó el jugador.
- **GP %**: Porcentaje de partidos en los que participó el jugador.
- **POS**: Posición del jugador.
- **TEAM**: Equipo al que pertenece el jugador.
- Victorias del equipo en el año.

¹⁷Solo se toman en cuenta los equipos que pertenecen a la *MLB*.

- *Playoffs* ganados por el equipo a lo largo de toda su historia.
- *Pennants* ganados por el equipo a lo largo de toda su historia.
- *Series Mundiales* ganadas por el equipo a lo largo de toda su historia.

También se añadieron algunas variables de interés como son la antigüedad del contrato del jugador, el control, dominio y comando de los lanzadores. Por motivo de escalas y buenas prácticas de programación, se agregaron los logaritmos naturales de las siguientes variables de compensación salarial: Salario base, salario de nómina y salario ajustado. En particular, estas sugerencias son dadas por el artículo de Congdon Hohman ([Congdon-Hohman and Lanning \(2018\)](#)).

Por último, cabe especificar las dimensiones de los paneles tras estas consideraciones. Para la base de datos que contiene las observaciones de los bateadores se tienen 7216 observaciones; mientras que para los fildeadores, se cuentan con 6021 observaciones. Cabe aclarar que dependiendo de las regresiones que se quieran estimar o el problema a analizar, los paneles resultantes tendrán una cantidad distinta de observaciones, pero siempre menor a la de los paneles originales.

2.2. Imputación de datos

Una de las razones por las que se destina una sección para hablar sobre la imputación de datos es que no se hizo de la manera estándar para todos los casos que se presentaron en el proceso *ETL*,¹⁸ es decir, no se imputó con valores iguales a cero o a la media para todas las situaciones donde se encuentren valores del tipo *NaN*, *NULL* o *None* ([Enders \(2022\)](#)). Razón por la cual solo nos limitaremos a los casos especiales. Entonces, ofrecemos una serie de consideraciones que son relevantes para nuestras bases de datos, así como para futuros trabajos en deportes que tengan una agencia libre.

- **Edad al firmar:** En caso de que no se encuentre este dato, se puede obtener a partir de la edad del jugador, el año de la observación y la cantidad de temporadas que cubre el contrato.
- **Edad:** La edad es una importante variable de control puesto que más adelante veremos que entre mayor sea el jugador, más probable es que sea un agente libre. Además, como se aprecia en la estadística descriptiva sobre las compensaciones salariales, todos los agentes libres tienen un mayor sueldo que los jugadores que no lo son. Por lo tanto, si el sueldo de este jugador está por encima del de los jugadores que no son agentes libres, se le imputa una edad mínima de 24 años y de 18 en otro caso. Esto es con motivo de ser coherentes con el contexto de las variables. En caso de que estén los datos de la edad al firmar, el año de la observación y la cantidad de temporadas de contratación; se puede calcular la edad.

¹⁸Las siglas se refieren a *Extract-Transform-Load*. Este proceso define todo tratamiento que reciben las bases de datos para limpiarlas y transformar los datos con el propósito de dejarlas listas para el análisis concernientes.

- **Logaritmos de compensaciones salariales:** Dependiendo del tipo de variable salarial salarial, pueden ser o no distintas a cero. Entonces, se tomará el logaritmo de 1 en caso de que la variable salarial sea igual a cero. La razón es que tomando el logaritmo de cero generará valores infinitos que provocarán errores en el software cuando se quieran obtener las regresiones. Por otro lado, debido a que se quiere reflejar el nulo valor de la variable, una buena aproximación es suponer que toman el valor de 1 dolar.
- **Estadísticas deportivas:** De manera general, hay variables que no pueden ser nulas o tienen cotas máximas, mínimas o ambas. Entonces, de acuerdo a la definición de estas, se imputarán de manera distinta.

3. Estadísticas descriptivas

Debido a la gran cantidad de estadísticas descriptivas que se han calculado, la totalidad de las mismas se presentan en el apéndice (E). así como categorías de datos y formatos, es que haremos nuestro análisis de acuerdo a las categorías de datos que nos permitan ser más claros en nuestra explicaciones. Cabe mencionar que la razón por la que se omitieron los diagramas de caja y series de tiempo de los promedios anuales de las estadísticas deportivas en la sección de estadísticas descriptivas (E) es que por el proceso *ETL* para generar las gráficas de los promedios anuales son equivalentes a las acumulativas tanto cualitativa como cuantitativamente. Por otro lado, algo que puede ser lógico preguntarse es si hay en nuestro datos evidencia que muestre que hay una diferencia entre la compensación salarial que perciben los agentes libres y los jugadores que no los son, así como en el desempeño deportivo o cualidades físicas tales como estatura, edad o peso. Entonces, el propósito de esta sección se enfoca en responder esas interrogantes. El orden en el que se abordarán los datos es tratar primero las compensaciones salariales y datos contractuales, luego las estadísticas deportivas y, por último, atributos físicos de los jugadores.¹⁹

Antes de abordar las preguntas que se plantearon anteriormente, cabe resaltar un comportamiento que tienen los datos en el año 2020 en las gráficas de las series de tiempo

¹⁹Muchas de las estadísticas descriptivas de las medidas de desempeño no se muestran en el apéndice. Sin embargo, se proporciona un enlace a un repositorio donde se pueden observar todas las medidas de desempeño con sus respectivas estadísticas descriptivas.

Bateadores		
Base de datos	Dirección del efecto	Variables
Acumulativa	Positivo	%Hit, %GS, OPS, OBP, %GP, SLG.
	Negativo	AB, 2B, HR, RBI, 3B, WAR.
Por juego	Positivo	AB, %Hit, 2B, HR, RBI, %GS, OPS, OBP, %GP, SLG.
	Negativo	Ninguno.
Fildeadores		
Base de datos	Dirección del efecto	Variables
Acumulativa	Positivo	H, R, ER, IP, GS, L, S, W, WAR.
	Negativo	Ninguno.
Por juego	Positivo	L(FA), S, SO(FA).
	Negativo	L(NFA), SO(NFA).

Cuadro 2: En este cuadro se muestran los efectos de la pandemia en los promedios anuales de las estadísticas deportivas de los bateadores y fildeadores de la *MLB*. Las siglas *FA* denota la agencia libre, mientras *NFA* indica a los jugadores que no pertenecen a dicha agencia.

(E.3), tanto para bateadores como fildeadores. A pesar de tratar con promedios de datos normalizados, se aprecia un efecto negativo en variables nominales²⁰ y un efecto positivo en aquellas estadísticas que contemplan su rendimiento promedio a lo largo de la temporada, es decir, basadas en promedios. En el cuadro 2 se muestran los efectos en la base de datos acumulativa (son los mismos que la que tiene datos promediados anuales), entre las bases de datos por juego y cada diez juegos, donde se parecía la dirección del efecto, las variables afectadas y la base de datos a las que pertenecen.

De manera general, lo que podemos decir es que la pandemia afectó estas estadísticas puesto que al haber un encierro se cancelaron mucho partidos. Entonces, el hecho de que la mayoría de los jugadores que normalmente participan entre 161 y 163 partidos por temporada, pasara a jugar entre 35 y 66 partidos en el año 2020 afecta drásticamente las variables que se basan en el total acumulado de dicha estadística a lo largo de la temporada; por lo cual no es coincidencia que todas las variables en las que el efecto fue negativo sean de este tipo. Por el contrario, se obtuvo el efecto opuesto en variables basadas en promedios a lo largo de la temporada como se muestra en el cuadro anterior (2), entonces si hay quienes jugaron relativamente menos partidos al resto de los jugadores en dicho año, habrá quienes tengan un buen desempeño y al haber participado en una menor cantidad de juegos, estos serán acreditados con un gran *score* en dichas estadísticas. Sin embargo, no todo el efecto del shock en las estadísticas deportivas se puede deber a la disminución de la cantidad de partidos en ese año.

Recordemos que el año de inicio oficial de la pandemia del *COVID-19* fue el 2020, la cual causó un confinamiento obligatorio en casi todo el mundo por los primeros meses. Centrándonos en los jugadores de la *MLB*, es de esperarse que este encierro provocase un efecto negativo en su condición física, así como en su habilidad, puesto que no se podía practicar en casa como es debido. Ya sea por el estrés, ansiedad, falta de interacción social u otro factor, también se sufrieron efectos en la salud mental de las personas a nivel global ([Ferwana and Varshney \(2021\)](#)) por lo que es de esperarse que igualmente se vieron afectados los beisbolistas. No obstante, de acuerdo a estudios y análisis hechos en las últimas décadas ([Ohrnberger et al. \(2017\)](#), [Yoon \(2019\)](#)) han mostrado que la salud mental tiene impacto en la salud física, razón por la cual se refuerza el hecho de que la pandemia tuvo un efecto en el desempeño de los jugadores de la *MLB* en las temporadas posteriores al confinamiento. Por lo tanto, podemos pensar que hay dos hechos que pueden ser los responsables de los efectos observados en esos años.

Más allá de detallar las posibles causas del cambio en el desempeño de los jugadores, es importante ver si estos shocks definen un cambio estructural o solo en el año del encierro por *COVID*. En el caso de las bases de datos basadas en acumulación y promedios anuales, se aprecia que después del 2020, los datos se comportan de forma similar previa a la pandemia. Sin embargo, vemos que en el caso de las estadísticas por juego y cada diez juegos, el comportamiento no tiende a ser así en todos los casos ya que hay comportamientos donde la pandemia ocasiona un efecto negativo para luego tener una tendencia

²⁰Entenderemos con variables nominales aquellas en las que se basan en la acumulación de alguna estadística deportiva sin tomar en cuenta la cantidad de partidos jugados en los que se acreditó el jugador con aquellas estadísticas.

mayor a las que había en años anteriores, es decir, en el año 2021 y 2022 se registraron mejores desempeños por juego que en años anteriores. Como particularidad, se aprecia que en el caso de los fildeadores solo se obtuvieron efectos negativos, a diferencia de los bateadores donde las variables porcentuales sufrieron shocks positivos y negativos en las que se basan en acumulación. A pesar de que no haya una razón clara del porqué, esto último podría sugerir un cambio estructural, por lo que se tendrá que analizar con una prueba de *Hausman*.

Por último, vale la pena mencionar otra anomalía que se observan en los diagramas de caja y densidad sobre los años de contrato promedio entre agentes libres y no agentes, en la sección E, donde se observa que hay varios casos atípicos de jugadores que no son agentes recibiendo contratos por más de seis años. Esto llama la atención al tener en cuenta que la mayoría de los agentes libres son contratados por tan solo una temporada. A pesar de que hay casos, aunque raros, sobre jugadores que no son agentes recibiendo contratos por más de seis años,²¹ puede que este anomalía se deba también a que las bases de datos sobre los agentes libres están incompletas. Por ejemplo, en el año 2011, solo tenemos la observación de un agente libre (el receptor Wil Nieves), pero de acuerdo *ESPN*²² hay al menos 42 agentes libres de manera oficial registrados en ese año. Por lo tanto, no se sabe qué tanto puede afectar este hecho en las comparaciones ya que por limitaciones de tiempo no es posible añadir manualmente cada jugador a las bases de datos. No obstante, esto no representa un gran problema en las estimaciones que se llevarán a cabo puesto que solo trabajaremos con los agentes libres; esta falta de datos solo se reflejará en menos datos para la estimación, independientemente de si se es bateador o fildeador.

3.1. Compensaciones salariales

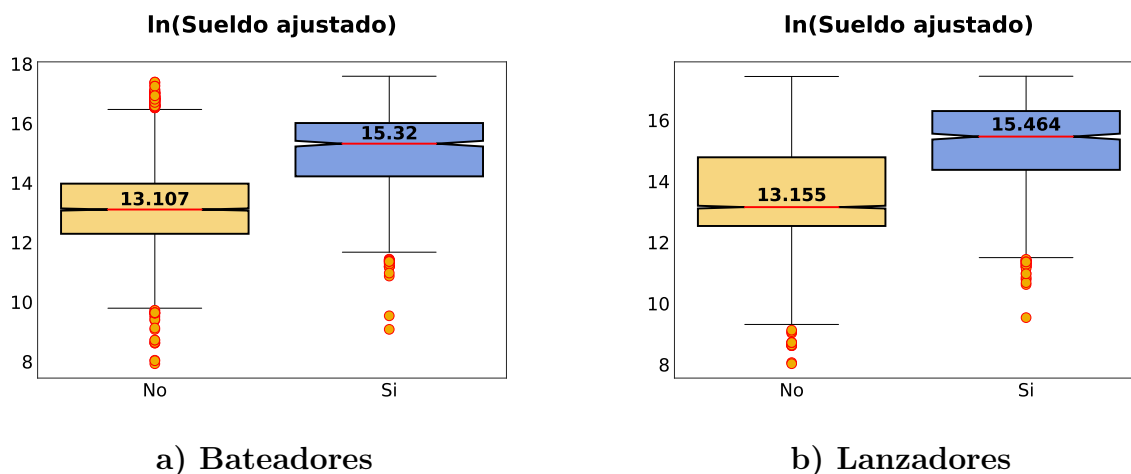
Algo que no causa sorpresa es la diferencia en agregado entre las compensaciones salariales de los jugadores²³ que son agentes libres y aquellos quienes no lo son, tanto para bateadores como fildeadores. De los diagramas de caja (cuadro 3) se aprecia que el promedio del logaritmo del salario base, regular y ajustado son mucho mayor para los agentes libres, donde lo mismo ocurre para en cada año de acuerdo a los gráficos de series de tiempo. En el caso de los agentes el promedio a entre el 2011 y el 2022 es 15,32 y 13,1 para los no-agentes, esto en el caso de los bateadores; mientras que para los fildeadores el promedio es 15,46 y 13,15. Más aún, en los diagramas de densidad, se observa que no hay presencia datos atípicos que indiquen que haya no-agentes que ganan lo mínimo que sí gana un agente libre en los años que abarca el panel. Esto resalta la importancia que tiene hacer un análisis sobre la dinámica del poder de negociación de los agentes libres puesto debido a la gran cantidad de dinero que tiene que gastar un equipo en sus salarios sujeto a la imposición de un *salary cap* aunado a que en la *MLB* se ofrecen en agregado mayores

²¹Un ejemplo es el jugador Bryce Harper quien fue contratado por trece años para los *Philadelphia Phillies* con un contrato cuyo valor es \$330 millones de dolares en 2019. Otro caso es el del jugador Giancarlo Stanton contratado por la misma cantidad de años con un contrato de \$325 millones de dolares en el 2014.

²²https://www.espn.com/mlb/freeagents/_/year/2011

²³Para este trabajo se tomarán las compensaciones salariales ajustadas por la inflación con respecto al año 2022. Se uso este año base debido a que es el último año que abarcan los paneles.

salarios en la agencia libre que en el resto de deportes en EUA. Así mismo, también se aprecia la propiedad de que los contratos ofrecidos están totalmente asegurados o cubiertos puesto que el comportamiento en los gráficos de series de tiempo no sufre cambio en sus tendencias incluso en el año de la pandemia de la *COVID-19* a pesar del cambio tan drástico en los resultados de los jugadores durante los partidos.



Cuadro 3: Diagramas de caja de los promedios de los logaritmos del sueldo ajustado para agentes libres y no-agentes.

Adicionalmente, vemos que para los pocos jugadores que se les ofrece un bono por firmar tanto para agentes libres como para jugadores que no lo son reciben cantidades grandes de dinero al firmar con un equipo. Análogamente a los casos de jugadores que no pertenecen a la agencia libre y reciben contratos por más temporadas que el promedio de los agentes, también se ofrecen bonos por firmar a jugadores que no son agentes. Por lo general, esto ocurre cuando hay jugadores muy buenos en un equipo o se quiere conseguir talento extranjero, por lo que los bonos sirven como un incentivo para que entre a un equipo en particular. Relacionado a esto, tampoco se aprecia ninguna diferencia significativa en la antigüedad que tienen agentes y no-agentes bajo el contrato en el que están, sin importar si son bateadores o fildeadores.

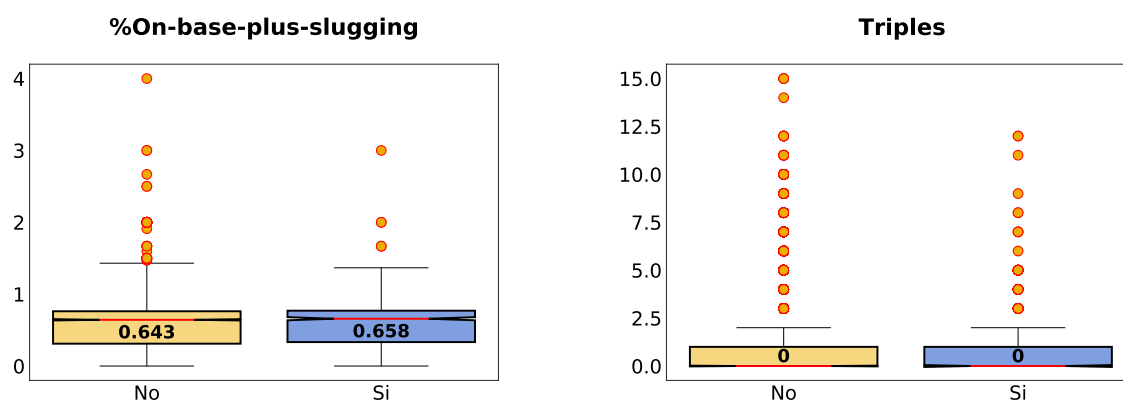
A pesar de los casos excepcionales como los que se han mencionado en esta subsección, es de esperarse que en agregado se aprecie una diferencia considerable en las características del contrato puesto que a los que son agentes libres se les ofrecen en promedio una mayor cantidad de temporadas para trabajar. En promedio, los agentes reciben 2 años frente al año que reciben los no-agentes, tanto para bateadores como fildeadores. Esto se puede explicar debido a que los equipos quieren conservar más años a sus agentes les ofrecen salarios más grandes, así como para incentivarlos. Sin embargo, al estar completamente cubiertos los contratos, puede que los equipos no ofrezcan en promedio contratos aún más largo por temor a que los incentivos a bajar el rendimiento de sus agentes provoque que se haya hecho una inversión que se vuelva contraproducente. En la siguiente subsección veremos si el desempeño deportivo de los agentes libres podría justificar la diferencia de salarios tan grande.

3.2. Estadísticas deportivas

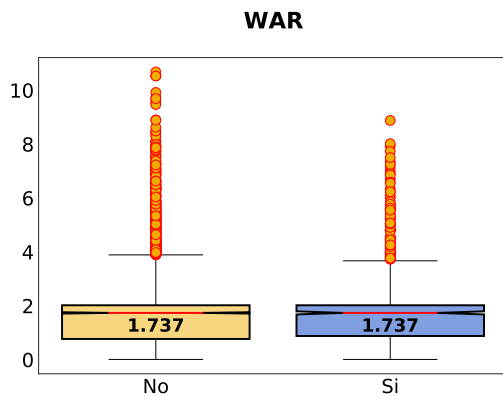
Algo que cabe abordar en primer lugar es el hecho de que en términos de participación o presencia, tanto bateadores como fildeadores, tiende a ser mayor cuando se trata de agentes libres; no solo en agregado, sino también en todos los años. Esto es comprensible puesto que si los equipos están haciendo una inversión grande de dinero para contratar a dichos jugadores, es de esperarse que quieran que tengan una mayor presencia y participación con tal de mejorar las posibilidades de victoria del equipo. Esto se observa en los diagramas de caja donde la cantidad de partidos jugados, nominal y porcentual, es en agregado mayor para los agentes libres independientemente de si la base de datos es acumulativa o por juegos. En el caso de los bateadores se observa que los agentes tienen más apariciones al bate; equivalentemente, los fildeadores poseen una mayor cantidad de *inning pitched*. En ambos casos, sin importar el tipo de base de datos, los agentes libres poseen una mayor presencia en campo. A diferencia del caso de las compensaciones salariales, aquí si se observa que la distribuciones de los agentes y no-agentes se empalman en algunos cuantiles, donde no se debe siempre a la presencia de datos atípicos, es decir, no siempre se debe a casos excepcionales.

3.2.1. Bateadores

Debido a que las estadísticas deportivas con las se evalúa el desempeño de los bateadores son muy distintas a las de los fildeadores, se verán ambos casos por separado. En el caso de los bateadores prestaremos atención a las siguientes: **AB**, **T**, **OBP**, **OPS**, **SLG** y **WAR** (algunas de estas se observan en el cuadro 4). La razón de prestar atención principalmente a estas variables se debe a que son las más comunes para determinar qué tan bueno es un bateador. Cabe mencionar que por limitaciones de las bases de datos, no se cuentan con estadísticas más avanzadas como el *Weighted Runs Created Plus* (**wRC+**). Una de las posibles consecuencias es que perdamos alguna variable que pudiera ser significativa para el modelo y que no toda la varianza que esta explique pueda ser explicada por otra variable.²⁴



²⁴Esto se vería reflejado en las estimaciones conjuntas o análisis por componentes principales.



Cuadro 4: Diagramas de caja de algunas de las medidas de desempeño de los bateadores.

Resulta interesante y desafortunado que al ver los promedios entre agentes y no-agentes en el **WAR**, se observan los mismos promedios de la base de datos acumulativa, aunque en sus distribuciones se logra apreciar que el centro de la campana de los agentes está más a la derecha que la de los no-agentes, esto sugiere que hay datos atípicos que causan un sesgo hacia el cero en el promedio de los agentes, pero se observa que los agentes contribuyen más a la victoria de su equipo con respecto a un remplazo que los no-agentes. Esto último también se ve en los gráficos de series de tiempo donde los no-agentes poseían un **WAR** mayor hasta el año 2016; en todos los años consecutivos los agentes libres se desempeñaron mejor en esa estadística.

El resto de las gráficas tienen un comportamiento similar, el desempeño promedio de los agentes libres es en agregado mayor al de los jugadores que no son agentes, independientemente del tipo de base de datos. En estos casos, la diferencia es más clara tanto en los diagramas de caja como en los diagramas de densidad; en estos últimos se logra observar que el centro de las distribuciones de los agentes se encuentra a la derecha del centro de los no-agentes. Análogo al caso anterior, esto sugiere que en general los agentes lo hacen mejor que los no-agentes y que el promedio no se debe exclusivamente a datos atípicos. A diferencia del **WAR**, estas estadísticas no consideran tantos factores, pero al menos nos dan información sobre que tienen más propensión a llegar a las bases cuando están al bate. Esto se refuerza cuando observamos las carreras limpias que realizan, en la base de datos acumulativa, los agentes realizan más carreras, sin beneficios de error, que los no-agentes; con promedios de 10 y 19, respectivamente. Cualitativamente se nota el mismo comportamiento en la base de datos que considera cada 10 juegos el promedio de las estadísticas, pero no en la que las considera cada juego puesto que el promedio de ambos es cero. En todos los casos, se observa en las distribuciones que el desempeño en dichas estadísticas es mejor en el caso de los agentes libres. Por último, se tiene que mencionar que esta superioridad en agregado no se mantiene a lo largo de todos los años, aunque sí en la mayoría. La única excepción es en el caso de las carreras limpias de la base de datos acumulativa donde se mantiene la superioridad de los agentes en todos los años.

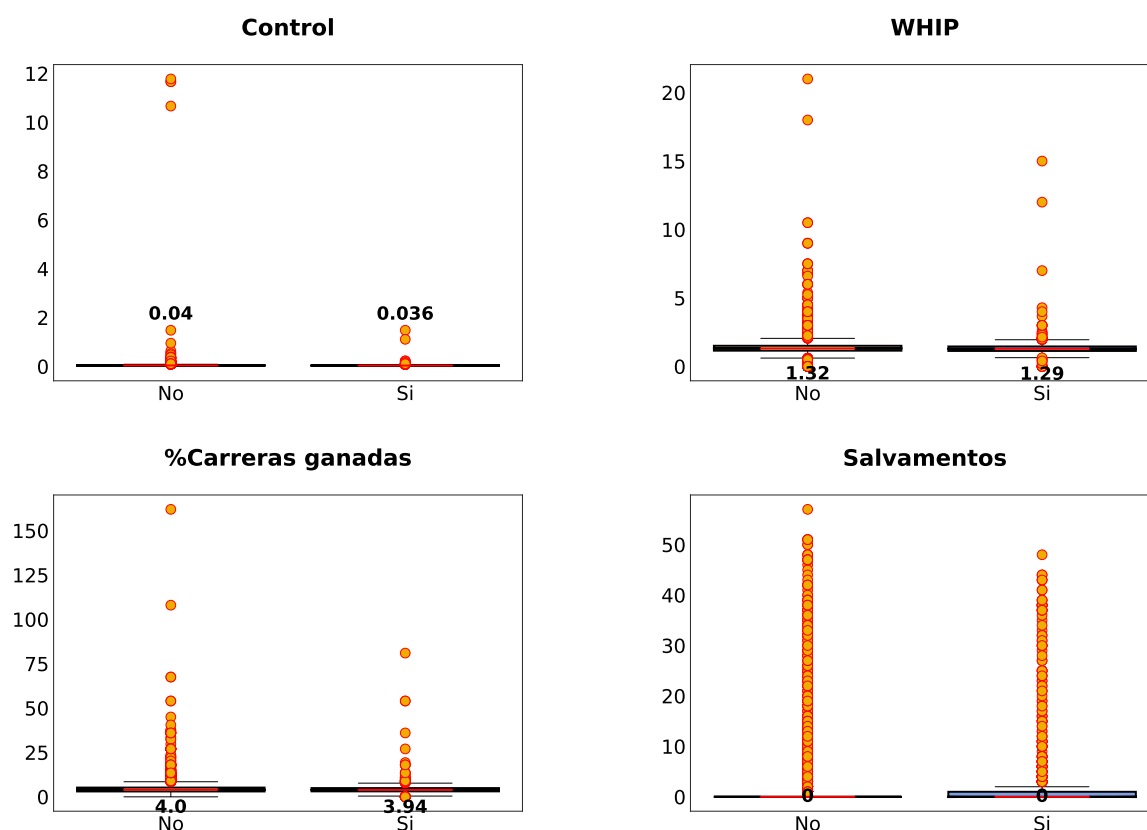
Por otro lado, el mismo comportamiento se aprecia en estadísticas tales como *home runs*

y dobles, donde la superioridad de los agentes se aprecia más en los *home runs*. Sin embargo, hay una estadística en la que los agentes no tienen un mayor desempeño y es en los triples. Llama la atención que sin importar el tipo de base de datos, el desempeño promedio en los triples de aquellos jugadores quienes no son agentes libres no solo es más alto en agregado, sino que es mayor que el de los agentes libres en todos los años. Es aún más interesante cuando esto no se observa en los dobles a pesar de no ser tan marcada la diferencia. Pueden haber algunas hipótesis tomando en cuenta que casi siempre los equipos tienden a abrir en sus líneas principales con los agentes libres, así como que estos están más presentes en el juego que aquellos que no son agentes. Recordando que los triples son raros de realizar, puede que se deba a que estos tienden a realizarse en las últimas entradas (*innings*) y al poder rotar más la ofensiva entre sus jugadores en comparación con los defensivos, entonces están más cansados estos últimos en comparación con los bateadores. Además, cabe mencionar que en la liga americana de la *MLB* está la exclusividad de un bateador designado por lo que al tener jugadores especializados en solo batear y no cubrir ningún puesto defensivo, estos se pueden enfocar más en sus tareas específicas tanto en los partidos como en los entrenamientos por lo que es de esperarse que tenga un mejor desempeño en la mayoría de las estadísticas ofensivas en comparación con aquellos que no son bateadores designados. Sin embargo, esta posición no es exclusiva de los no-agentes por lo que las razones de este comportamiento quedan algo inconclusas.

Sin embargo, al menos gráficamente no se puede determinar si esta variable o no será de utilidad para el problema causal, pero no se puede descartar puesto que a nosotros nos interesa transformar las estadísticas para el modelo y después atender esta cuestión de explicar las diferencias en el cambio del poder de negociación. Esta misma conclusión se aplica a los resultados que se mostrarán de las estadísticas de los fildeadores.

3.2.2. Fildeadores

Para evaluar a los fildeadores existen muchas estadísticas dependiendo de la posición que se ocupe en la defensiva. Sin embargo, las variables que contienen nuestras bases de datos se enfocan más al desempeño de los lanzadores puesto que la mayoría de las estadísticas deportivas enfocadas a evaluar exclusivamente el rendimiento de estos están presentes, lo cual no ocurre con las demás posiciones. Esto no excluye la posibilidad de evaluar el rendimiento de jugadores que ocupen posiciones como receptores, bases o jardineros; solo no hay muchas estadísticas especializadas para estos. En este caso pondremos atención especialmente a la **ERA**, **WHIP**, **SO** y **WAR** (en el cuadro 4 se observan algunas de estas). Esta última estadística está creada para evaluar tanto bateadores como fildeadores por lo que puede que las conclusiones sean análogas a las que se encontraron en el caso de los *hitters*.



Cuadro 5: Diagramas de caja de algunas de las medidas de desempeño de los lanzadores iniciales.

A diferencia del caso de los bateadores, los fildeadores que son agentes libres poseen un promedio mayor al de los no-agentes en **WAR** en el caso de las bases de datos acumulativas y promedios anuales; pero el mismo promedio en las bases promediadas por juego o cada diez juegos. Sin embargo, en todos los casos se aprecia que la distribución de los agentes está más sesgada a la derecha que la distribución de los no-agentes por lo que podemos deducir que la superioridad en los promedios no se debe solo a datos atípicos. En cambio, en el caso del panel acumulativo esta diferencia se mantiene a favor de los agentes en todos los años, salvo uno; mientras que en las bases por juego y cada diez juegos no se mantiene la ventaja todos los años, pero sí la mayoría a partir del 2014.

Al ver todos los diagramas de caja tanto de las carreras ganadas (**ER**) como del porcentaje de las carreras ganadas (**ERA**), ambas por entrada (*inning*), el promedio de los agentes libres es menor que el de los no-agentes, esto nos dice que los agentes permiten menos carreras en contra sin beneficio de error **E**, en agregado. Sin embargo, cuando se ven las distribuciones de las bases de datos acumulativas y de promedios anuales se observa que sus promedios se ven afectados por datos atípicos puesto que el centro de la distribución de los jugadores que no son agentes está a la izquierda del centro de la otra distribución. Mientras que en las bases de datos promediadas por juegos las distribuciones son muy parecidas, donde no se aprecia ninguna diferencia significativa entre el desempeño de ambos grupos. De esto podemos concluir que el desempeño defensivo por juego tiende

a ser ligeramente mejor que el de los no agentes. Más aún, sin importar el formato de la base de datos, en todos los años el desempeño de los defensivos que son agentes es superior al de los no-agentes, es decir, el promedio de los agentes es menor en todos los años. Análogamente, existe otra estadística que en lugar de medir las carreras en contra sin beneficio de error mide las bases por bola y bateos permitidas en contra por entrada (*inning*), es decir, el **WHIP**. En este caso, la superioridad de los agentes frente a los no-agentes para defender es bastante marcada puesto que en todos los casos el promedio de los agentes es menor, aunque en los diagramas de caja no es mucha la diferencia. Sin embargo, al ver las distribuciones no solo se observa que el desempeño de los agentes es superior al ver los centros de las distribuciones, sino que la pequeña diferencia en los promedios agregados de los diagramas de caja se debe a que datos atípicos en contra de los agentes por lo que es de esperarse que el comportamiento promedio de los agentes es mejor de lo que se observa en estos diagramas, donde en la base de datos por cada diez juegos se aprecia más la diferencia de desempeños entre estos grupos. Más aún, cuando se observan las series de tiempo, en todas las bases de datos los agentes tienen un promedio menor al otro grupo en todos los años, la única excepción es para la acumulativa donde en un solo año no se cumplió. En resumen, se puede concluir que el comportamiento agregado de los agentes por entrada es mucho mejor que el de los no-agentes para impedir avanzar a los ofensivos. Esto puede contrastar cuando se observan las carreras, bases por bola y bateos en contra de forma nominal puesto que el promedio de los agentes es mayor. Sin embargo, esto se debe a que los agentes libres juegan más entradas que aquellos quienes no lo son. Entonces, si aquellos jugadores que no son agentes tienen pocas estadística acumuladas en contra, es porque no tienen tantas entradas jugando (**IP**) en comparación con los agentes.

Con respecto a la última de las medidas de desempeño relevantes anteriormente mencionadas tenemos una de las estadísticas deportivas más famosas, los *strike outs* (**SO**). Las distribuciones de todas las bases muestran que los agentes tienen un mejor desempeño para sacar bateadores del juego sin importar qué bases de datos se tomen en cuenta. Sin embargo, los promedios en agregado muestran que los agentes tienen una media menor a los no-agentes, aunque esto se debe a algunos datos atípicos que favorecen el promedio de este último grupo. Por otro lado, al ver las series de tiempo se observa que en casi todos los años los agentes tienen un mejor promedio que los no-agentes muy marcado. No obstante, las estadísticas tales como el comando, control y dominio no nos servirán mucho en este análisis puesto que favorecen matemáticamente a los jugadores novatos ya que si estos no tienen muchas apariciones en el campo o pocos *strike outs* entre otras que aparecen dividiendo otras cantidades, obtendrán estadísticas muy altas. Entonces, si los novatos no tienden a obtener muchas de estas estadísticas, podrán obtener puntajes más altos que incluso agentes libres ya que, tal y como vimos antes, los agentes tienen una mayor participación en los partidos que aquellos quienes no lo son.

Viendo el resto de las estadísticas tales como los salvamentos (**S**), *wins* (**W**) y *losses* (**L**) se observa que tanto el promedio como las distribuciones tienen el mismo comportamiento no solo en agregado a lo largo del panel, sino de todos los años. Tanto las mayores contribuciones a las victorias como a las pérdidas se les atribuyen a los agentes libres. Tal vez parezca contradictorio que obtengan los mayores promedios en las tres y no solo en

las que contribuyan a la victoria o viceversa. Esto se puede deber a la enorme cantidad de entradas que juegan los agentes en comparación con los no-agentes. Entonces, al cubrir más turnos o entradas es más probable que estos se desgasten conforme pase el partido dando pie a estas situaciones. Sin embargo, esto sería algo que los equipos tendrían que notar, razón por la cual se hace la hipótesis a que los ofensivos se van adaptando al estilo de juego de los defensivos ya que estos rotan menos, así mismo, también se les puede estudiar más fácil durante el desarrollo de un partido.

3.3. Resumen de la sección

Hasta aquí es claro que en agregado, tanto agentes libres que son bateadores como fildeadores se desempeñan mejor en juego en comparación con los jugadores que no han entrado a la agencia libre. Al ver la edad promedio de los agentes libres, 32, notamos que hay una ventaja promedio frente a la edad de los no-agentes, 27. Esta ventaja se podría traducir en agregado como mayor experiencia por lo que esto se refleja en sus estadísticas deportivas. No obstante, este factor puede ser solo una de varias causas puesto que también los agentes libres reciben un salario muy superior frente aquellos que no lo son, sin mencionar que la mayoría de los contratos en la agencia libre son por uno o dos años; entonces, esto puede generar incentivos a que realicen un mayor esfuerzo para el equipo a pesar de que no se refleje en sus **WAR**. Cabe resaltar que el resto de atributos físicos como son el peso y la estatura son prácticamente los mismos entre agentes y no-agentes. A pesar de poder pensarse que estos factores no pudieron haber explicado algunas de las diferencias en los desempeños, hay posiciones que exigen ciertos atributos físicos a los jugadores por lo que esto podría tener impacto en los modelos cuando se segmente por posición.

Pudimos mostrar en nuestros datos la importancia que tienen los contratos de los agentes libres debido a la gran cantidad de dinero que se invierte a ellos en comparación con los jugadores que no pertenecen a la agencia libre; sin mencionar las restricciones que tiene cada equipo para conseguir talento nacional e internacional. También se apreció que en la mayoría de las estadísticas deportivas de interés los agentes libres fueron superiores no solo en agregado, sino también en la mayoría de los años. Una posible explicación de esto se puede deber a la ventaja de experiencia que poseen los agentes libres. Por último, es importante resaltar que a pesar de que las estadísticas descriptivas muestren que los agentes son superiores en algunas medidas de desempeño, estas no necesariamente serán significativas en nuestros modelos puesto que el proceso mostrado en el modelo empírico trabaja con las transformaciones de estas estadísticas deportivas.

4. Modelo empírico

En secciones anteriores se menciona que la compensación salarial del agente libre puede depender de más variables aparte del nivel de esfuerzo que el agente realice, a . Sin embargo, dicho modelo de acción oculta presenta otra limitación y es que la variable de estado del agente libre se limita a su utilidad de reserva, \hat{u} , pero esto se puede extender a otras variables tales como el cambio del poder de negociación que no es observable, δ , la cual es nuestro objeto de interés.

Si bien no se han hecho modelos para medir el poder de negociación, δ , en problemas de agencia libre en el deporte, se cuenta con un modelo empírico (DiGiannatale et al (2023)) para estimar de forma indirecta dicha variable al medir los cambios en el poder de negociación entre periodos, definido como la solución al problema de optimización de Bellman (5). No obstante, cabe mencionar que dicho modelo no está especializado para el sector del deporte, sino que está pensado para cualquier sector donde exista una relación principal-agente con un principal neutro al riesgo y un agente averso al riesgo. Dicho modelo analiza la dinámica en tiempo discreto del cambio en el poder de negociación de un agente libre bajo un contrato de duración infinita,

$$\sqrt{\omega_{t+1}} - \sqrt{\omega_t} = \varepsilon \left((-1)^{I_t} \frac{y_t}{\sqrt{y_H}} \right) + u_t \quad (10)$$

Tenemos que ω_t denota la compensación salarial que recibe el agente libre de la *MLB* en el periodo t . Similarmente, y_t es el *output* que genera el dicho agente en el periodo t , en el presente trabajo se refiere a las estadísticas deportivas del jugador, independientemente de si se es bateador o fildeador (véase el apartado E del apéndice). Relacionado a la variable anterior, y_H se refiere al *output* más grande que se haya generado en el equipo de béisbol al que pertenezca el agente a lo largo de todo el periodo de análisis. Por ejemplo, supongamos que nuestro periodo de análisis en los datos abarca del año 2004 al 2017 y que estamos interesados en obtener la y_H de *home runs* de un agente que juega para los *Cubs* de Chicago; si en ese periodo de tiempo la mayor cantidad o *record* de *home runs* que un bateador pudo registrar en cualquier temporada es 35, entonces $y_H = 35$ para cualquier agente que pertenezca a los *Chicago Cubs*, independientemente de si este bateador es el mismo o no. Por último, I_t es una variable dicotómica definida de la siguiente manera

$$I_t = \begin{cases} 0 & \text{dist}(y_t, y_{\text{máx}}) > \frac{1}{2}|y_{\text{máx}} - y_{\text{mín}}| \\ 1 & \text{en otro caso} \end{cases} \quad (11)$$

Hasta aquí, no se aprecia cómo se puede medir el cambio en el poder de negociación a partir de la ecuación (10), pero se sabe que $\Delta\sqrt{\omega_{t+1}} = \Delta\delta_{t+1}$, es decir,

$$\delta_{t+1} - \delta_t = \varepsilon \left((-1)^{I_t} \frac{y_t}{\sqrt{y_H}} \right) + u_t \quad (12)$$

Tal y como se puede intuir, el poder de negociación no se puede observar directamente a partir de los datos. Es esta razón por la cual se usará la ecuación dada por (10) ya que dicho modelo se puede estimar a partir de datos observables como es el caso de la compensación salarial del agente y sus estadísticas deportivas que obtiene en una temporada.

Entonces, estimando ε , podemos obtener la dirección y magnitud del cambio en el poder de negociación, $\delta \in [0, 1]$. El parámetro $\varepsilon \in [0, 1]$ se puede interpretar como la medida de poder de incentivos.

Por otro lado, los errores satisfacen una relación específica

$$u_t = (e_t - e_{t+1}) \sqrt{y_H} \quad (13)$$

Además, se hará el supuesto de que los errores, e_t , poseen una distribución normal $\mathcal{N}(0, \sigma^2)$ e independientes. Por lo general, en las series de tiempo se asumen errores correlacionados por lo que al suponer que los errores, dados en la ecuación (13), son esféricos impone un supuesto fuerte y una restricción en los modelos que se pueden emplear para la estimación del cambio en el poder de negociación. A pesar de que este supuesto sea fuerte y presente una gran restricción para los modelos que se puedan emplear, los resultados encontrados por los autores de este modelo fueron satisfactorios al aplicar un modelo lineal tipo *pooled* y un *ARMA* donde se aplicó en un contexto de agencia libre en distintas empresa. Sin embargo, dichos agentes no se salieron de la empresa durante el periodo que abarca el panel que usaron los autores, es decir, no renunciaron ni se cambiaron de empresa. Esta característica no la presentan los agentes libres en la *MLB* que están en nuestras bases de datos puesto que ninguno juega para siquiera un equipo a lo largo de todos los años que abarca el panel, incluso hay quienes solo tienen observaciones para una temporada. Más aún, hay jugadores que cambian de equipos de una temporada a otra; pero estos hechos no implican directamente que no se puedan obtener buenos resultados para la estimación y el problema de causalidad para explicar las diferencias en el poder de negociación con respecto a la estadísticas deportivas.

Por último, aún está presente el problema de endogeniedad para medir el poder en el cambio de negociación puesto que es equivalente a explicar la diferencias entre los distintos salarios de acuerdo a la habilidad del jugador. El problema de explicar las diferencias en los salarios acuerdo a la habilidad de quien se contrata es bastante conocido.²⁵

4.1. Implementación del modelo empírico

El propósito de esta subsección es explicar cómo se implementará el modelo, así como las distintas especificaciones que se aplicarán a las muestras con las que se harán las regresiones y cómo estas afectan los resultados del modelo.

Cuando se quiere contratar a un jugador que es agente en el periodo t_0 , el principal tiene que proponer inicialmente un contrato donde se estipula la paga que el agente libre recibirá a lo largo de ese temporada (t_0) y las demás que abarque el contrato ($t > t_0$). Sin embargo, el principal no conoce de manera *ex-ante* cómo será el desempeño del agente durante el periodo de contratación, por lo que tendrá que basarse en el desempeño del jugador en periodos previos $t < t_0$. Entonces, en nuestro modelo se traduce a que si queremos explicar las diferencias en el poder de negociación en el periodo t de acuerdo a las medidas de desempeño deportivas, tendremos que usar las medidas del periodo previo

²⁵cameron2005microeconometrics

al de contratación, es decir, $t - 1$. Notemos que esto no afecta el proceso descrito por (10) puesto que solo es un reescalamiento de los índices del tiempo. Por último, de acuerdo a (Hoffman (2014)), estadísticas deportivas de hasta dos temporadas previas ($t - 2$) ofrecen un buen poder de predicción para los salarios de los beisbolistas. A pesar de que nos interesa la significancia estadística, este hecho sugiere que dichas variables pueden reducir la varianza de nuestro modelo, es decir, las usaremos como controles para las medidas de desempeño del periodo $t - 1$. Similarmente, de acuerdo a Congdon-Hohman and Lanning (2018), la compensación salarial del agente libre, ω_t , también depende del cuadrado de las medidas de desempeño, tanto para bateadores como fildeadores. Entonces, también se considerarán las variables descritas en el apéndice (*Medidas de desempeño deportivas*) para las regresiones del cambio en el poder de negociación²⁶.

Por otro lado, se tienen que aplicar los modelos a distintos grupos de jugadores puesto que no todas los jugadores poseen las mismas medidas de desempeño, esto depende de la posición que ocupe el jugador en cuestión, es decir, controlaremos la varianza con la posición que ocupa el jugador como otro control. Por ejemplo, un receptor no puede ser evaluado de acuerdo a la cantidad de *home-runs* que este realice puesto que en esta posición no se batea, así como no tiene sentido evaluar el desempeño de un bateador a través del número veces que atrapa la bola. Siendo los bateadores y *filderos* las dos categorías principales en las que pueda participar un jugador durante el partido, razón por la cual las bases de datos sobre las estadísticas deportivas de los jugadores están separadas de acuerdo a estas dos clasificaciones. Estos grupos tienen algunas variables en común y otras propias de su clasificación. Sin embargo, esta idea se apoya con el hecho de que los bonos por firmar con un equipo (*signing bonus*) solo se les da a agentes libres con el mejor rendimiento de acuerdo a su posición²⁷ (Brown (2009)). Los grupos en los que dividiremos a los agentes libres es de acuerdo a la posición que estos puedan ocupar durante un partido son en lanzadores iniciales, lanzadores de relevo, *infielders* y *outfielders*; en el caso de los fildeadores. Para el caso de los bateadores se dividirán en bateadores regulares bateadores designados (ver sección de posiciones de jugadores en el apéndice).

En caso de que un jugador cambie de posición entre una temporada y la siguiente, se tomará la última de estas, es decir, la del periodo $t + 1$ como está descrito en nuestro modelo.

²⁶Se omitirán la cantidad de partidos jugados, y sus derivados, puesto que estas no revelan ninguna información sobre su desempeño ni sobre sus salarios puesto que casi todos los jugadores tienden a jugar la misma cantidad de partidos. La excepción ocurre cuando hay una lesión, pero al estar completamente asegurados en sus contratos, no afecta la compensación salarial que estos puedan recibir en esa temporada y todas las que abarque el contrato.

²⁷El bono solo se les otorga a agente libre del tipo A, los cuales son quienes están en el 20 % de los agente libres con mejor rendimiento de acuerdo a su categoría. Existe otra categoría llamada agente libres del tipo B que se refiere a los jugadores que están entre el 21 % y 40 % mejores con respecto a la categoría con la que se analizan sus estadísticas deportivas. Las categorías en las que evalúan a los agentes libres son receptores (C) en el primer grupo; bateadores designados (DH), primeras bases (1B) y *outfielders* en el segundo grupo; segunda bases (2B), terceras bases (3B) y campo cortos (SS) en el tercer grupo; lanzadores iniciales (SP) en otro grupo; y por lanzadores de relevo (RP) en el último grupo.

Dejamos hasta el último la consideración más importante para el modelo. Tal y como se había mencionado anteriormente, los agentes libres de la *MLB* tienen un contrato que los asegura completamente a lo largo de todas las temporadas que cubre el contrato independientemente de su desempeño en los partidos o incapacitaciones que este sufra. Entonces, esto provoca que un efecto ambiguo sobre los incentivos en el agente ya que este podría a bajar su desempeño si sabe que recibirá la misma paga o podría esforzarse cada vez más para optar por un mejor salario en el siguiente partido; en ambos casos, va a recibir el mismo salario y por ende, estas diferencias en su desempeño no se apreciarán en su compensación salarial. Como consecuencia, el único cambio en los sueldos que se observa entre temporadas es debido al ajuste por la inflación. Sin embargo, la consecuencia más importante es que teóricamente no se refleja el cambio en el poder de negociación del agente puesto que su salario será el mismo desde que firmó su contrato con el equipo, es decir, no vuelve a negociar su salario después de la primera temporada. Por lo tanto, no tiene sentido teórico considerar más de un año del periodo de contratación. Por ejemplo, supongamos que a un agente se le contrata por cinco temporadas, entonces, solo consideraremos las observaciones de la primera temporada para estimar las regresiones.

4.2. Casos de interés

Aparte de nuestro objetivo principal alrededor del problema de indentificación, también han surgido ciertos comportamientos que aumentan los casos de interés a investigar. Las hipótesis de mayor interés que vamos a investigar son las siguientes:

- H.1 Las medidas de desempeño de los jugadores explican parcialmente las diferencias en el cambio del poder de negociación, $\Delta\delta_i$.
 - H.1.a No todas las variables identificadas son necesarias de manera conjunta para explicar $\Delta\delta_i$, es decir, no se necesitan todas las variables para explicar la varianza entre ellas.
 - H.1.b Las variables de desempeño que consideren más aspectos del jugador tenderán a ser las más significativas en el modelo empírico.
- H.2 El convertirse en agente libre tiene un impacto positivo en $\Delta\delta_i$.
- H.3 La edad del agente libre tiene efectos en $\Delta\delta_i$.
 - H.3.a El impacto de la edad cambia a lo largo del tiempo.
 - H.3.b Los efectos de la edad no tienen que ser los mismos entre las distintas posiciones que puede ocupar un jugador.
- H.4 Hay un cambio estructural en el modelo empírico causado por el año oficial del inicio de la pandemia del COVID-19.

5. Análisis econométrico

En esta sección se mostrarán los resultados e interpretaciones de las distintas estimaciones sobre los casos de interés. En todos los casos se usará un estimador de primeras-diferencias cuyos resultados serán los principales. No obstante, se contrastarán con las estimaciones obtenidas con los siguientes estimadores: *pooling*, *within* y efectos aleatorios. La razón del porqué nos centraremos en los resultados del estimador en primeras-diferencias se debe a que más del 85 % de los jugadores tienen dos o un año de observaciones, independientemente del panel, por lo cual es más eficiente este estimador en lugar del *within*. Debido a que los estimadores *pooling* y el de efectos aleatorios imponen más restricciones o supuestos fuertes en comparación con el estimador en primeras-diferencias es que no se preferirán aquellos estimadores. Cabe mencionar, que los resultados que se discutirán en esta sección apoyarán esta decisión puesto que no se encontrará evidencia explícita en contra de nuestra decisión.

Primero, se abordará el problema de indentificación donde se mostrarán los estimadores, errores robustos y significancias de la estimación para cada una de las estadísticas deportivas transformadas a X_{var} del modelo empírico, así como los controles. Después, se llevará a cabo la estimación conjunta para cada uno de los estimadores usando las X_{var} que fueron estadísticamente significativas en el problema de indentificación. En la segunda parte estudiamos el efecto de la edad en el cambio en el poder de negociación en tres distintos periodos. Por último, se analizan otros casos de interés como es el impacto de la pandemia del COVID-19 en la dinámica del poder de negociación, los beneficios de usar componentes principales y el estudiar el impacto de ser agente libre mediante diferencias-en-diferencias.

5.0.1. Identificación

El objetivo principal de este trabajo gira alrededor del problema de identificación, es decir, de identificar qué variables pueden explicar las diferencias en el poder de negociación, δ . Por lo tanto, nos interesa encontrar qué variables son estadísticamente significativas para el modelo empírico (12) de forma individual, pero con sus respectivas variables de control. Notemos que identificar qué estadísticas deportivas son significativas es equivalente a identificar las variables X_{var} ya que es a partir de las variables de desempeño que se obtienen estas últimas.

Los controles que se usarán en todas las estimaciones son la edad ($Edad_t$), los años de duración del contrato en el que se encuentre el jugador en el periodo t ($Años\ contrato_t$), una dummy del equipo al que pertenece el jugador ($Equipo_t$), una dummy que indica si el jugador es un agente libre ($Agente_t$)²⁸ y las estadística deportiva transformada del periodo $t - 1$ ($X_{var_{t-1}}$). Además, el tiempo, t , se refiere solo al primer año de cada contrato distinto en el que un agente libre esté.

²⁸En todas las estimaciones, salvo las del apartado de los efectos de ser agente libre, abarcan solo los periodos en los que el jugador pertenece a la agencia libre por cuestiones teóricas. Por lo tanto, todos los jugadores en estas estimaciones tendrán el mismo valor para dicha variable independientemente del periodo que se analice.

Por último, antes de abordar el análisis, cabe aclarar que la estrategia en el proceso de identificación será escoger aquellas estadísticas deportivas que sean significativas pertenecientes al periodo t o $t - 1$. La razón por la que también consideramos las variables de un periodo previo se deben a que el modelo teórico toma en cuenta la historia de todos los *outputs* del agente libre; además, es sabido que variables de hasta dos periodos anteriores al de negociación también afectan la decisión sobre su salario.

Entonces, procedamos al análisis de las estimaciones para bateadores y lanzadores iniciales. Lo primero que resalta al ver las tablas de la subsección de robustez (sección F.2 del apéndice) es que el estimador en primeras-diferencias captura en la mayoría de sus variables identificadas el efectos de las variables del periodo $t - 1$ independientemente de la posición del jugador. Estos resultados ofrecen evidencia de la validez empírica de los supuestos del modelo empírico puesto sobre que el historial de *outputs* del agente son relevantes para la dinámica en el poder de negociación. Inclusive, hay casos para ambas posiciones donde la variable del periodo $t - 1$ es estadísticamente significativa, pero no la del periodo t .

Sin embargo, sabemos que hay variables especializadas comúnmente empleadas para evaluar el desempeño del jugador de acuerdo a la posición que este ocupe, salvo el caso del **WAR** que engloba distintos aspectos del jugador independientemente de la posición que este ocupe. En el caso de esta última, se aprecia que **WAR** y **WAR**² son significativas para el modelo, pero solo para el último periodo antes de un nuevo contrato por lo que sugiere que el principal le da más peso a la última valuación general del jugador, es decir, la información más reciente sobre su desempeño global en el equipo. Ahunado a esto, cobra más fuerza esta afirmación si recordamos que dichas estadísticas deportivas miden el impacto del jugador en el desempeño general del equipo, es decir, qué tanto contribuye a la victoria de su equipo con respecto a un reemplazo. No obstante, hay más variables que son comumente empleadas para evaluar el desempeño de un bateador tales como los bateos promedio (**BA**), **OPS**, **OBP**, **HR** o **RBI** las cuales fueron identificadas por el estimador de primeras-diferencias ya sea con la variable lineal X o su cuadrática X^2 . Esto resulta conveniente para la validez de este estimador puesto que concuerda con la evidencia empírica. Sin embargo, no fue el caso del resto de estimadores puesto que no lograron captar los efectos de estas variables en la dinámica de δ , aunque también aparecen identificadas variables del periodo $t - 1$, razón por la cual se puede argumentar que hay robustez entre los estimadores para concluir que empíricamente sí importa el historial del agente libre.²⁹

Similarmente para los lanzadores iniciales, hay estadísticas especializadas para los lanzadores iniciales tales como **ERA**, **WHIP**, *strike-outs* **SO** o **comando**. Análogamente al caso de los bateadores, todas estas variables fueron identificadas por el estimador de primeras-diferencias por lo que otra vez el modelo está acorde con la evidencia empírica. Sin embargo, lo más sorprendente es que los estimadores ε de las estadísticas **ERA** y **WHIP** son negativos, siendo que esto no lo prevee la base teórica del modelo. Más aún, tanto la significancia como el signo del estimador son robustos ante los distintos

²⁹No obstante, algunos autores sostienen que se le da más peso en las negociaciones a la estadísticas deportivas que sean más recientes.

estimadores (ver cuadro 7).³⁰ No obstante, esto no contradice la teoría debido a que estas variables impactan de manera negativa en el desempeño del agente libre puesto que estas estadísticas miden carreras, bases alcanzadas o bateos en contra del lanzador en turno. Por lo tanto, es de esperarse que entre mayor sean estas variables, peor impactan en el desempeño general del equipo por lo que intuitivamente disminuiría el poder de negociación del jugador para la siguiente negociación. Cabe mencionar que también todos los estimadores coinciden en la relevancia estadística de estas variables cuando corresponden al periodo $t - 1$, incluso hay casos como las del **WHIP** donde solo no se identifica para ningún estimador ningún efecto del periodo t .

Por otro lado, a pesar de uno esperaría que solo las estadísticas deportivas que por definición tengan un impacto negativo en la valuación del desempeño posean un estimador negativo, este no siempre es el caso. Por ejemplo, consideremos el caso de los bateadores donde no poseen variables de desempeño negativas en ninguna estimación, pero aún así hay estimadores negativos para la mayoría de los estimadores, salvo el *within*. Esto va en contra de la intuición puesto que uno esperaría que variables como los *home-runs* tengan un impacto positivo en la dinámica del poder de negociación, pero los resultados muestran que no necesariamente; similar con otras estadísticas. Sin embargo, algo importante que tenemos que mencionar es que este comportamiento no es robusto entre los distintos estimadores, por ejemplo, las variables asociadas a los triples T poseen un estimador negativo para el caso del estimador en priemras diferencias, pero no para el caso del estimador *pooling*. El resultado que sí robusto con respecto a la significancia, signo del estimador y periodo identificado es la estadística **WAR** la cual fue identificada por todos los estimadores (cuadro 6); con el periodo t en común, concordando lo anteriormente explicado sobre el peso de las valuaciones del jugador más recientes.

Variable	<i>Estimador</i>			
	Pooling	Within	RE	FD
AB				$t - 1$
AB²				
H	t			t
H²				$t - 1, t$
BA				$t - 1$
BA²				$t - 1$
RBI	t			
RBI²				
D				
D²				
HR				t
HR²				t
GS				$t - 1$
GS²				

³⁰La única excepción es el *WHIP* en el caso del estimador *within*.

Variable	<i>Estimador</i>			
	Pooling	Within	RE	FD
OPS				$t - 1$
OPS ²	t		t	
OBP	t			$t - 1$
OBP ²	t			$t - 1, t$
%SLG				
%SLG ²	t			$t - 1$
T	$t - 1$		$t - 1$	t
T ²				$t - 1$
WAR	$t - 1, t$	t	$t - 1, t$	t
WAR ²	$t - 1$		$t - 1$	t

Cuadro 6: Resultados de las estimaciones individuales del problema de indentificación para los bateadores. En caso de la estadística deportiva asociada a X_{var} haya sido significativa para el modelo, se indica el periodo de la variable que lo fue (pueden ser ambos periodos).

Similarmente, este comportamiento de obtener estimadores cuyo signo choca con la intuición también aparece en el caso de los lanzadores iniciales. Sin embargo, en este caso, hay más variables que son robustas al menos para tres de cuatro estimadores, donde aquellas que son robustas para los cuatro estimadores son la **ERA** y los salvamentos (**S** y **S**²). De nuevo, ambas variables robustas con respecto al signo del estimador, significancia estadística y en periodos identificados.³¹ Análogamente, hay resultados robustos con respecto a las mismas características sobre que variables no son relevantes para el modelo empírico tales como los dobles (**D**) o juegos iniciados (**GS**), en el caso de los bateadores; sin embargo, esto no sucede con los lanzadores puesto que nuestro estimador de primeras-diferencias es capaz de identificar el impacto de cada una de las estadísticas deportivas en la dinámica del poder de negociación.

Variable	<i>Estimador</i>			
	Pooling	Within	RE	FD
BB				$t - 1, t$
BB ²				$t - 1$
H				t
H ²				t

³¹La excepción es para los estimadores de los salvamentos para el estimador *pooling*.

Variable	<i>Estimador</i>			
	Pooling	Within	RE	FD
R		t		$t - 1$
R²				
ER				$t - 1$
ER²				t
Comando			$t - 1$	$t - 1, t$
Comando²		$t - 1, t$	$t - 1$	$t - 1$
Control	$t - 1$		$t - 1$	$t - 1$
Control²	$t - 1$		$t - 1$	$t - 1, t$
Dominio	$t - 1$		$t - 1$	$t - 1, t$
Dominio²	$t - 1$		$t - 1$	$t - 1, t$
IP				$t - 1, t$
IP²				t
GS				
GS²				
L				
L²				t
ERA	$t - 1, t$	$t - 1$	$t - 1$	$t - 1$
ERA²	$t - 1$		$t - 1$	
S	$t - 1$	t	$t - 1, t$	t
S²	$t - 1$	$t - 1, t$	$t - 1, t$	$t - 1, t$
SO				$t - 1, t$
SO²				$t - 1, t$
WAR				
WAR²				$t - 1$
WHIP	$t - 1$		$t - 1$	$t - 1$
WHIP²	$t - 1$		$t - 1$	$t - 1$
W				$t - 1$
W²		$t - 1$		

Cuadro 7: Resultados de las estimaciones individuales del problema de indentificación para los lanzadores iniciales. En caso de la estadística deportiva asociada a X_{var} haya sido significativa para el modelo, se indica el periodo de la variable que lo fue (pueden ser ambos periodos).

5.1. Estimaciones conjuntas

La razón de considerar las estimaciones conjuntas se debe a que se quiere determinar con qué variables disponibles en nuestro páneles nos bastan para poder determinar un conjunto

mínimo de estas para explicar las diferencias en el poder de negociación. Además, de ver el efecto en el cambio del poder de negociación de cada una de las variables asociadas a estas estadísticas deportivas cuando se consideran de manera conjunta.

La estrategia para llegar al mínimo de variables necesarias consiste solo usar las variables que fueron estadísticamente significativas en el problema de identificación, independientemente del periodo. Posteriormente, se hace la estimación con todas esas variables y aquellas que no fueron significativas para el modelo se descartan. Se seguirá este proceso hasta que solo queden aquellas que son relevante estadísticamente para el modelo independientemente de si el signo de su estimador va acorde o no con la intuición.

Los resultados se pueden apreciar en el apéndice, en el apartado de estimaciones conjuntas (sección G). De nuevo, es interesante como tras el proceso de refinamiento, todos los estimadores dan como resultado que la variable asociada al **WAR** del jugador del periodo t es estadísticamente relevante y robusto para el modelo en el caso de los bateadores. Más aún, salvo el estimador *within*, el resto de estimadores tienen aproximadamente la misma magnitud. Similarmente para los lanzadores, se aprecia lo mismo con la variable asociada a **ERA** del periodo $t - 1$ que no solo es robusta de forma individual, sino también de manera conjunta. Nótese que el signo negativo de sus estimadores también es robusto, pero que dependiente de si se asume un modelo de efectos fijos o no, la magnitud de los estimadores será distinta, aunque aproximadamente igual para los distintos estimadores asociados al mismo modelo.

<i>Cambio en el poder de negociación: Y_{it}</i>				
	Pooling	Within	RE	FD
$Edad_t$	-0.006** (0.003)	-0.006 (0.004)	-0.006** (0.003)	-0.011*** (0.002)
Años contrato $_t$	-0.004 (0.004)	-0.039*** (0.012)	-0.007* (0.004)	-0.050*** (0.009)
Equipo $_t$	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.002*** (0.001)
$X_{T_{t-1}}$	0.010* (0.005)			
X_{AB_t}				0.003*** (0.001)
$X_{AB_{t-1}}$				-0.002*** (0.0004)
X_{H_t}				-0.003*** (0.001)
$X_{H_{t-1}^2}$				-0.0005*** (0.0001)
X_{GS_t}				-0.004** (0.002)
$X_{GS_{t-1}}$				0.006***

<i>Cambio en el poder de negociación: Y_{it}</i>				
	Pooling	Within	RE	FD
$X_{OPS_{t-1}}$				(0.001) -0.056***
$X_{OBP_{t-1}}$				(0.012) 0.113***
$X_{OBP_t^2}$				(0.012) 0.063***
X_{T_t}				(0.019) -0.067***
$X_{T_t^2}$				(0.010) 0.025***
X_{WAR_t}	0.016** (0.007)	0.035*** (0.009)	0.019*** (0.006)	(0.005) 0.015***
$X_{WAR_{t-1}}$				(0.004) 0.008*
$X_{WAR_t^2}$				(0.005) 0.010**
$X_{HR_{t-1}}$				(0.005) -0.006***
$X_{RBI_{t-1}}$				(0.002) 0.004**
Agente _t	0.187** (0.081)		0.181** (0.082)	(0.002)

Note: *p<0.1; **p<0.05; ***p<0.01

Cuadro 8: Estimaciones conjuntas con las medidas de desempeño de los bateadores para los estimadores *pooling*, *within*, efectos aleatorios (RE) y primeras-diferencias (FD).

Por último, en la sección G del apéndice, se muestran los resultados de aplicar la prueba de Hausman para cambio estructural entre las distintas estimaciones conjuntas. Resalta el hecho de que entre todas las estimaciones conjuntas para los bateadores hay evidencia de un cambio estructural (a 1 %). Sin embargo, debido a las razones explicadas en un inicio sobre porqué empleamos como estimador principal el de primeras-diferencias es que preferiremos este último sobre los demás. Por otro lado, para los lanzadores solo hay evidencia de un cambio estructural (a 1 %) entre el estimador *within* y primeras-diferencias lo cual resulta curioso debido a no hay evidencia con respecto a los otros estimadores que parte de un modelo distinto. Análogamente, no se encuentra evidencia en contra de usar el estimador de primeras-diferencias para ambas posiciones. Por otro lado, también resulta

preferible utilizar este estimador debido a que es capaz de identificar muchas más variables que el resto de estimadores, incluso identifica estadísticas deportivas que otros modelos no, pero cuya relevancia está respaldada por distintas fuentes.

<i>Cambio en el poder de negociación: Y_{it}</i>				
	Pooling	Within	RE	FD
$Edad_t$	-0.008** (0.004)	-0.020* (0.012)	-0.009** (0.004)	-0.016*** (0.004)
Años contrato $_t$	-0.013* (0.007)	-0.017 (0.020)	-0.013* (0.007)	-0.058*** (0.012)
Equipo $_t$	0.002 (0.001)	0.004 (0.002)	0.002 (0.001)	0.002* (0.001)
$X_{Control_t^2}$	-0.157** (0.071)		-0.148** (0.071)	
$X_{Control_t}$	0.091** (0.041)		0.084** (0.041)	
$X_{H_t^2}$				0.0005** (0.0002)
$X_{H_{t-1}^2}$				-0.0004*** (0.0001)
X_{H_t}				0.020*** (0.002)
$X_{ER_t^2}$				-0.001*** (0.0003)
$X_{Dominio_{t-1}}$	0.047*** (0.014)		0.043*** (0.014)	0.042*** (0.009)
$X_{IP_t^2}$				-0.001*** (0.0001)
$X_{L_{t-1}^2}$				-0.003*** (0.001)
$X_{ERA_{t-1}}$	-0.019*** (0.006)	-0.034*** (0.011)	-0.019*** (0.006)	-0.036*** (0.006)
X_{R_t}				-0.023*** (0.003)
$X_{Comando_{t-1}^2}$				0.0004*** (0.0001)
$X_{Comando_t}$				0.048*** (0.006)
$X_{Comando_{t-1}}$				-0.046*** (0.006)
$X_{Control_{t-1}^2}$				-0.098*** (0.013)
$X_{Control_{t-1}}$				-0.053***

<i>Cambio en el poder de negociación: Y_{it}</i>				
	Pooling	Within	RE	FD
$X_{Dominio_t^2}$				(0.012) -0.151***
$X_{Dominio_t}$				(0.011) 0.134***
$X_{Dominio_{t-1}^2}$				(0.020) -0.084***
X_{ERA_t}	-0.013** (0.006)		-0.012** (0.006)	(0.011) -0.046***
$X_{S_t^2}$		-1.883*** (0.656)		(0.007) -2.435***
$X_{S_{t-1}^2}$	-0.194** (0.090)	0.066*** (0.019)	-0.170** (0.083)	(0.439)
X_{S_t}	0.374** (0.159)		0.332** (0.145)	
$X_{S_{t-1}}$		1.447*** (0.465)		1.770*** (0.295)
$X_{SO_{t-1}^2}$				0.001*** (0.0001)
X_{SO_t}				0.005*** (0.001)
$X_{SO_{t-1}}$				-0.005*** (0.001)
$X_{WAR_{t-1}^2}$		-0.008** (0.003)		-0.017*** (0.002)
$X_{WAR_t^2}$				0.081*** (0.012)
$X_{BB_t^2}$				0.001*** (0.0002)
X_{BB_t}				0.006*** (0.002)
$Agente_t$	0.257** (0.123)		0.275** (0.132)	

Note: *p<0.1; **p<0.05; ***p<0.01

5.2. Efectos de la edad en el poder de negociación

Una característica que no se ha abordado aún es la robustez de los estimadores de los controles con respecto a la significancia estadística, signo o magnitud. En particular, nos centraremos los efectos de la edad del agente libre (Edad_t) sobre la dinámica del poder de negociación (δ). No obstante, cabe mencionar que no ignoramos que hay robustez con respecto a estas características para los estimadores de los años de contrato (Años contrato_t) o la dummy del equipo al que pertenecen (Equipo_t). Basta observar para cada posición (batedores y lanzadores) que en todas estimaciones con el estimador de primeras-diferencias tienden a poseer la misma significancia, signo del estimador y magnitud independiente de la estadística deportiva asociada. Se puede observar un comportamiento similar para el resto de estimadores en el problema de identificación. Además, se mantiene la robustez ante modelos en las estimaciones conjuntas con respecto a la significancia y signo del estimador; mientras que la magnitud tiende ser similar para tres de los cuatro estimadores, incluso en los resultados con componentes principales. El efecto de la dummy asociada a ser agente libre (Agente_t) se tratará en otra parte del análisis.

Con el propósito de estudiar cómo cambia el impacto de la edad del agente libre en el poder de negociación con respecto al tiempo es que se realizarán dos estimaciones por estadística deportiva (tablas de la sección H del apéndice). La primera estimación abarca los primeros dos periodos del jugador como agente libre y, por ende, la otra regresión abarca el resto de años. Es importante tomar en cuenta que estos periodos, t , no son necesariamente temporadas o años consecutivos ya que cada periodo solo abarca el primer año de cada contrato distintos que un agente tenga a lo largo de su carrera.³² Entonces, una de las ventajas que ofrece esta definición para los periodos, aunado al hecho de que existe una correlación fuerte y positiva entre el valor del contrato y los años que este abarca, es que teóricamente provoca que los efectos de la edad sean más apreciables. Esto se debe a que no son muchos los agentes libres que tengan más observaciones bajo estas especificaciones y son muchos de ellos quienes llegan a tener contratos de más de cuatro años, entonces esto tiende a evitar problemas para identificar si las variaciones en las estimaciones se deben al efecto de la edad o por otros factores.

En el apéndice se pueden observar los resultados del problema de identificación para el caso del estimador en primeras-diferencias (cuadros 30 y 31), pero también se proporciona el enlace para observar el resto de regresiones con los otros tres estimadores. En el caso de primeras-diferencias se logra apreciar en las tablas de resultados que para los bateadores (cuadro 10) que hay un comportamiento común e independiente de la estadística deportiva con la que se realice la estimación, y es la dirección del cambio entre los estimadores para el control de la edad. Se aprecia que los estimadores pasan de ser positivos y no significativos a negativos y estadísticamente significativos (a un nivel del 5 %) donde para la estimación con los periodos $t > 2$ el estimador de la edad tiende a poseer la misma magnitud en la mayoría de los casos. No obstante, vemos que esto no se tiende a cumplir para aquellas variables en las que se encontró evidencia de un cambio estructural, bajo un test de

³²Nótese que para realizar las estimaciones del modelo empírico es indispensable que el jugador cuente con al menos dos contratos. Además, al emplear variables del periodo $t - 1$ se implica que ahora se necesitan mínimo tres periodos o contratos.

Hausman. Salvo el **WAR** que es una variable análoga a un componente principal, el resto de variable que rechazan la prueba tienden a ser aquellas relacionadas con la propensión a hacer bases o carreras. En conclusión, esto nos dice que conforme pasan el tiempo, la edad tienen a poseer un efecto negativo en la dinámica del poder de negociación. Es interesante notar que la dirección del cambio para los lanzadores es opuesta a la de los bateadores puesto que en la mayoría de los casos el impacto es positivo. Si bien hay una excepción (**Dominio**), en magnitud sí se cumple para todas las variables de desempeño que el estimador crece después de dos periodos donde, además, estas pasan a ser estadísticamente significativas. En este caso, se observa que la edad posee un efecto mucho menor en los primeros dos periodos que en las estimaciones para el resto de periodos. Por lo tanto, la conclusión es opuesta a la hecha en el caso de los bateadores con respecto al efecto de la edad. Con respecto a una posible relación entre las magnitudes de los estimadores y el p -value del test de Hausman, en la mayoría de los casos (salvo **ER**, **Control**² y **Dominio**²) hay evidencia de una cambio estructural para un nivel de tolerancia del 5 %.

Variable	Primeros dos periodos	Periodos restantes	$\Delta\beta_{\text{Edad}}$	p-value
AB	0.011	-0.016***	Negativo	0.4813
AB ²	0.007	-0.015***	Negativo	0.8949
H	0.008	-0.016***	Negativo	0.9425
H ²	0.007	-0.015***	Negativo	0.9979
BA	0.007	-0.015***	Negativo	0.3685
BA ²	0.007	-0.015***	Negativo	0.1045
RBI	0.008	-0.015***	Negativo	0.4797
RBI ²	0.007	-0.015***	Negativo	0.2255
D	0.009	-0.016***	Negativo	0.4156
D ²	0.011	-0.015***	Negativo	0.9985
HR	0.017	-0.015***	Negativo	0.9828
HR ²	0.008	-0.015***	Negativo	0.03681
GS	0.009	-0.015***	Negativo	0.1121
GS ²	0.008	-0.015***	Negativo	0.4797
OPS	0.006	-0.012***	Negativo	5.109e-05
OPS ²	0.006	-0.015***	Negativo	3.621e-09
OBP	0.003	-0.025***	Negativo	0.04995
OBP ²	0.005	-0.016***	Negativo	0.0002568
%SLG	0.006	-0.014***	Negativo	0.0001872
%SLG ²	0.008	-0.016***	Negativo	0.4813
T	0.006	-0.012***	Negativo	5.109e-05
T ²	0.006	-0.015***	Negativo	3.621e-09
WAR	0.003	-0.025***	Negativo	0.04995
WAR ²	0.005	-0.016***	Negativo	0.0002568

Variable	Primeros dos periodos	Periodos restantes	$\Delta\beta_{\text{Edad}}$	p-value
Cuadro 10: Estimadores de las regresiones para ambos periodos de los bateadores. Se muestra la estadística de portiva asociada a los estimadores, la dirección del cambio del estimador de la edad y el p-value de aplicar un test de Hausman a dichos periodos.				

Una posible explicación de la diferencia de conclusiones entre bateadores y lanzadores se puede deber a las exigencias que requiere cada una de las posiciones. Por ejemplo, del lado de los fildeadores se necesita trabajar de manera conjunta para evitar que el bateador en turno anote una carrera por lo cual el desgaste del cuerpo de aquellos que sean mayores se compense con el de aquellos que sean jóvenes debido a la coordinación que estos deben de pooser para trabajar adecuadamente en equipo dando más peso a la experiencia de los jugadores defensivos que el envejecimiento de estos. Por otro lado, un bateador juega de forma individual³³ donde además se tiene que correr rápido para llegar a la siguiente base antes de que se le marque un *out*. Entonces, es de esperarse en general que entre mayor sea el jugador, más se le dificultará llegar a las bases lo cual se apoya con los resultados de las variables donde hay cambios estructurales ya que muchas giran entorno a esta habilidad del bateador.

Variable	Primeros dos periodos	Periodos restantes	$\Delta\beta_{\text{Edad}}$	p-value
BB	0.003	0.061***	Positivo	0.996
BB ²	0.003	0.071***	Positivo	0.7317
H	-0.005	0.054**	Positivo	0.02113
H ²	-0.004	0.070***	Positivo	0.9072
R	0.003	0.055**	Positivo	1
R ²	-0.002	0.066**	Positivo	0.9085
ER	0.002	0.067**	Positivo	0.9461
ER ²	-0.002	0.070***	Positivo	0.8504
Comando	-0.001	0.083***	Positivo	0.8257
Comando ²	-0.005	0.084**	Positivo	0.9484
Control	-0.001	0.043**	Positivo	0.6326
Control ²	-0.003	0.079***	Positivo	0.000964
Dominio	-0.005	-0.017***	Positivo	8.982e-05
Dominio ²	-0.003	0.003	Positivo	0.003822
IP	-0.002	0.067**	Positivo	0.9163
IP ²	-0.003	0.062**	Positivo	0.6518
GS	-0.004	0.070***	Positivo	0.9072
GS ²	-0.004	0.070***	Positivo	0.9072
L	-0.002	0.060**	Positivo	0.7528

³³Si bien depende de otros bateadores para poder intentar avanzar a otra base, el que llegue a la siguiente base depende principalmente de la habilidad de este

Variable	Primeros dos periodos	Periodos restantes	$\Delta\beta_{\text{Edad}}$	p-value
\bar{L}^2	-0.003	0.067**	Positivo	0.846
ERA	-0.002	0.106***	Positivo	0.1358
ERA ²	0.002	0.090**	Positivo	0.8939
S	-0.004	0.075***	Positivo	0.8224
S ²	-0.004	0.072***	Positivo	0.8863
SO	-0.003	0.066**	Positivo	0.8194
SO ²	-0.001	0.065**	Positivo	0.7941
WAR	-0.002	0.069***	Positivo	0.9259
WAR ²	-0.003	0.080***	Positivo	0.8731
WHIP	0.003	0.048**	Positivo	0.17
WHIP ²	0.007	0.044***	Positivo	0.6765
W	-0.003	0.064**	Positivo	0.9559
W ²	-0.004	0.072**	Positivo	0.4133

Cuadro 11: Estimadores de las regresiones para ambos periodos de los lanzadores iniciales. Se muestra la estadística deportiva asociada a los estimadores, la dirección del cambio del estimador de la edad y el p-value de aplicar un test de Hausman a dichos periodos.

Sin embargo, recalquemos que estas conclusiones depende de asumir que se cumplen los supuestos del estimador en primeras-diferencias, aunque son menos restrictivos que el resto y más eficiente dadas las características de los paneles. Incluso, hay casos como para el estimador *pooling* con los bateadores se obtienen conclusiones totalmente opuestas a las que se mencionaron en este apartado. No obstante, si bien hay comportamientos distintos para el efecto de la edad con el resto de estimadores, no se puede negar que el impacto de la edad cambia con respecto al tiempo.

6. Efectos de convertirse en agente libre

Alguna de las cuestiones más importantes a analizar es si el convertirse en agen libre tiene, en efecto, un impacto positivo en el cambio del poder de negociación, $\Delta\delta$. Para responder esto, estimamos $\Delta\delta$ para el último periodo del jugador antes de convertirse agente libre y el primer periodo como agente libre. En la sección I se pueden apreciar los resultados para los cambios estructurales entre estas estimaciones y las estimaciones de los primeros dos años como agente libre, asumiendo un estimador *pooling*.³⁴

Recordemos que en las estimaciones del problema de indentificación, estimaciones con-

³⁴No se usaron el resto de estimadores o modelos debido a que solo hay un periodo de observación, estos estimadores son equivalente al *pooling* o no se pueden obtener por definición.

juntas y estimaciones entre periodos, independientemente de la posición, la dummy de la edad no fue solo significativa ante el estimador *pooling* y el de efectos aleatorios, sino que tienden a poseer la misma magnitud (de signo positivo). En otras palabras, este resultado es robusto en significancia estadística ante estimadores, medidas de desempeño y posición del jugador. Por lo tanto, afirmar que el convertirse en agente libre tiene un efecto positivo en el cambio en el poder de negociación. Además, si observamos los estimadores de esta dummy para el periodo de transición, se observa que en general es mayor en magnitud que al de resto de estimaciones. Si bien no se puede interpretar cuantitativamente, ordinalmente puede sugerir que hay un mayor impacto en ser agente libre en comparación con el resto de periodos.

Por otro lado, se puede observar que independientemente de la posición, todas las medidas de desempeño (salvo **OBP** para los bateadores y **Comando**²) hay evidencia de un cambio estructural usando el test de Hausman. De esto podemos afirmar que sí hay una diferencia para los estimadores, en general, de $\Delta\delta$ por lo que se refuerza la hipótesis de que convertirse en agente libre tiene un impacto en $\Delta\delta$.

7. Otros resultados

En este apartado se tratarán otros ejercicios con respecto a formas de analizar el cambio en el poder de negociación bajo otros métodos de análisis o escenarios.

7.1. Efectos del COVID-19

Una característica que se observó en los gráficos de series de tiempo para las estadísticas registradas por juegos (sección E.3) fue la presencia de un shock en estas para el año del 2020 que aparentemente cambiaba la tendencia de la serie de tiempo. Sin embargo, recordemos que los páneles con estadísticas registradas por juego no pasaron el proceso *ETL* debido a que no quedaban con observaciones suficientes para tener una muestra significativa³⁵. A pesar de ello, se aplicó una prueba de Hausman en los otros páneles por la presencia de un shock temporal situado únicamente en el 2020, como puede observarse en las gráficas de las series de tiempo de las estadísticas deportivas.

Si bien se citaron algunos estudios que sugieren un impacto negativo en el desempeño de los beisbolistas, esto no nos asegura que exista un cambio estructural entre las estimaciones usando todo el panel y aquellas que excluyen el año de la pandemia (donde ocurrió el encierro mundial). Es sabido que el desempeño de un jugador puede fluctuar de una temporada a otra; entonces nos interesa determinar un cambio estructural de forma conjunta, es decir, donde se consideren distintos aspectos del desempeño deportivo del jugador y no solo un aspecto de su desempeño en partido. Esta es la razón por la que se evaluarán las pruebas de cambio estructural entre los modelos refinados y no los individuales para cada variable de desempeño, como fue el caso del problema de identificación.

³⁵En ambos casos quedaron menos de 15 de observaciones

Los resultados se muestran en la sección K del apéndice. Se puede apreciar que no hay evidencia de ningún cambio estructural para un nivel de significancia del 10 %. Este resultado es robusto ante distintos estimadores y la posición que ocupe el jugador (bateador y/o fildeador), entonces podemos interpretar que de manera conjunta los jugadores pudieron recuperar un nivel de desempeño en el 2020 tal que no representara una gran diferencia comparado con otros años. Sin embargo, esta conclusión se restringe al caso del formato de las estadísticas de nuestro pánels, las cuales son acumulativas.

7.2. Análisis por componentes principales

Por la enorme cantidad de estadísticas deportivas que cuenta nuestros pánels para bateadores y lanzadores es natural que consideremos emplear componentes principales puesto que este método suele reducir la cantidad de variables a solo unas cuentas que logren explicar cierto porcentaje de la varianza entre estas variables (suele elegirse niveles menores al 95 % para evitaar *overfitting*).

Entonces, para la estimaciones hechas con PCA, usaremos las transformaciones individuales de las estadísticas deportivas, es decir, las variables X_{it} asociadas a cada una de las variables deportivas sin importar si es la estadística, r_{it} , o su cuadrado, r_{it}^2 . La única distinción que se hizo para construir los componentes principales es separar las variables del periodo t y las del $t - 1$ ya que uno de los supuestos del modelo empírico es que toda la historia del jugador importa para la dinámica del cambio en el poder de negociación. Por lo tanto, queremos separar los efectos de las variables en distintos periodos.

Sabemos que hay distintos métodos para determinar el número óptimo de componentes a obtener por PCA tales como el método del *codo* y validación cruzada. En nuestro caso, usamos ambos para llegar a un consenso entre los métodos, aunque hubo diferencias de una o dos unidades se obtuvieron dos componentes óptimos para los lanzadores y un componente óptimo para los bateadores, independientemente del periodo.

A pesar de las ventajas operativas del método PCA y el utilizar el óptimo de componentes, no se obtuvieron resultados estadísticamente significativos para todos los estimadores o modelos como en el caso del problema de indentificación (sección J) donde se logró observar la significancia individual de cada medida de desempeño. Si bien los componentes óptimos pueden interpretarse como una variable que toma en cuenta la contribución de todas las estadísticas deportivas en los pánels, ya contamos con variables que hacen este trabajo tal y como es el caso de la variable **WAR**, la cual toma en cuenta la contribución a la victoria del equipo de cada jugador sin importar la posición en la que este juegue; análogamente hay otras variables resumen como el **OPS** o el **WHIP**. Pensamos que es por esta razón que a pesar de haber una gran cantidad de variables, el número óptimo a usar tiende a ser muy bajo, entre uno y dos componentes independientemente de si son del periodo t o $t - 1$, ya que hay variables que por si solas ya explican gran parte de la varianza entre estas.

7.2.1. Diferencias-en-Diferencias

Si bien ya hemos estimado el efecto de ser agente libre en el modelo empírico, también podemos hacerlo mediante herramientas más especializadas como es el caso del método de diferencias-en-diferencias (DID) para estimar el efecto promedio de ser agente libre, es decir, el tratamiento será el convertirse en agente.

Un indicio de si se cumple o no el supuesto de tendencias paralelas es por medio de las gráficas de series de tiempo mostradas en el apéndice. A pesar de haber utilizado los mismos controles que en el problema de identificación para reducir la varianza, no se pudo aislar el efecto de ser agente libre. Los resultados se pueden ver en la sección L del apéndice.

8. Conclusión

En la presente tesis se llevaron a cabo diversos análisis empíricos con el objetivo de entender mejor el cambio en el poder de negociación de los agentes libres de la *MLB* y los factores que lo influyen. A pesar de las limitaciones teóricas en el modelo empírico, se lograron obtener resultados robustos que apoyan la hipótesis de que las medidas de desempeño tienen un impacto significativo en el cambio del poder de negociación de los jugadores. Es importante destacar que, aunque no todas las medidas de desempeño son necesarias para explicar las diferencias en el cambio del poder de negociación, su identificación individual permite a los jugadores enfocarse en aquellas en las que deben mejorar para optimizar su desempeño.

En cuanto a la influencia de la edad, se encontró que esta variable tiene un efecto negativo en el cambio del poder de negociación de los jugadores. Sin embargo, el impacto de la edad varía a lo largo del tiempo y depende de la posición que ocupe el jugador en el campo. Es decir, el efecto de la edad es diferente para un jugador de campo corto que para un lanzador, por ejemplo. Este resultado es importante para los jugadores y sus agentes, ya que les permite entender mejor el impacto que la edad tiene en su carrera y en su poder de negociación.

Por otro lado, se estimó el efecto de convertirse en agente libre en el cambio del poder de negociación, a través de una dummy que indica si el jugador es o no agente libre. Los resultados indican que convertirse en agente libre tiene un impacto positivo en el cambio del poder de negociación de los jugadores. Además, se observó evidencia de cambios estructurales entre las estimaciones del periodo de transición a agente libre y las estimaciones como agente libre.

En general, se concluye que los resultados obtenidos en esta tesis son importantes para los jugadores, sus agentes y para la *MLB* en general, ya que permiten entender mejor los factores que influyen en el cambio del poder de negociación de los jugadores y cómo estos factores varían a lo largo del tiempo y según la posición en el campo. Asimismo, los resultados de esta tesis resaltan la importancia de elegir cuidadosamente los modelos y estimadores utilizados en los análisis empíricos para evitar posibles errores y obtener conclusiones sólidas y robustas.

Apéndice

El propósito de abordar la terminología y reglas que rigen un partido de béisbol en este apéndice es proporcionar al lector una base con la cual pueda entender mejor el análisis, decisiones y conclusiones que se hacen a lo largo de este trabajo. Sin embargo, tampoco se explicarán términos que no sean considerados útiles para estos propósitos.

A. Equipos

En este apartado se muestran los equipos que se ocuparon para el análisis y construcción de los modelos. Los nombres oficiales y más recientes, así como la liga a la que pertenecen (americana o nacional), de cada equipo fueron obtenidos de *ESPN*³⁶, *Baseball-reference*³⁷ y *Spotrac*³⁸. Por otro lado, para mantener consistencia con la fuente de las bases de datos principal (*Spotrac*), solo se optó por dejar la abreviación dada por esta página para *Washington Nationals*.

■ National League East:

- **CHC**: Chicago Cubs
- **CIN**: Cincinnati Reds
- **MIL**: Milwaukee Brewers
- **PIT**: Pittsburgh Pirates
- **STL**: St. Louis Cardinals

■ National League Central:

- **BAL**: Baltimore Orioles
- **BOS**: Boston Red Sox
- **NYM**: New York Yankees
- **TB**: Tampa Bay Rays
- **TOR**: Toronto Blue Jays

■ National League West:

- **ARI**: Arizona Diamondbacks
- **COL**: Colorado Rockies
- **LAD**: Los Angeles Dodgers
- **SD**: San Diego Padres
- **SF**: San Francisco Giants

■ American League East:

- **HOU**: Houston Astros
- **LAA**: Los Angeles Angels
- **OAK**: Oakland Athletics
- **SEA**: Seattle Mariners
- **TEX**: Texas Rangers

■ American League West:

- **ATL**: Atlanta Braves
- **MIA**: Miami Marlins
- **NYM**: New York Mets
- **PHI**: Philadelphia Phillies
- **WSH**³⁹: Washington Nationals

■ American League Central:

- **CHW**: Chicago White Sox
- **CL**: Cleveland Guardians
- **DET**: Detroit Tigers
- **KC**: Kansas City Royals
- **MIN**: Minnesota Twins

³⁶<https://www.espn.com/mlb/teams>

³⁷https://www.baseball-reference.com/about/team_IDs.shtml

³⁸<https://www.spotrac.com/mlb/free-agents/>

³⁹La abreviación actual y oficial es *WSN*

Todos los equipos se encuentran en Estados Unidos de América, a excepción de *Toronto Blue Jays*.

B. Posiciones de jugadores

A continuación, se ofrece con detalle la explicación de las tareas que se tienen que desempeñar en cada una de las posiciones que un jugador pueda ocupar durante un partido. Así mismo, también se proporcionan los nombres y abreviaturas en el idioma original debido a que en la mayoría de las fuentes citadas dichos términos están en dicho idioma, es decir, el inglés. El número que identifica a cada jugador (Fig.B.1) está dado por la enumeración de la lista.

B.1. Fildeadores

Los fildeadores, mejor conocidos en inglés como *fielders*, son quienes juegan a la defensiva, teniendo como objetivo evitar que el equipo contrario realice carreras. Consta de nueve tipo de posiciones distintas donde cada una se puede identificar con un número entre el 1 y el 9 tal y como se muestra en la *Figura* (B.1).

■ Batería:

1. **Lanzador/Pitcher (P)**: Se encarga de lanzar la bola hacia el receptor con el objetivo de marcar *strike outs* para terminar el *inning* cuanto antes.
 - **Lanzador inicial/Starting pitcher (SP)**: Aquel jugador que empieza el partido como lanzador.
 - **Lanzador de relevo/Relief pitcher (RP)**: Es quien releva al lanzador inicial. Dependiendo de en qué punto del partido se quiera hacer el relevo, el *relief pitcher* se puede dividir en dos subcategorías: *Middle relief pitcher* (MRP) o *long relief pitcher* (LRP).
2. **Receptor/Catcher (C)**: Se le denomina así al jugador que yace agachado atrás del bateador y adelante del árbitro (*umpire*). Su función es atrapar las bolas que lance el *pitcher*. Además, también puede hacer *tag outs* para evitar que el equipo contrario anote carreras.

■ Infield⁴⁰:

3. **Primera base/First baseman (1F)**: También conocido como *inicialista*, es aquel jugador situado en la primera base que se encarga de evitar que los jugadores de la ofensiva lleguen a dicha base.
4. **Segunda base/Second baseman (2F)**: Similar a la posición anterior, este jugador no solo se encarga de evitar que lleguen a su base (la segunda) o hacer dobles, sino que también de cortar los tiros del jardinero derecho para hacer más rápidos los *outs*. Así mismo, puede o no estar situado en la segunda base, depende de la estrategia que siga con el campo corto (*shortstop*).

⁴⁰También existe la subcategoría abreviada **CI** que engloba a cualquier jugador que sea primera o tercera base.

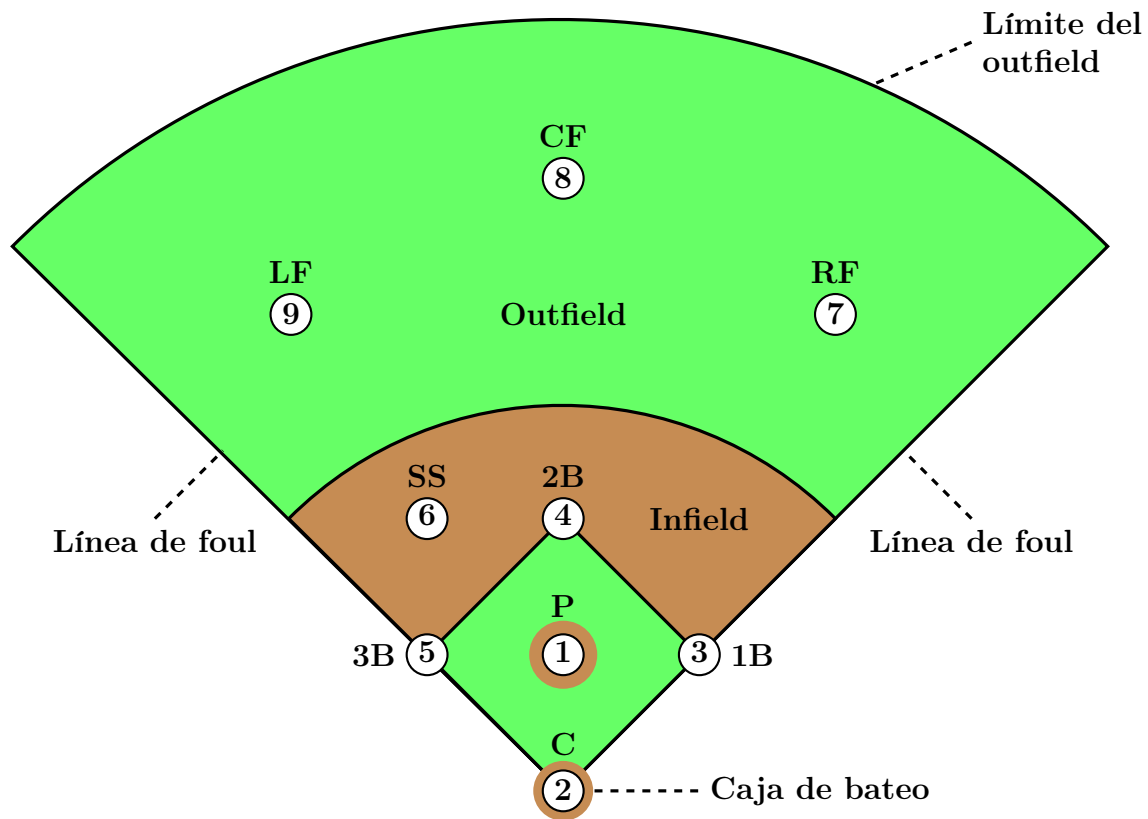


Figura 5: Configuración usual de las posiciones de los jugadores en un campo de béisbol. [MLB \(2023\)](#)

5. **Tercera base/Third baseman (3F)**: Es el defensivo que ocupa la tercera base, conocido también como *antesalista*, cuya función es evitar las bolas bateadas pasen al jardín izquierdo y contribuir a marcar *outs*.
6. **Campo corto/Shortstop (SS)**: Este jugador se halla entre la segunda y la tercera base. Las tareas que desempeña son las mismas que los jugadores de la segunda y tercera base. Sin embargo, cabe mencionar que esta posición es más exigente puesto que el espacio desde donde tiene que cortar tiros es grande en comparación con el resto de jugadores en el *infield*.

■ **Outfield**⁴¹:

7. **Jardinero derecho/Right fielder (RF)**: Situado en el jardín derecho. Es el defensivo que se encarga de atrapar los batazos altos en su lado respectivo del campo.
8. **Jardinero central/Center fielder (CF)**: Análogo a la posición anterior, solo que este jugador defiende la parte central del jardín.

⁴¹Estos jugadores también se identifican bajo la abreviatura **OF** que denota *outfield*.

9. **Jardinero izquierdo/Left fielder (LF)**: Similar al resto de posiciones en el *outfield*, este jardinero se encarga del jardín izquierdo.

B.2. Bateadores

También conocidos como *hitters* o *batters* (**H**), situados en la caja de bateo, juegan en la ofensiva y se encargan de conectar *hit* con el bate. Cabe mencionar que en la base de datos también aparece una posición derivada llamada *bateador designado* (**DH**), son jugadores que reemplazan durante varias entradas en la ofensiva a otros para batear, por lo general a *pitchers*⁴². Sin embargo, el bateador designado no puede jugar en ninguna posición en la defensiva.

⁴²Actualmente hay una gran controversia ya que esta posición afecta el partido debido a que, en agregado, los lanzadores batean peor que un jugador promedio, es decir, realizan pocos *hits*.

C. Medidas de desempeño

En este apartado se explica el significado de cada una de las medidas de desempeño⁴³ contenidas en las bases de datos empleadas para los modelos tanto para los fildeadores como los bateadores⁴⁴. Así mismo, se dividen estas medidas de desempeño en las que son medidas de forma individual y las que son medidas de forma agregada por equipo.

C.1. Medidas de desempeño deportivas

Las medidas de desempeño deportivas presentes en la base de datos son las que se muestran en las siguientes listas⁴⁵: A continuación, se ofrece una descripción de cada una de estas variables

- **Partidos jugados (Games played - GP)**: Son la cantidad de juegos en los que ha participado un jugador a lo largo de la temporada. Para que se considere que un jugador participó en un partido este tiene que aparecer en cualquier punto del partido sin importar si es un inicialista (*starter*) o reemplazo.
- **Porcentaje de partidos jugados (Percentage of games played - GP %)**: Son la cantidad de juegos jugados por el jugador entre la cantidad de juegos totales que hubo en la temporada⁴⁶.
- **Victorias por encima del nivel de reemplazo (Wins above replacement - WAR)**⁴⁷: Es la estadística más importante para evaluar a cualquier jugador en varios aspectos de su desempeño independientemente de la posición que este ocupe. Sirve para determinar si el impacto del jugador fue positivo o negativo para el equipo en comparación con el caso donde algún otro jugador lo reemplazara en dicho equipo⁴⁸. Se hacen dos distinciones a partir de las posiciones

- **Bateadores:**

$$\begin{aligned} \text{WAR} = & \text{RBI} + \text{Fielding Runs Above Average} \\ & + \text{Positional Adjustment} + \text{League Adjustment} + \text{Base Running Runs} \\ & + \frac{\text{Runs Added or Lost Due to Grounding into Double Plays}}{\text{Runs Per Win}} \end{aligned}$$

- **Fildeadores:** Existen dos variantes:

⁴³Las definiciones se obtuvieron del glosario oficial de la MLB (<https://www.mlb.com/glossary>) y el diccionario llamado *The New Dickson Baseball Dictionary* Dickson (1999)

⁴⁴Se omitió la explicación de las variables que son transformaciones cuadráticas de otras estadísticas puesto que su interpretación se obtiene de manera directa por medio de la estadística deportiva empleada.

⁴⁵En caso de existir alguna traducción oficial al castellano se usará; en caso contrario, se dejará el nombre en inglés.

⁴⁶En una temporada típica hay 162 partidos por jugador, pero un jugador puede jugar más de esta cantidad cuando hay desempates, es decir, 163 partidos.

⁴⁷En inglés también es conocida como *wins above replacement player*, abreviada como WARP.

⁴⁸Se suelen usar el promedio estadísticas deportivas de la temporada en la que se quiere calcular el WAR. Así mismo, el jugador reemplazo se limita, dependiendo del caso, a la posición y liga del jugador en cuestión.

Bateadores	Fildeadores
<ul style="list-style-type: none"> • Al bate • Bateos conectados • Bateos promedio • Carreras implusadas • Dobles • <i>Home runs</i> • Juegos iniciados • <i>On base plus slugging percentage</i> • Porcentaje de embasado • Porcentaje de partidos jugados • Porcentaje de juegos iniciados • Porcentaje de <i>slugging</i> • Triples • Victorias por encima del nivel de reemplazo (<i>Wins above replacement</i>) 	<ul style="list-style-type: none"> • Base por bola en contra. • Bateos permitidos. • Carreras permitidas. • Carreras limpias. • Comando. • Control. • Dominio. • <i>Inning pitched</i>. • Juegos iniciados. • <i>Losses</i>. • Partidos jugados. • Porcentaje de juegos iniciados. • Porcentaje de carreras limpias. • Porcentaje de partidos jugados. • Salvamentos. • <i>Strike-outs</i>. • Victorias por encima del nivel de reemplazo. • <i>Walks and Hits per Inning Pitched</i>. • <i>Wins</i>.

Cuadro 12: Medidas de desempeño deportivas en las bases de datos.

- **RA9 WAR:** Se enfoca en las carreras que el lanzador de turno permite por entrada (*inning*):

$$\text{RA9 WAR} = \frac{\text{Runs Saved Above Average}}{\text{Runs Per Win}}$$

- **WAR:** Tienen un enfoque sobre más estadísticas con el propósito de observar las *carreras salvadas arriba del promedio*:

$$\text{WAR} = [\text{IP} \times [\text{League ERA} - (\text{ERA} + \text{Park Factor})] + \frac{\text{Defense-Independent Pitching Statistics}}{\text{Runs Per Win}}]$$

La ventaja de esta estadística es que puede tomar en cuenta las distintas medidas de desempeño sin importar la posición que este ocupe. Sin embargo, se debe de tomar

en cuenta la liga del jugador, entres otros factores que se indican en la fórmula correspondiente.

C.1.1. Bateadores

- **Al bate (At-bats - AB)**: Son el número de veces que el bateador aparece en el plato (*plate appearance*) menos los sacrificios, bases por bola o *hits by pitch*. Un jugador se acredita con esta estadística cuando llega a alguna base a través de un *fielder's choice*, bateo, error⁴⁹ o cuando el bateador es puesto en una jugada de no-sacrificio.
- **Bateos conectados (Hits - H o HC)**: Es la cantidad total de veces que el *hitter* batea una bola y puede llegar a una base. Sin embargo, si llega a la base mediante un error de algún fildero o por una jugada de sacrificio, no se acreditará con esta estadística.
- **Bateos promedio (Average bating - BA o AVG)**: Se define como el número de bateos entre la cantidad dada por los al bates. Se obtiene con la siguiente fórmula

$$AVG = \frac{H}{AB}$$

Esta estadística es usada frecuentemente como comparación entre jugadores para evuluar su capacidad de bateo.

- **Carreras impulsadas (Run batted in - RBI)**: Esta estadística se le acredita a un bateador cuando se anota una carrera como consecuencia de su aparición al plato. Las excepciones ocurren si la carrera es anotada mediante un error o un *double play* forzado.
- **Dobles (Doubles - 2B)**: Se cuenta como doble cuando al batear la bola, el *hitter* es capaz de llegar a segunda base. Si llega a segunda base por un error o es puesto en *out*, no se le otorga el doble al bateador.
- **Home runs - HR**: Cualquier bateo que haga el *hitter* que le permita pasar por las cuatro bases que no sea mediante algún error o se haya registrado un *out* en el proceso. Hay tres formas principales de conseguir hacer un *home run*:
 - Cuando la bola bateada pasa la barda del *outfield* o le pega a algunos de los postes del *outfield*.
 - Si el bateador realiza una jugada de tal suerte que puede pasar por las cuatro bases. Este tipo de eventos pasan muy pocas veces.
 - Puede ocurrir cuando una bola justa (*fair ball*) rebota del campo a territorio fuera de este (sin importar cómo).
- **Juegos iniciados (Games started - GS)**: Esta medida de desempeño se le otorga a un bateador cuando forma parte de la alineación principal (*starting lineup*) en un partido y tiene una aparición en el plato.

⁴⁹La excepción ocurre cuando el error es debido a la interferencia del receptor.

- ***On-base plus slugging percentage* - OPS**: Es la suma de los porcentajes dados por **OBP** y **SLG**, es decir,

$$OPS = OBP + SLG$$

A diferencia de las otras estadísticas deportivas que se refieren a porcentajes, esta está acotada entre el cero y el dos, $OPS \in [0, 2]$.

- **Porcentaje de embasado (*On-base percentage* - OBP)**: Esta estadística es usada frecuentemente para describir la propensión de un *hitter* para llegar a una base cuando tiene un turno al bate (**AB**). Se define como la cantidad de veces que un bateador avanza a alguna base como resultado de realizar un bateo conectado (**HC**), una base por bola (**BB**), ser golpeado por el lanzador (**HBP**) entre la suma de los turnos al bate (**AB**), los *fly* de sacrificio (**SF**), los **BB** y los **HBP**. Equivalentemente se puede escribir como la siguiente expresión

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF}$$

- **Porcentaje de juegos iniciados (*Games started percentage* - GS%)**: Se define como los juegos iniciados por temporada, es decir, con respecto a los juegos totales en los que participo el bateador. Se define matemáticamente como

$$GS\% = \frac{GS}{GP}$$

- **Porcentaje de *slugging* (*Slugging percentage* - SLG o SA)**: Este promedio se define como el total de bases a las que llega un ofensivo (**TB**) entre el total de turno al bate (**AB**), es decir,

$$SLG = \frac{TB}{AB}$$

Por lo general, jugadores con mayor **SLG** tienden a anotar más dobles, triples o *home runs*.

- **Triples (*Triples* - 3B)**: Es completamente análoga a la estadística de los dobles (**2B**) salvo que para acreditarse con un triple, el bateador tiene que llegar a tercera base.

Adicionalmente, se definen otras medidas de desempeño que no se hayan en el panel de los bateadores, pero que sirven como auxiliares para otros términos de manera más intuitiva.

- **Base por bola (*Walks* - BB)**: Cuando el *pitcher* lanza cuatro bolas malas (*balls*) mientras un jugador ofensivo está en su turno al bate, entonces, el bateador avanza a primera base. Sin embargo, no se le acredita como base, sino como base por bola (*walk*).
- **Carreras (*Runs* - R)**: Si un jugador de la ofensiva que ya se encontrase en alguna de las bases cruza el plato, se considera como una carrera. Cabe mencionar que no importa cómo cruce el plato, ya sea a través de una carrera limpia o por algún error de los fildeadores.

C.1.2. Fildeadores

- **Carreras limpias (Earned runs - ER)**: Se define como la cantidad de carreras (**R**) anotadas en contra del lanzador en turno donde no hay beneficio de error (**E**) o un *passed ball* (**PB**).
- **Comando (Command)**: Esta estadística refleja la capacidad del lanzador para lanzar bolas precisas, es decir, poder lanzar cuando quiere dentro o fuera de la zona de *strike*. Matemáticamente se obtiene de la siguiente manera

$$\text{Comando} = \frac{\text{SO}}{\text{BB}}$$

Entre mejor sea el comando de un lanzador, menos bases por bola (**BB**) alcanzarán los bateadores en turno y mayor será la cantidad de jugadores que saque mediante *strike outs* (**SO**).

- **Control (Control)**: Es la habilidad del lanzador para evitar que algún ofensivo alcance una base por medio de bases por bola (**BB**). Se calcula mediante la siguiente fórmula

$$\text{Control} = \frac{\text{BB}}{\text{IP}}$$

Esta medida de desempeño contribuye a obtener el **WHIP** por lo que se usa más esta última debido a que ofrece un panorama más amplio sobre el perfil del jugador defensivo.

- **Dominio (Dominance)**: Se entiende como la capacidad de un lanzador para sacar a bateadores del juego mediante *outs* (**O**). Se define como la cantidad de *strike outs* (**SO**) en comparación con las entradas en las que pichea, es decir, los *innings pitched* (**IP**):

$$\text{Dominio} = \frac{\text{SO}}{\text{IP}}$$

es decir, se define como un promedio entre los *strike outs* (**SO**) y los *innings pitched* (**IP**).

- **Innings lanzado (Innings pitched - IP)**: Son el número de entradas (*innings*) que un lanzador permanece en el campo. Para que se le registre esta medida de desempeño el *pitcher* tiene que marcar al menos tres *strike outs*. Cada *SO* define un tercio de la entrada, es decir, un tercio de un *inning pitched*.
- **Juegos iniciados (Games started - GS)**: Son el número de partidos en una temporada donde el *pitcher* fue el primero en lanzar una bola para su equipo. No importa si esa bola lanzada marca un *out* (**O**) o no.
- **Losses - L**: Se le atribuye un *loss* a algún lanzador a quien le anotaron una carrera en su turno en la defensiva. Esta carrera tiene que ser tal que pone a la cabeza al equipo contrario y mantienen la ventaja el resto del partido. Cabe mencionar que no es una medida adecuada para evaluar el desempeño de algún *pitcher* puesto que también se le puede otorgar esta estadística a algún lanzador con un buen desempeño, pero tener un equipo defensivo malo.

- **Promedio de carreras limpias (*Earned run average* - *ERA*):** Promedio de las carreras ganadas con respecto a las entradas picheadas (**IP**), es decir,

$$ERA = \frac{ER}{IP}$$

- **Salvamentos (*Saves* - **S**):** Se le otorga a un lanzador de relevo (*relief pitcher*) que acaba el juego estando de parte del equipo ganador. Sin embargo, se tienen que satisfacer las siguientes condiciones para que se le conceda un salvamento:

- Debe ser el último lanzador en el juego: El lanzador debe entrar en el juego después del último lanzador del equipo contrario y trabajar por lo menos un tres entradas. Cuando entra al partido su equipo no tiene que poseer una ventaja mayor a tres carreras (**R**).
- Debe mantener la ventaja: El lanzador debe mantener la ventaja del equipo durante su aparición, es decir, no permitir que el equipo contrario empate o adelante en el marcador.
- Debe retirarse sin perder la ventaja: El lanzador debe retirarse del juego sin perder la ventaja y sin permitir que el equipo contrario empate o adelante en el marcador.

La estadística de salvamento es un indicador importante del rendimiento de un lanzador en situaciones de presión, ya que refleja su capacidad para mantener la ventaja y ayudar a su equipo a ganar juegos.

- ***Strike outs* - **SO** o **K**:** Se registra cuando el lanzador hace cualquier combinación exitosa de tres *swings* o *strikes*. Como consecuencia se marca un *out* al bateador y este queda fuera de la jugada.
- ***Walks and hits per inning pitched* - **WHIP**:** Se define como el promedio del número de bases por bolas (**BB**) más bateos permitidos (**HP**) por cada *inning* lanzado (**IP**).

$$WHIP = \frac{BB + HP}{IP}$$

El **WHIP** es una de las estadísticas más usadas para evaluar el desempeño general de algún lanzador.

- ***Wins* - **W**:** Esta medida de desempeño se le otorga a un lanzador si su equipo gana el partido y, además, este estuvo picheado la mayor parte del partido. En un juego tradicional de nueve entradas (*innings*), el *pitcher* tendría que estar presente en el campo un mínimo de cinco entradas. rel

Las siguientes medidas de desempeño aparecen en las bases de datos de los lanzadores, pero se refieren a estadísticas anotadas en contra del lanzador en turno.

- **Bases por bola (*Walks* - **BB**).**

- Bateos (**Hits - H**)⁵⁰.
- Carreras (**Runs - R**).

C.1.3. Definiciones misceláneas

Debido a que en muchas de las definiciones dadas en las subsecciones usan términos tecnicismos del béisbol es que se crea esta sección, para proporcionar un mayor entendimiento sobre las medidas de desempeño. Por otro lado, también se abordarán términos que conciernen a las reglas y dinámica de un partido de béisbol puesto que se considera más apropiado abordarlos en esta subsección

- **Bolas:** Cuando un *hitter* batea una bola, esta solo podrá ser válida (*fair ball*) si cae en el área que está entre las *líneas de foul* (Fig.B.1). De lo contrario, se denominará como *foul*. Hay algunas definiciones derivadas que tienen relación con otros conceptos presentados en estas sección.
 - **Passed ball - PB:** Es un lanzamiento por parte del *pitcher* hacia el receptor, pero que comete algún fallo al intentar atraparla. Como consecuencia, permita que algún jugador en la ofensiva avance a primera base o cualquier otra base.
 - **Bola mala (Balls):** Ocurre cada que el *pitcher* lanza una bola a una zona fuera de *strike* y el bateador no realiza un *swing*, es decir, no intenta batear la bola.
 - **Bola justa (Fair ball):** Cualquier bola bateada que haya caído primero en cualquier zona válida del campo (entre las líneas de *foul*).
- **Error (Error - E):** Este término se usa en particular cada que por culpa de una mala maniobra, por parte de la defensiva, permite que un jugador de la ofensiva llegue a una base adicional. Por ejemplo, cuando el *pitcher* lanza una bola a una área que no sea la zona de *strike* y el *hitter* no batea, entonces se le permite al jugador de la ofensiva avanzar a la siguiente base (*walk*). También ocurre si el bateador es golpeado directamente por alguna bola que haya picheado el lanzador. Estos dos casos son los únicos que tienen relación con las medidas deportivas en la base de datos usada.
- **Fielder's choice - FC:** Se refiere a una situación en la que algún jugador defensivo tiene la opción de elegir retirar a un corredor en base o permitir que otro corredor avance. Esto puede ocurrir cuando el bateador se pone en *out*, pero el defensa elige retirar a otro corredor en lugar del bateador, lo que permite que un corredor adicional avance a una base. El *fielder's choice* no se registra como un bateo (**H**) o un error (**E**), ya que el bateador no está eligiendo dónde correr, sino que está en el bateador y depende del defensa para retener o no. En cambio, el *fielder's choice* se registra en el juego como una "jugada de corredor" contribuye a las estadísticas de bateo y carreras.

⁵⁰También se conocen como bateos/*hits* permitidos (**HP**).

- **Outs - O**: Es cualquier jugada en la que el árbitro (*umpire*) determina que el jugador de la ofensiva se retira. Existen dos categorías:
 - **Out por regla**: Ocurre cuando el jugador de la ofensiva incumple con el reglamento del béisbol.
 - **Out forzado**: Se hacen una vez el *hitter* haya bateado la bola exitosamente. Hay tres maneras de realizar este tipo de *outs*:
 - **Pop out - PO**: Si el *hitter* batea y la bola es atrapada por alguno de los base (jugadores en el área *infield*) sin que esta haya rebotado previamente, entonces se denomina como *pop out*.
 - **Fly out**: Es análogo al *pop out* salvo la diferencia que en este caso quienes la atrapan son los jardineros, es decir, los jugadores que se encuentran en la parte del campo llamada *outfield*.
 - **Tag out**: Para que este tipo de *out* suceda, se tiene que cumplir que el *hitter* batee y la bola aterrice en territorio válido sin que ningún fildeador la haya atrapado. Entonces, si alguno de los jugadores en la defensiva, con la bola en la mano, toca al jugador de la ofensiva que intenta llegar a base, este último sale del juego por *tag-out*.
 - **Put out - P**: Similar a la estadística anterior. Sin embargo, esta se registra cuando el base, con bola en mano, llega primero a la base antes que el ofensivo que intenta alcanzar a dicha base.
- **Sacrificio (*Sacrifice* - S)**: Es una acción en la que un jugador da su turno al bate para permitir que un compañero de equipo avance a una base. Una jugada de sacrificio se registra como un *out* (**O**) en el registro personal del jugador que hizo la jugada, pero no afecta negativamente la estadística de carreras limpias del lanzador (**ER**). Las jugadas de sacrificio son una estrategia común en el béisbol, especialmente cuando un equipo está tratando de anotar una carrera (**R**) con un hombre en base y sin *outs* (**O**).

D. Variables salariales

A pesar de contar con bases de datos especializadas en agentes libres, las variables con respecto a los salarios son las mismas. Esta es la razón por la que se explicarán dichas variables en la presente sección donde se señalarán aparte los datos que conciernen a la agencia libre.

- **Efectivo (*Cash*)**: Cantidad de dinero total en efectivo que recibe un jugador en un año dado.
- **Ganancias (*Earnings*)**: Cantidad de dinero en efectivo que recibe (sin importar de qué equipo provenga el dinero) más cualquier otra compensación salarial, aunque no sea en efectivo.
- **Salario base (*Base salary*)**: Es el salario que recibe cualquier jugador de la *MLB*, pero no se contemplan bonos por firmar u otro tipo de compensaciones.

- **Salario regular (*Payroll salary*)**: Es el salario total que recibe un jugador, toma en cuenta el salario base, bonos por firma o cualquier otra compensación salarial que este reciba.
- **Salario ajustado (*Adjusted salary*)**: Se refiere al salario base, pero ajustado por cambios en la inflación o costos de vida. Ayuda a ver el valor de un jugador a lo largo del tiempo.
- **Bono por firmar (*Signing bonus*)**: Es un pago único que se le da a un jugador por firmar el contrato con cierto equipo.
- **Salario % (*Salary %*)**: Representa el porcentaje del valor total del contrato que recibe en un año el jugador como salario.
- **Valor total del contrato (*Total value of player's contract - TVS*)**: Es el valor total del contrato que tiene un jugador, también se incluyen bonos por firmar o cualquier otro tipo de compensación extra en forma de dinero. Esta valor se encuentra entre el 0 y 1 ya que se obtiene al normalizar la base de datos del año en cuestión.

Algunas variables con respecto a los agentes libres se omitieron en el apéndice puesto que su significado se puede obtener completamente a partir del nombre. Estas variables omitidas se muestran en la sección sobre la base de datos para las bases de los agentes libres y de los salarios.

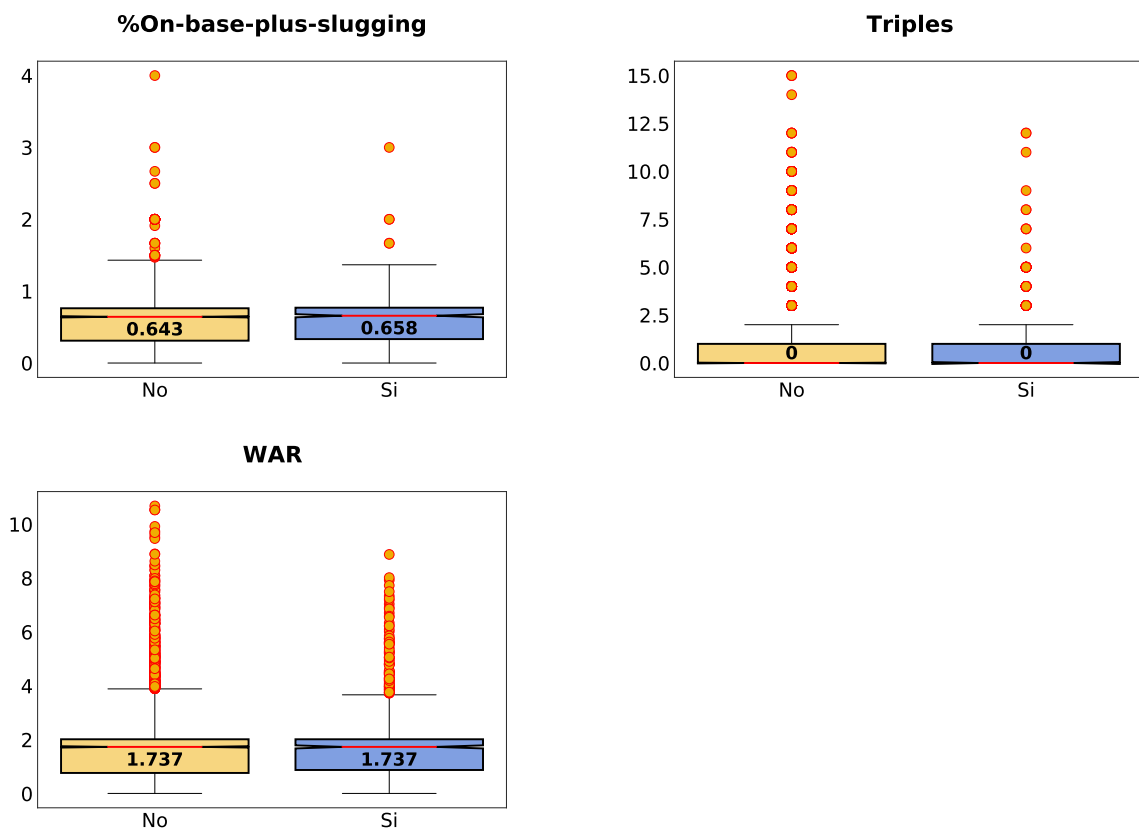
E. Estadísticas descriptivas

A lo largo de esta sección se mostrarán distintos gráficos relacionados a las medidas de desempeño. Por cuestiones de espacio, no se muestran todas las estadísticas deportivas. Sin embargo, se pueden consultar en el siguiente enlace [Estadísticas descriptivas completas](#).

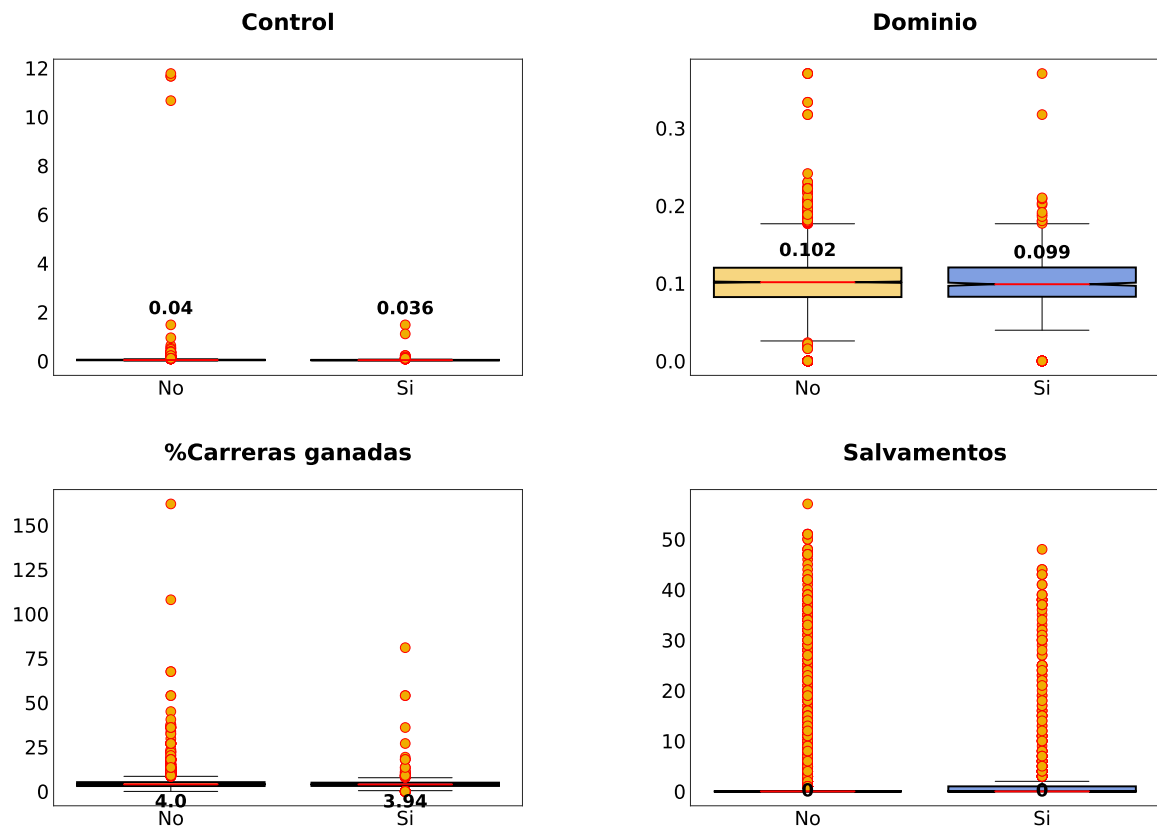
E.1. Diagramas de caja

Comparación con diagramas de caja entre los jugadores que son agentes libres (*Si*) y los no-agentes libres (*No*) con respecto a las estadísticas deportivas, salarios y datos generales. Se tomaron los promedios a lo largo de todos los años que abarca la base de datos.

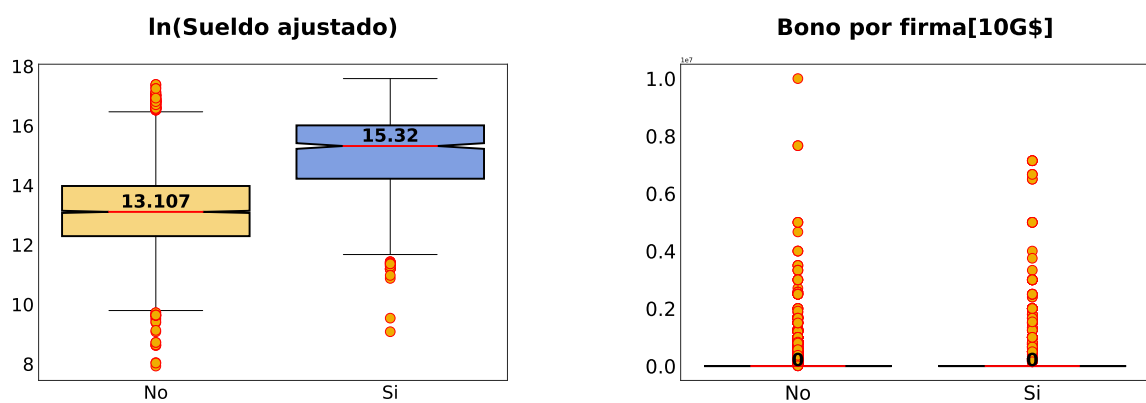
E.1.1. Estadísticas deportivas acumuladas - Bateadores

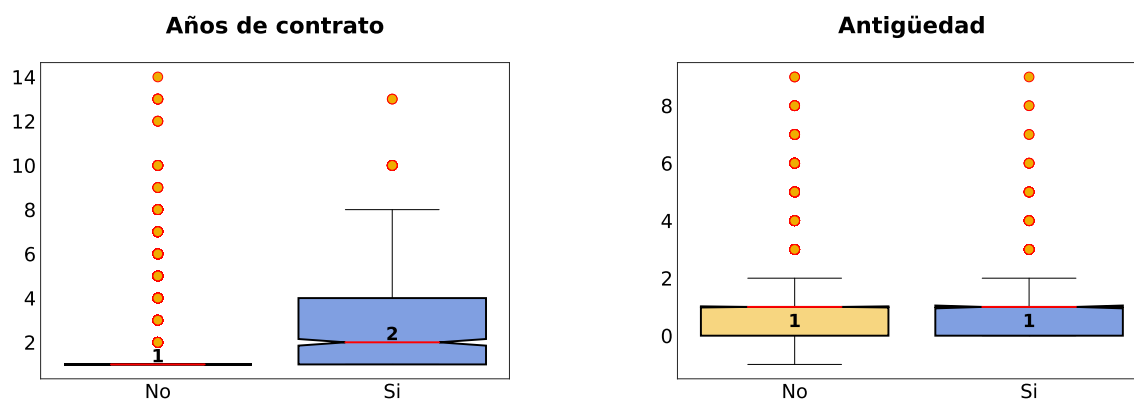


E.1.2. Estadísticas deportivas acumuladas - Filderos

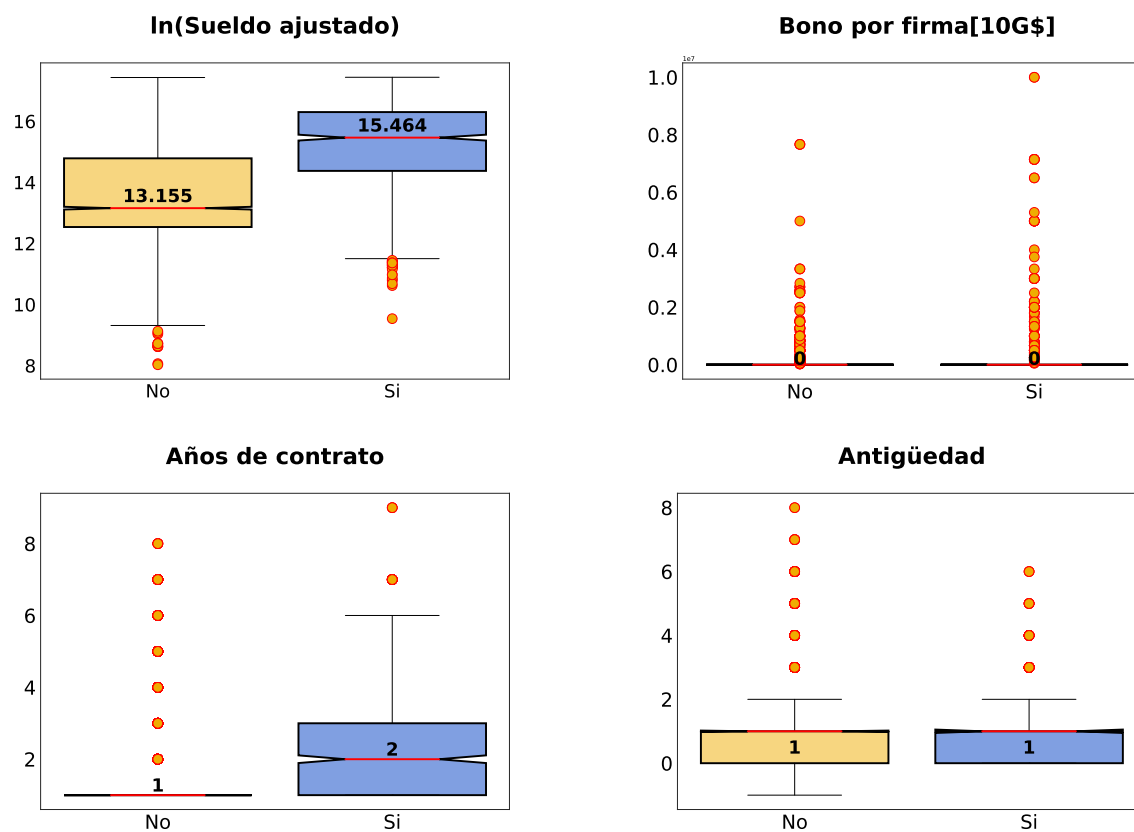


E.1.3. Datos contractuales - Bateadores

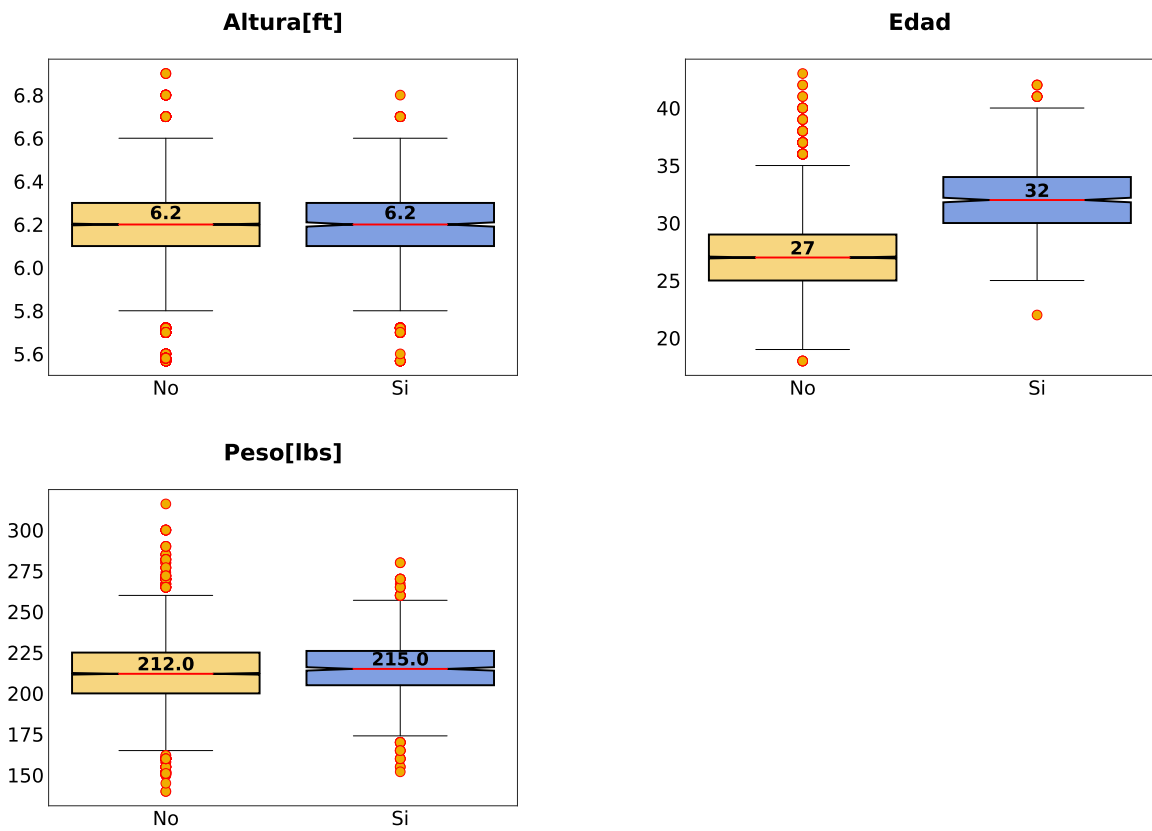




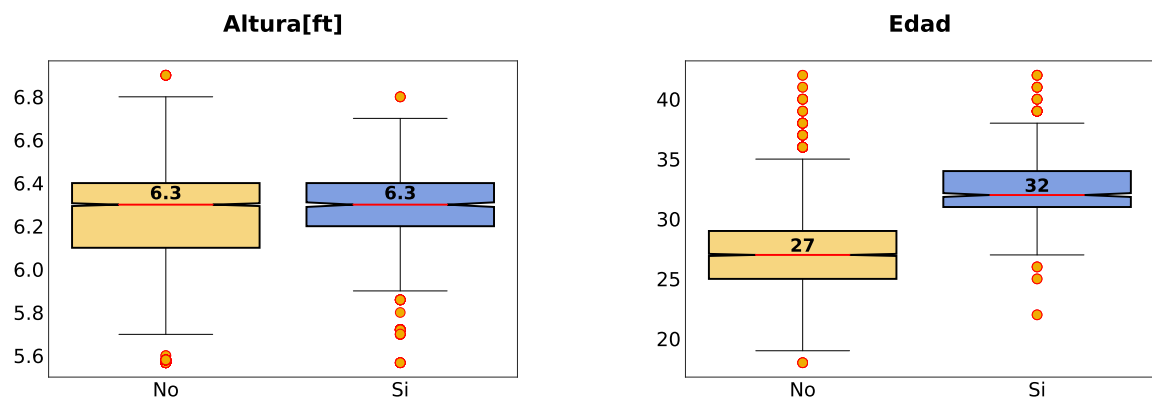
E.1.4. Datos contractuales - Filderos

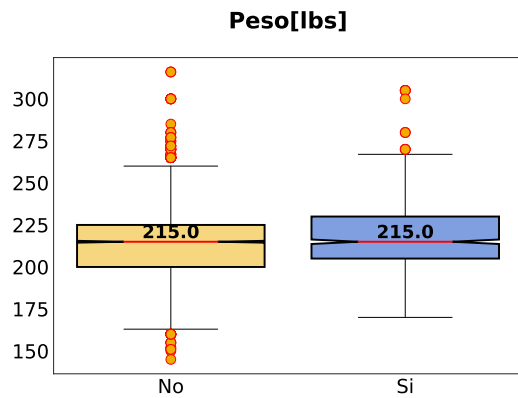


E.1.5. Datos generales - Bateadores



E.1.6. Datos generales - Filderos

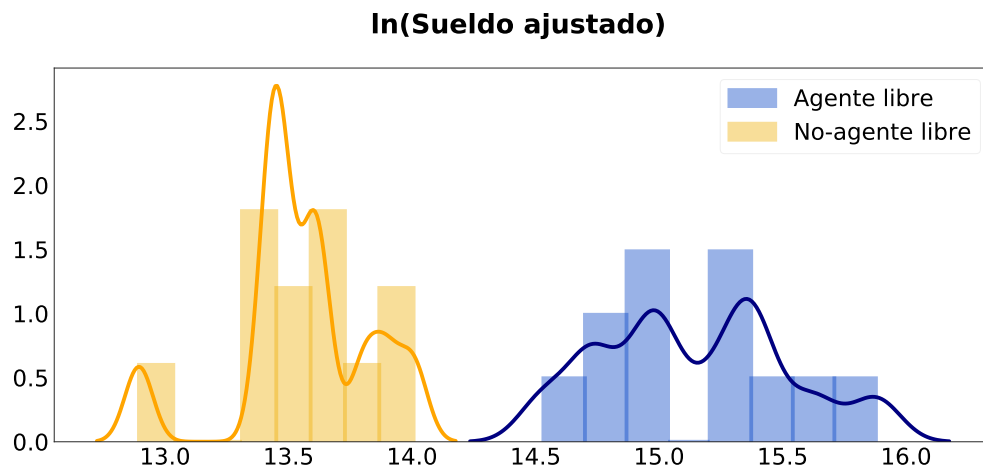




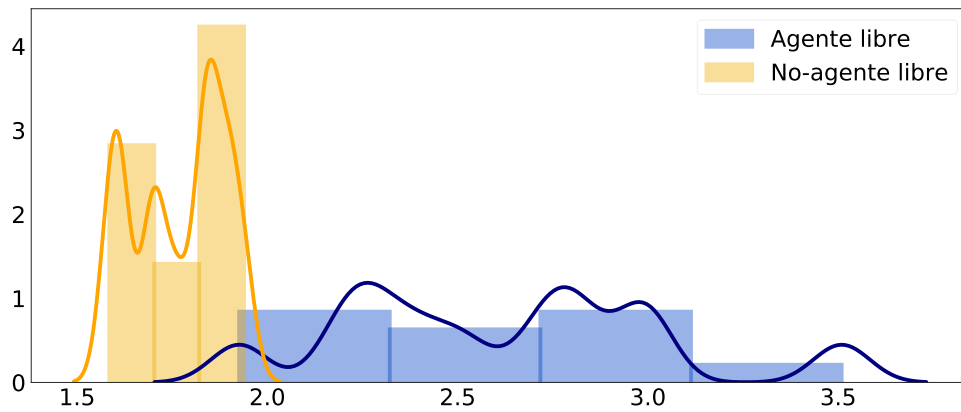
E.2. Diagramas de densidad

De manera similar a la sección sobre los diagramas de caja, los jugadores que son agentes libres se etiquetan con un *Sí* y, en caso contrario, con un *No*. Se muestran los diagramas de densidad de las estadísticas deportivas, salarios y datos generales. Se tomaron los promedios de todos los años que abarca la base de datos.

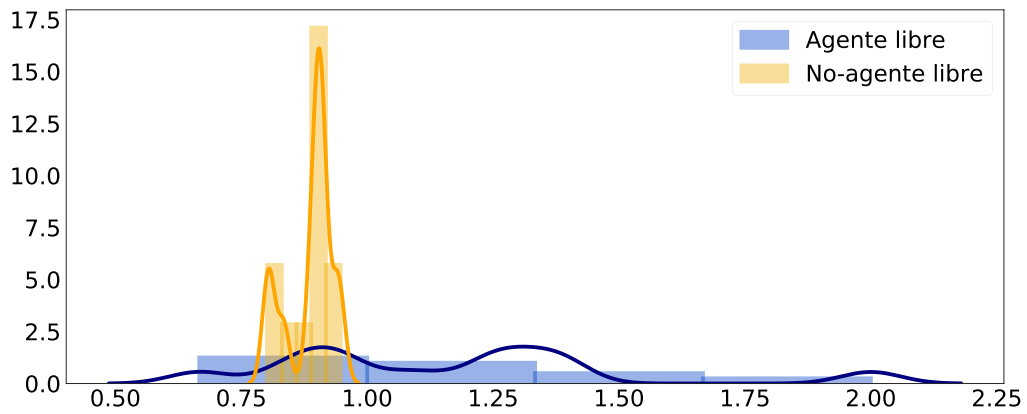
E.2.1. Datos contactuales - Bateadores



Años de contrato

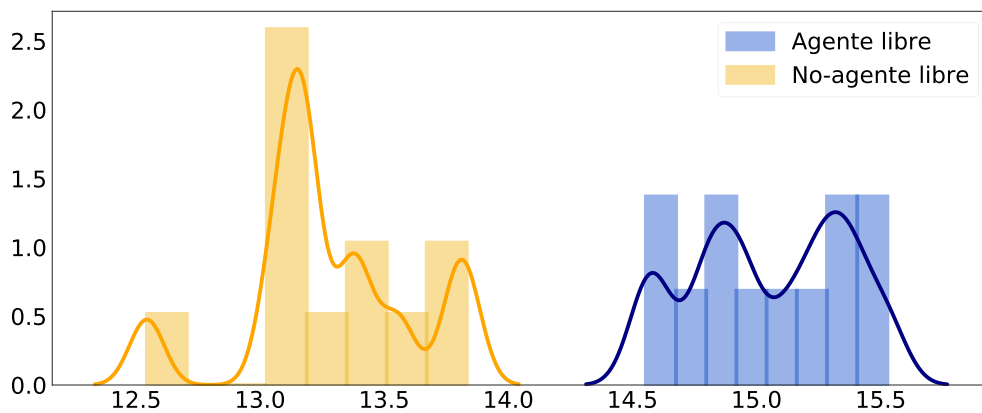


Antigüedad

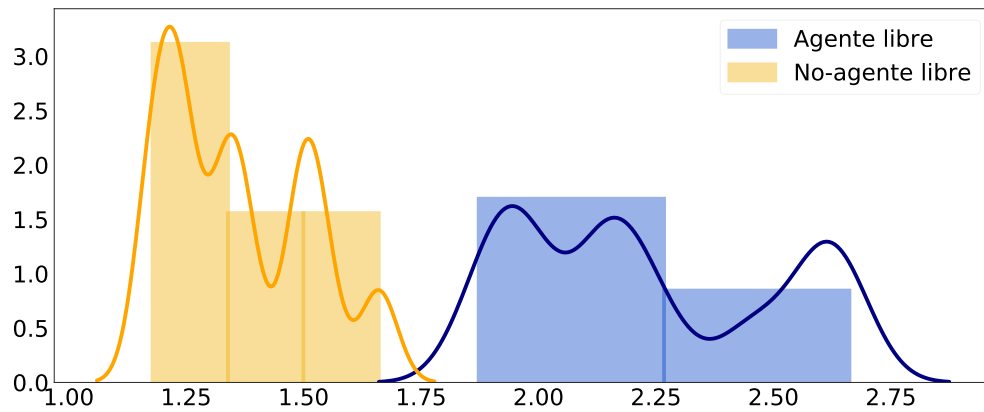


E.2.2. Datos contactuales - Filderos

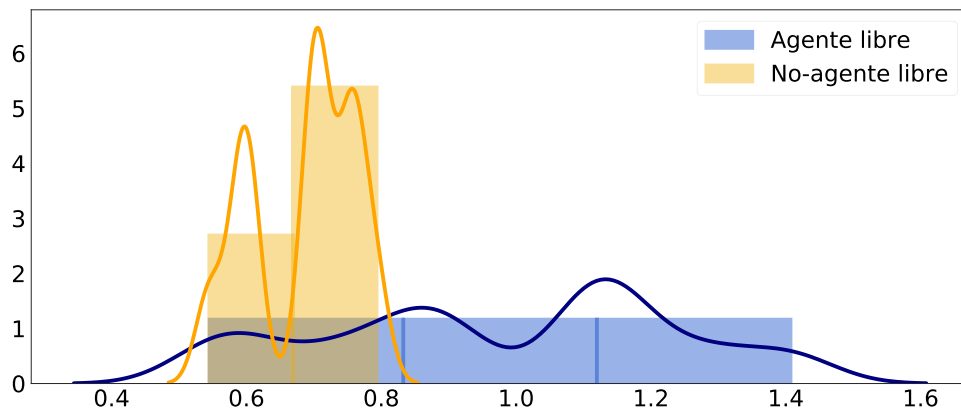
ln(Sueldo ajustado)



Años de contrato

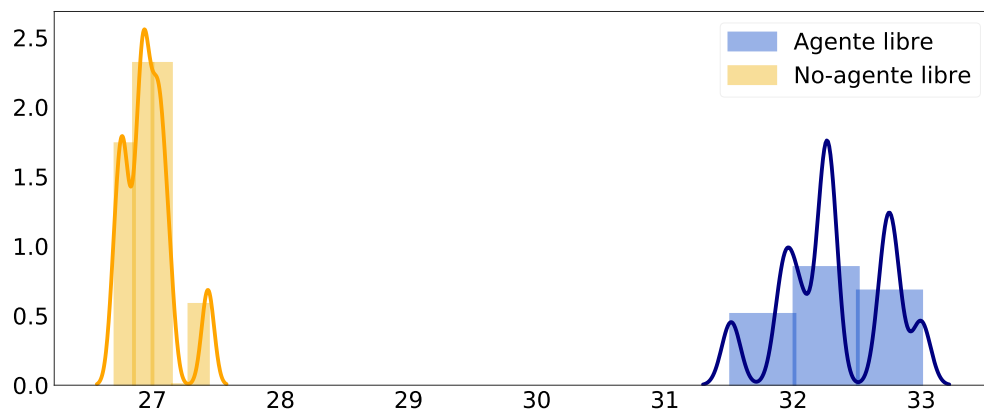


Antigüedad

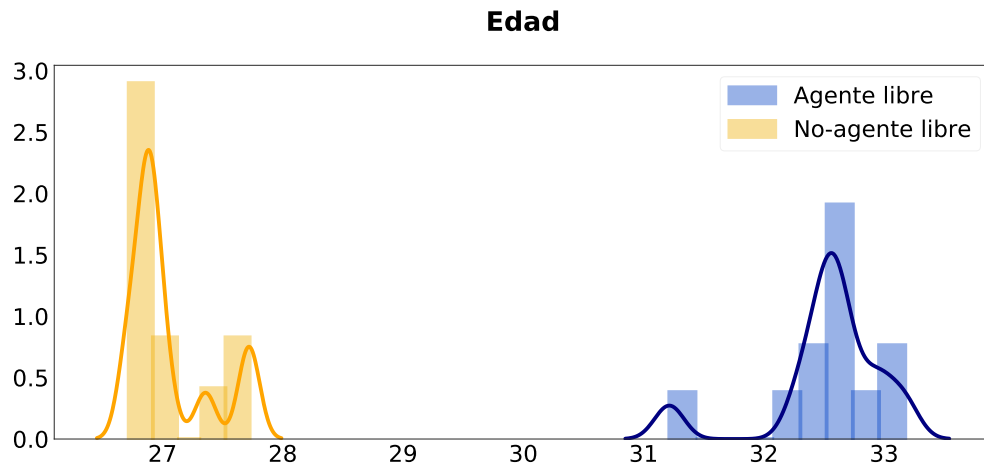


E.2.3. Datos generales - Bateadores

Edad



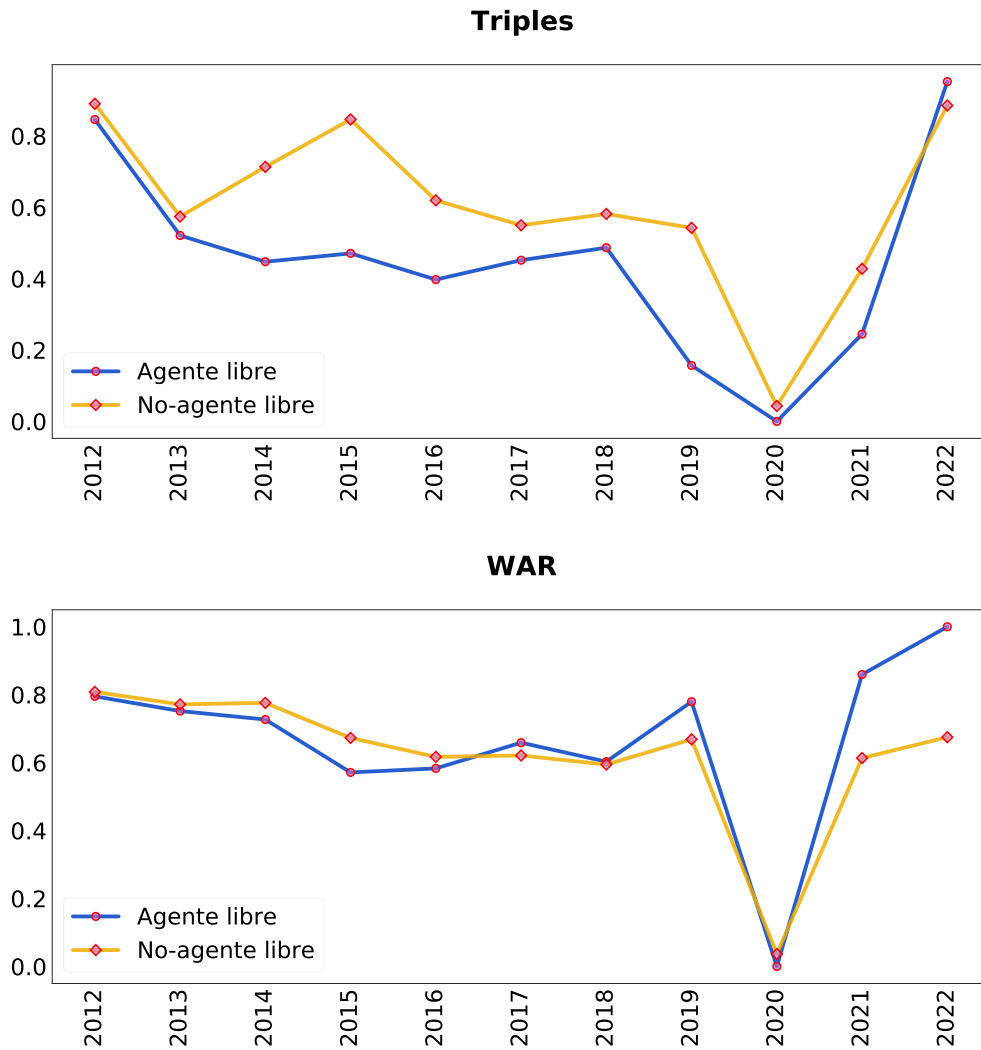
E.2.4. Datos generales - Filderos



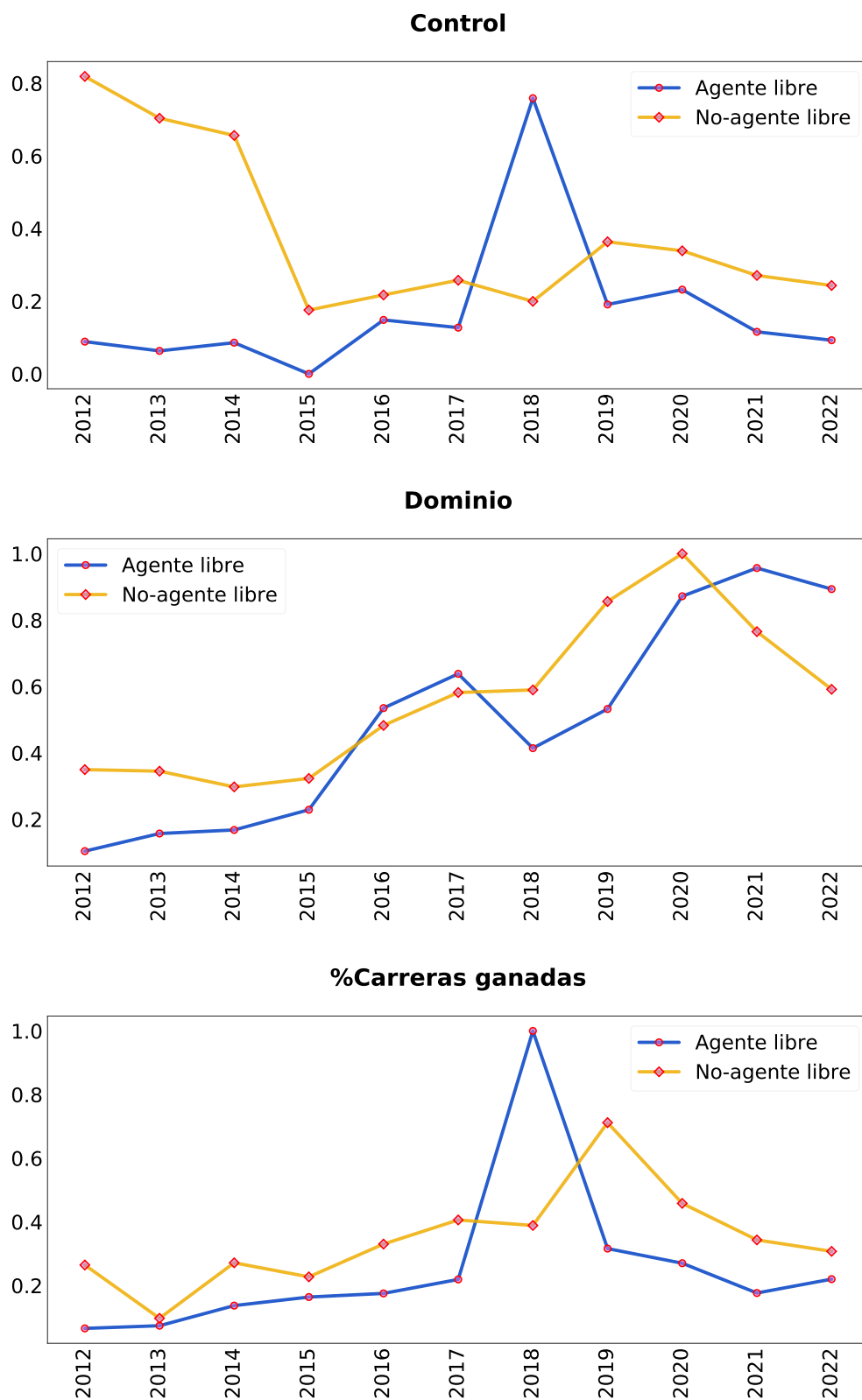
E.3. Series de tiempo

En esta sección se muestran los promedios anuales de las mismas variables usadas en las dos secciones anteriores. Previo a obtener los promedios, se normalizaron los datos para restringir los valores al intervalo $[0, 1]$, que abarca la base de datos.

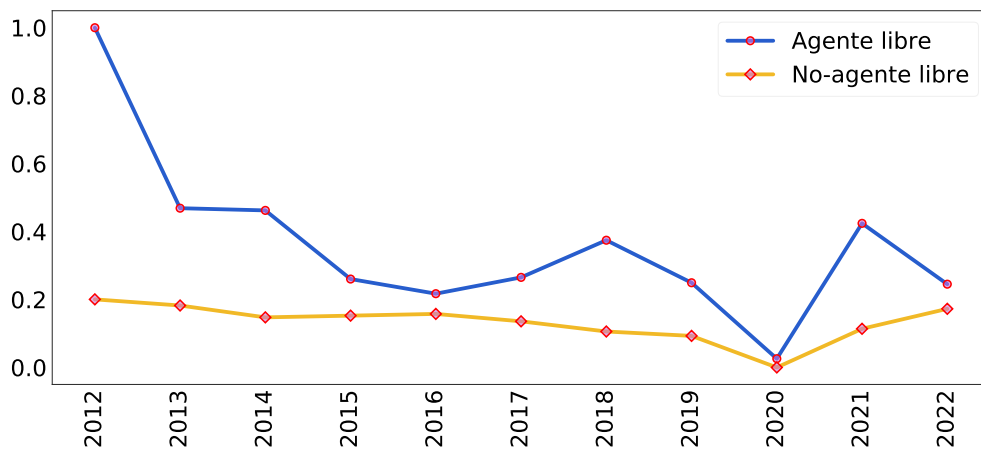
E.3.1. Estadísticas deportivas acumuladas - Bateadores



E.3.2. Estadísticas deportivas acumuladas - Filderos

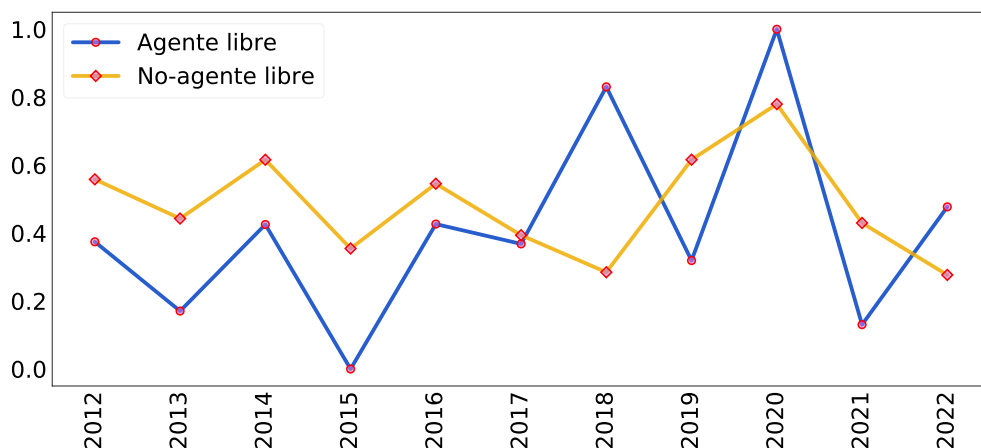


Salvamentos

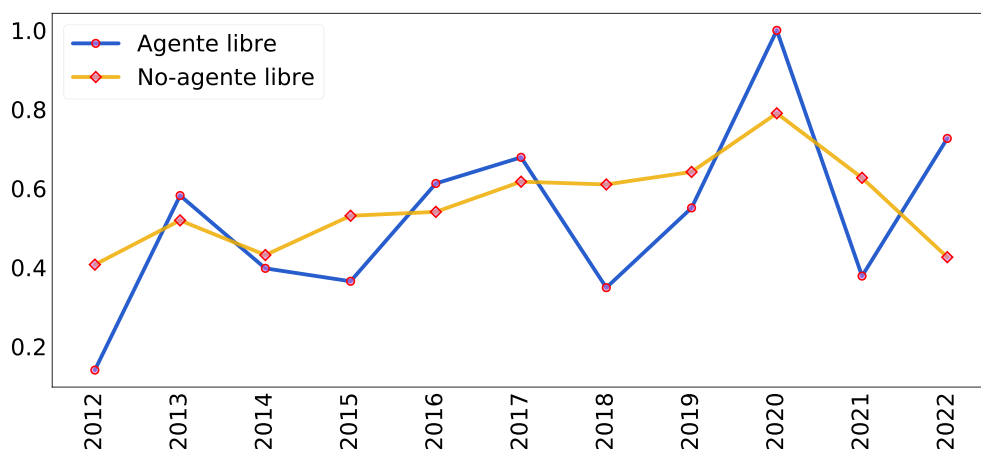


E.3.3. Estadísticas deportivas por juego - Filderos

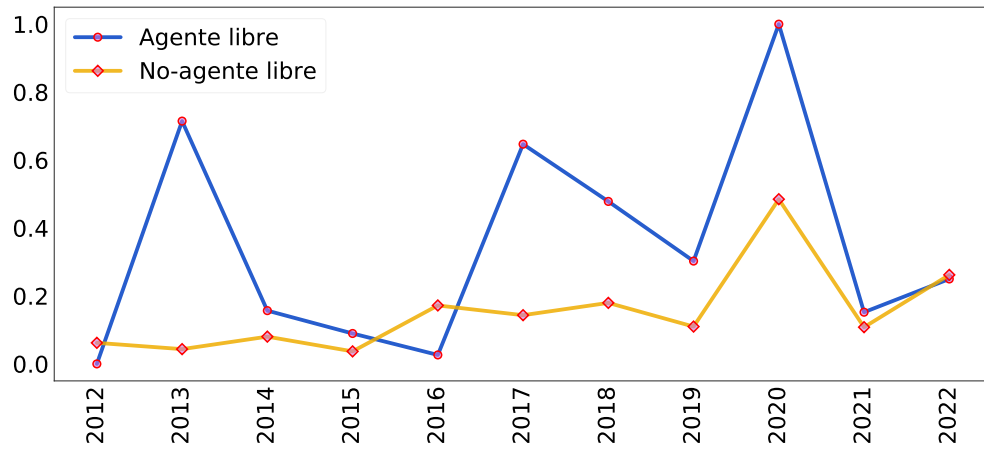
Control



Dominio

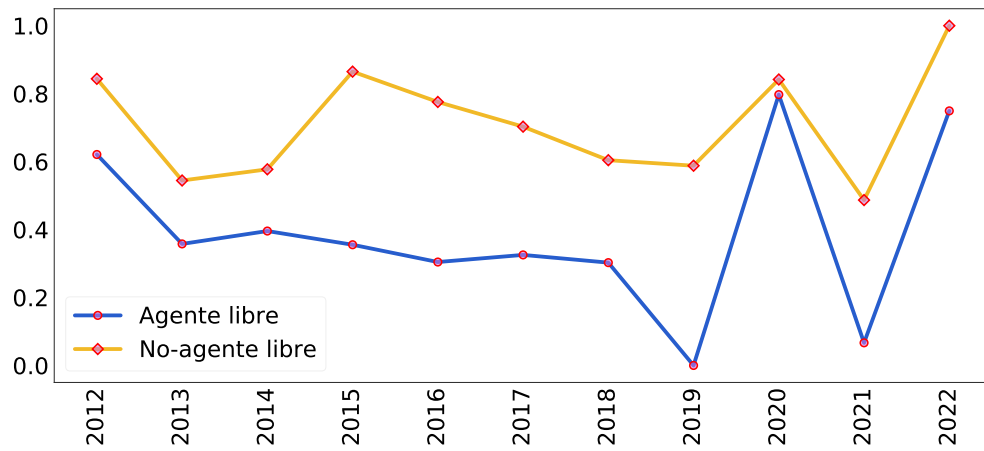


Salvamentos



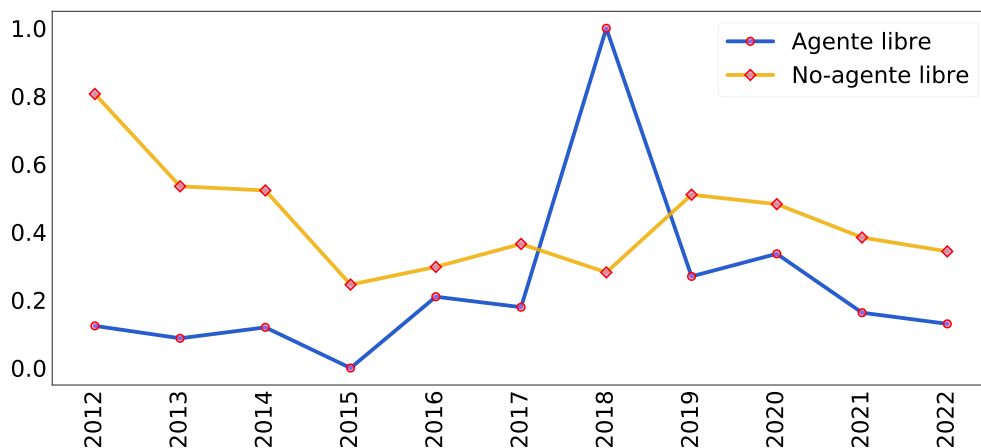
E.3.4. Estadísticas deportivas por cada 10 juegos - Bateadores

Triples

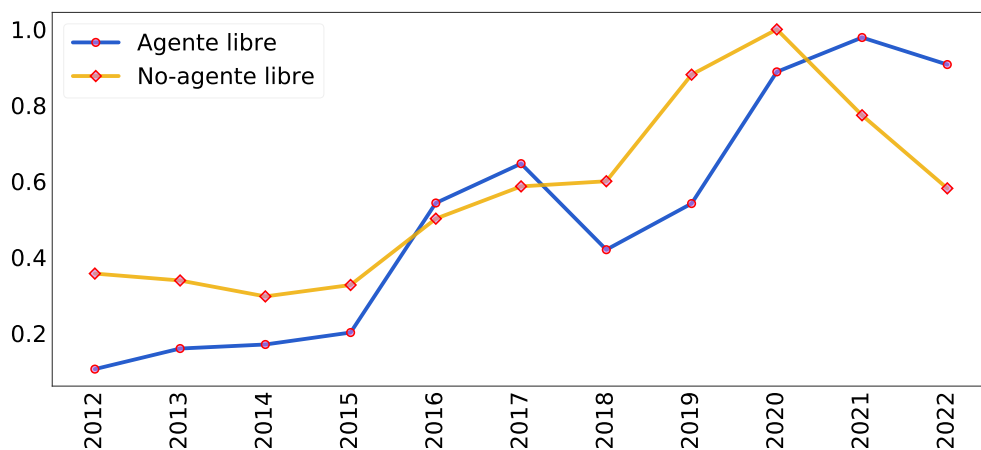


E.3.5. Estadísticas deportivas por cada 10 juegos - Filderos

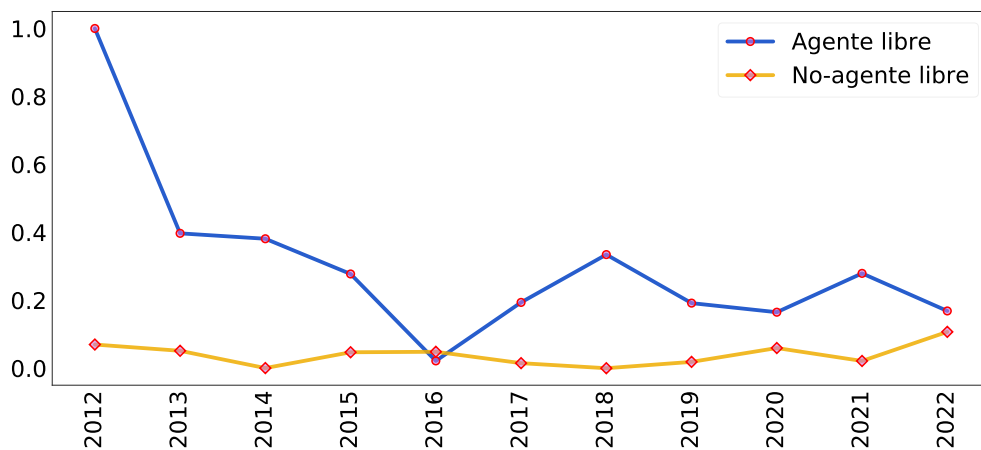
Control



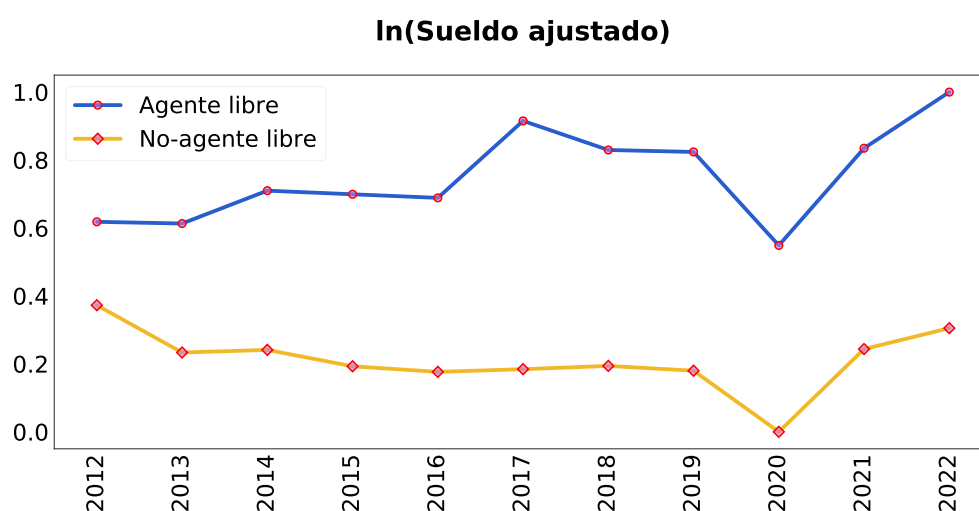
Dominio



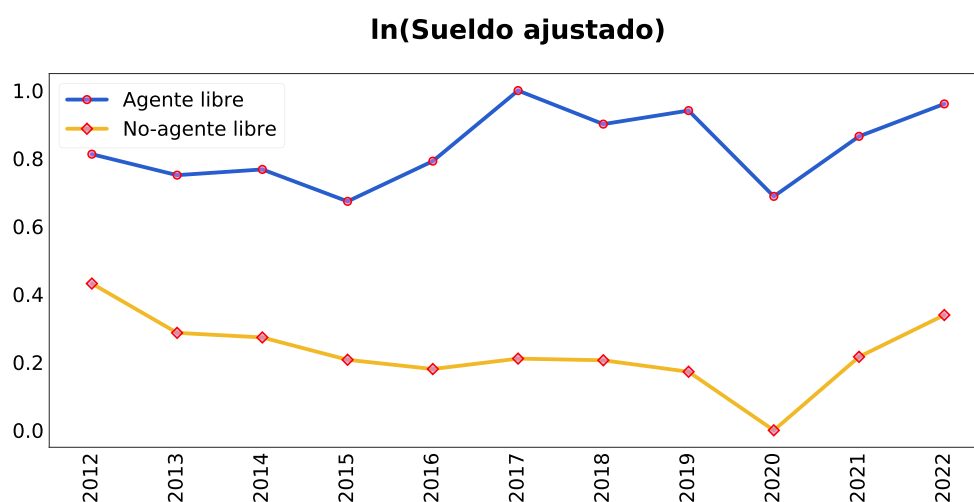
Salvamentos



E.3.6. Salarios - Bateadores



E.3.7. Salarios - Fielderos



F. Problema de identificación

En este apartado se muestran las estimaciones directas e individuales con cada una de las estadísticas deportivas transformadas, X , para el estimador de primeras-diferencias como modelo principal, así mismo, se comparan los resultados con otros modelos o estimadores distintos: *Pooling*, *within* y efectos aleatorios. Los resultados del análisis de robustez se encuentran después de las estimaciones en cada subsección. Por último, los resultados mostrados se concentran en los obtenidos por primeras-diferencias por lo que si se desean ver el resto de resultados para los demás estimadores, se recomienda ir al siguiente enlace: [Análisis de identificación](#).

Cabe aclarar que por motivos de presentación, se muestran los resultados hasta la siguiente página.

F.1. Estimaciones individuales

F.1.1. Bateadores

Cuadro 30: Bateadores: Estimaciones para cada una de las estadísticas deportivas transformadas, X_{var} , independientemente del periodo. Se asumió un estimador en primeras diferencias.

<i>Cambio en el poder de negociación: Y_{it}</i>						
Edad _t	−0.011*** (0.002)	−0.011*** (0.002)	−0.011*** (0.002)	−0.012*** (0.002)	−0.012*** (0.002)	−0.011*** (0.002)
Años contrato _t	−0.045*** (0.009)	−0.045*** (0.009)	−0.045*** (0.009)	−0.043*** (0.009)	−0.044*** (0.009)	−0.044*** (0.009)
Equipo _t	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)
X_{AB_t}	−0.0001 (0.0004)					
$X_{AB_{t-1}}$	0.001*** (0.0003)					
$X_{AB_t^2}$		−0.00002 (0.00001)				
$X_{AB_{t-1}^2}$		0.00001 (0.00003)				
X_{H_t}			−0.001* (0.001)			
$X_{H_{t-1}}$			0.001 (0.001)			
$X_{H_t^2}$				−0.0001*** (0.0001)		
$X_{H_{t-1}^2}$				−0.0002* (0.0001)		
X_{BA_t}					0.0001 (0.012)	
$X_{BA_{t-1}}$					0.039*** (0.010)	
$X_{BA_t^2}$						−0.004 (0.021)
$X_{BA_{t-1}^2}$						0.030*** (0.009)

Note:

*p<0.1; **p<0.05; ***p<0.01

<i>Cambio en el poder de negociación: Y_{it}</i>						
Edad _t	−0.011*** (0.002)	−0.011*** (0.002)	−0.011*** (0.002)	−0.012*** (0.002)	−0.011*** (0.002)	−0.011*** (0.002)
Años contrato _t	−0.045*** (0.009)	−0.045*** (0.009)	−0.047*** (0.010)	−0.049*** (0.010)	−0.046*** (0.009)	−0.045*** (0.009)
Equipo _t	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)
X_{D_t}	−0.002 (0.002)					
$X_{D_{t-1}}$	−0.001 (0.002)					
$X_{D_t^2}$		0.0001 (0.0004)				
$X_{D_{t-1}^2}$		−0.001 (0.0003)				
X_{HR_t}			0.006* (0.004)			
$X_{HR_{t-1}}$			0.001 (0.002)			
$X_{HR_t^2}$				0.001*** (0.0004)		
$X_{HR_{t-1}^2}$				0.0002 (0.0003)		
X_{GS_t}					−0.001 (0.001)	
$X_{GS_{t-1}}$					0.002*** (0.001)	
$X_{GS_t^2}$						−0.00003 (0.0001)
$X_{GS_{t-1}^2}$						0.00004 (0.0001)

Note:

*p<0.1; **p<0.05; ***p<0.01

<i>Cambio en el poder de negociación: Y_{it}</i>						
Edad _t	-0.012*** (0.002)	-0.011*** (0.002)	-0.011*** (0.002)	-0.012*** (0.002)	-0.011*** (0.002)	-0.012*** (0.002)
Años contrato _t	-0.044*** (0.009)	-0.043*** (0.009)	-0.045*** (0.009)	-0.046*** (0.009)	-0.044*** (0.009)	-0.044*** (0.009)
Equipo _t	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)
X_{OPS_t}	-0.007 (0.009)					
$X_{OPS_{t-1}}$	0.013* (0.007)					
$X_{OPS_t^2}$		-0.013 (0.008)				
$X_{OPS_{t-1}^2}$		-0.005 (0.006)				
X_{OBP_t}			0.017 (0.022)			
$X_{OBP_{t-1}}$			0.049*** (0.015)			
$X_{OBP_t^2}$				0.052** (0.026)		
$X_{OBP_{t-1}^2}$				0.029*** (0.010)		
X_{SLG_t}					-0.011 (0.012)	
$X_{SLG_{t-1}}$					-0.010 (0.014)	
$X_{SLG_t^2}$						-0.010 (0.014)
$X_{SLG_{t-1}^2}$						-0.023* (0.014)

Note:

*p<0.1; **p<0.05; ***p<0.01

<i>Cambio en el poder de negociación: Y_{it}</i>						
Edad _t	−0.011*** (0.002)	−0.012*** (0.002)	−0.009*** (0.002)	−0.011*** (0.002)	−0.014*** (0.002)	−0.012*** (0.002)
Años contrato _t	−0.046*** (0.009)	−0.045*** (0.009)	−0.045*** (0.009)	−0.044*** (0.009)	−0.051*** (0.009)	−0.050*** (0.009)
Equipo _t	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)
X_{RBI_t}	0.0004 (0.001)					
$X_{RBI_{t-1}}$	0.002 (0.001)					
$X_{RBI_t^2}$		0.0003 (0.0003)				
$X_{RBI_{t-1}^2}$		−0.0002 (0.0001)				
X_{T_t}			−0.029*** (0.007)			
$X_{T_{t-1}}$			0.002 (0.009)			
$X_{T_t^2}$				−0.002 (0.003)		
$X_{T_{t-1}^2}$				0.003** (0.001)		
X_{WAR_t}					0.030*** (0.003)	
$X_{WAR_{t-1}}$					0.004 (0.005)	
$X_{WAR_t^2}$						0.014*** (0.004)
$X_{WAR_{t-1}^2}$						0.0002 (0.001)

Note:

*p<0.1; **p<0.05; ***p<0.01

F.1.2. Lanzadores iniciales

Cuadro 31: Lanzadores iniciales: Estimaciones para cada una de las estadísticas deportivas transformadas, X_{var} , independientemente del periodo. Se asumió un estimador en primeras diferencias.

<i>Cambio en el poder de negociación: Y_{it}</i>						
Edad _t	−0.019** (0.009)	−0.018*** (0.007)	−0.019** (0.008)	−0.017** (0.008)	−0.015** (0.008)	−0.016** (0.007)
Años contrato _t	−0.025*** (0.009)	−0.043*** (0.007)	−0.035*** (0.007)	−0.033*** (0.007)	−0.033*** (0.008)	−0.036*** (0.008)
Equipo _t	0.002** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
$X_{H_t^2}$	−0.0003*** (0.0001)					
$X_{H_{t-1}^2}$	0.00002 (0.0001)					
X_{H_t}		0.003* (0.001)				
$X_{H_{t-1}}$		0.0005 (0.001)				
$X_{R_t^2}$			−0.0002 (0.0001)			
$X_{R_{t-1}^2}$			0.00003 (0.0001)			
$X_{ER_t^2}$				−0.0005*** (0.0002)		
$X_{ER_{t-1}^2}$				−0.00004 (0.0001)		
X_{ER_t}					−0.001 (0.001)	
$X_{ER_{t-1}}$					0.003*** (0.001)	
X_{R_t}						−0.0002 (0.001)
$X_{R_{t-1}}$						0.003** (0.001)

Note:

*p<0.1; **p<0.05; ***p<0.01

<i>Cambio en el poder de negociación: Y_{it}</i>						
$Edad_t$	-0.020** (0.008)	-0.019** (0.007)	-0.018** (0.007)	-0.016** (0.007)	-0.019*** (0.007)	-0.019*** (0.007)
$Años\ contrato_t$	-0.038*** (0.009)	-0.041*** (0.008)	-0.033*** (0.008)	-0.036*** (0.008)	-0.035*** (0.008)	-0.040*** (0.008)
$Equipo_t$	0.004*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.003*** (0.001)
$X_{Comando_t^2}$	-0.002 (0.003)					
$X_{Comando_{t-1}^2}$	0.00001*** (0.00000)					
$X_{Comando_t}$		0.017* (0.009)				
$X_{Comando_{t-1}}$		0.001*** (0.0003)				
$X_{Control_t^2}$			-0.069*** (0.018)			
$X_{Control_{t-1}^2}$			-0.026*** (0.005)			
$X_{Control_t}$				0.009 (0.034)		
$X_{Control_{t-1}}$				-0.058*** (0.016)		
$X_{Dominio_t^2}$					-0.010*** (0.003)	
$X_{Dominio_{t-1}^2}$					0.009*** (0.003)	
$X_{Dominio_t}$						0.030*** (0.006)
$X_{Dominio_{t-1}}$						0.012** (0.005)

Note:

*p<0.1; **p<0.05; ***p<0.01

<i>Cambio en el poder de negociación: Y_{it}</i>						
Edad _t	-0.016** (0.008)	-0.014* (0.007)	-0.017** (0.008)	-0.015* (0.008)	-0.020*** (0.007)	-0.018** (0.008)
Años contrato _t	-0.033*** (0.010)	-0.035*** (0.011)	-0.029*** (0.008)	-0.029*** (0.009)	-0.034*** (0.007)	-0.033*** (0.007)
Equipo _t	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
$X_{ERA_t^2}$	0.001 (0.002)					
$X_{ERA_{t-1}^2}$	-0.003 (0.003)					
X_{ERA_t}		-0.003 (0.009)				
$X_{ERA_{t-1}}$		-0.021*** (0.005)				
$X_{IP_t^2}$			-0.0002*** (0.0001)			
$X_{IP_{t-1}^2}$			0.00004 (0.0001)			
X_{IP_t}				-0.002** (0.001)		
$X_{IP_{t-1}}$				0.002* (0.001)		
$X_{L_t^2}$					-0.003* (0.002)	
$X_{L_{t-1}^2}$					-0.00002 (0.001)	
X_{L_t}						-0.007 (0.005)
$X_{L_{t-1}}$						-0.0005 (0.003)

Note:

*p<0.1; **p<0.05; ***p<0.01

<i>Cambio en el poder de negociación: Y_{it}</i>						
$Edad_t$	-0.019*** (0.007)	-0.018** (0.007)	-0.018** (0.007)	-0.017** (0.008)	-0.018** (0.007)	-0.017** (0.008)
Años contrato $_t$	-0.036*** (0.008)	-0.036*** (0.008)	-0.035*** (0.009)	-0.042*** (0.008)	-0.030*** (0.009)	-0.035*** (0.010)
Equipo $_t$	0.003*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
$X_{S_t^2}$	0.100*** (0.001)					
$X_{S_{t-1}^2}$	0.020*** (0.006)					
X_{S_t}		0.074*** (0.007)				
$X_{S_{t-1}}$		-0.014 (0.022)				
$X_{SO_t^2}$			-0.0001*** (0.00003)			
$X_{SO_{t-1}^2}$			0.0003*** (0.0001)			
X_{SO_t}				0.001* (0.0005)		
$X_{SO_{t-1}}$				0.002*** (0.001)		
$X_{WAR_t^2}$					-0.002 (0.002)	
$X_{WAR_{t-1}^2}$					-0.004*** (0.001)	
X_{WAR_t}						-0.005 (0.006)
$X_{WAR_{t-1}}$						0.005 (0.008)

Note:

*p<0.1; **p<0.05; ***p<0.01

<i>Cambio en el poder de negociación: Y_{it}</i>						
Edad _t	−0.013*	−0.014*	−0.017**	−0.015**	−0.018**	−0.015*
	(0.007)	(0.008)	(0.007)	(0.008)	(0.009)	(0.008)
Años contrato _t	−0.032***	−0.036***	−0.034***	−0.025***	−0.034***	−0.026***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.010)	(0.009)
Equipo _t	0.003***	0.004***	0.003***	0.004***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$X_{WHIP_t^2}$	0.003					
	(0.004)					
$X_{WHIP_{t-1}^2}$	−0.021***					
	(0.006)					
X_{WHIP_t}		−0.004				
		(0.007)				
$X_{WHIP_{t-1}}$		−0.034**				
		(0.013)				
$X_{BB_t^2}$			−0.0002			
			(0.0002)			
$X_{BB_{t-1}^2}$			0.0005**			
			(0.0002)			
X_{BB_t}				−0.005***		
				(0.001)		
$X_{BB_{t-1}}$				0.004***		
				(0.001)		
$X_{W_t^2}$					−0.001	
					(0.001)	
$X_{W_{t-1}^2}$					0.0003	
					(0.001)	
X_{W_t}						−0.010***
						(0.004)
$X_{W_{t-1}}$						0.003
						(0.003)

Note:

*p<0.1; **p<0.05; ***p<0.01

G. Estimaciones conjuntas

En el siguiente enlace [Análisis estructural](#) se muestran las tablas de bateadores y lanzadores iniciales con los resultados de las regresiones refinadas con los siguientes estimadores: *Pooling*, *within*, efectos aleatorios (RE) y primeras-diferencias (FD). Las variables utilizadas son aquellas obtenidas del problema de indentificación, es decir, aquellas variables X_{var} que fue estadísticamente significativa. Las estimaciones de forma conjunta se refinaron hasta que solo quedaran variables estadísticamente significativas de manera conjunta; para ver todo el proceso de refinamiento se puede consultar el siguiente enlace

Por último, en esta subsección se muestran las pruebas de Hausman para todas las parejas posibles de los modelos refinados de las estimaciones conjuntas anteriores.

Cuadro 32: Bateadores: Hausman Test Results

Modelos	χ^2	df	p-value
Pooling vs Within	24.791	4	5.542e-05
Pooling vs RE	34.85	4	4.988e-07
Pooling vs FD	29.901	4	5.128e-06
Within vs RE	19.316	4	0.0006812
Within vs FD	19.74	4	0.0005619
RE vs FD	26.893	4	2.089e-05

Cuadro 33: Lanzadores iniciales: Hausman Test Results

Modelos	χ^2	df	p-value
Pooling vs Within	4.2929	5	0.5081
Pooling vs RE	4.8623	10	0.9002
Pooling vs FD	9.4283	6	0.1509
Within vs RE	4.4388	5	0.4881
Within vs FD	101.17	7	< 2.2e-16
RE vs FD	9.501	6	0.1473

H. Efectos de la edad a través del tiempo

En esta parte del apéndice se muestran las estimaciones, asumiendo un modelo en primeras diferencias, de los primeros dos años como agente libre comparados con las estimaciones con el resto de años, además se muestran los resultados de la prueba de Hausman entre cada uno de los periodos para cada X_{var} . Debido a la gran cantidad de resultados y la extensión de estos, se tienen en el siguiente enlace [Análisis estructural](#). No obstante, se muestran las tablas con la dirección del crecimiento de los estimadores obtenidos por primeras-diferencias.

I. Estimaciones del periodo de transición a agente libre

Las estimaciones con las observaciones correspondientes al último periodo del jugador como no-agente libre y el primer periodo como agente libre se muestran aquí abajo. En este caso, se muestran los resultados obtenidos por un estimador *pooling*. Las estimaciones de cada medida de desempeño se muestran en el siguiente enlace .

I.1. Bateadores

Variable	χ^2	df	p-value
X_{AB}	59.168	5	1.805e-11
X_H^2	18.707	5	0.002179
X_H	11.385	5	0.04426
X_{BA}	11.884	5	0.03642
X_{BA}^2	8.2914	5	0.1409
X_{HR}	21.733	5	0.0005885
X_{HR}^2	12.028	5	0.0344
X_{IP}	73.223	5	2.184e-14
X_{OPS}	721.14	5	<2.2e-16
X_{OBP}	5.0247	5	0.4129
X_{OBP}^2	9.2569	5	0.09924
X_{RBI}	65.011	5	1.115e-12
X_{WAR}	9.4941	5	0.09091
X_{WAR}^2	10.736	5	0.05687

Cuadro 34: Caption

I.2. Lanzadores iniciales

Variable	χ^2	df	p-value
X_H^2	112.41	5	<2.2e-16
X_H	68.033	5	2.629e-13

Variable	χ^2	df	p-value
X_{ER}^2	40.822	5	1.019e-07
X_{ER}	4.0421	5	0.5434
X_{ERA}	106.44	5	<2.2e-16
X_R	28.166	5	3.378e-05
$X_{Comando}^2$	1.3237	5	0.9325
$X_{Comando}$	17.236	5	0.004074
$X_{Control}^2$	291.17	5	<2.2e-16
$X_{Control}$	210.26	5	<2.2e-16
$X_{Dominio}^2$	10.813	5	0.05521
$X_{Dominio}$	18.944	5	0.001969
X_{IP}^2	98.225	5	<2.2e-16
X_{IP}	91.178	5	<2.2e-16
X_L^2	119.05	5	<2.2e-16
X_{SO}^2	1490.6	5	<2.2e-16
X_{SO}	65.845	5	7.484e-13
X_{WAR}^2	46.886	5	5.993e-09
X_{WHIP}^2	29.987	5	1.483e-05
X_{WHIP}	231.55	5	<2.2e-16
X_{BB}^2	1425.3	5	<2.2e-16
X_{BB}	35.3	5	1.311e-06
X_W	586.52	5	<2.2e-16

Cuadro 35: Test para cambio estructural entre periodos

J. Análisis por PCA

En esta subsección se muestran las estimaciones hechas con PCA para bateadores regulares y lanzadores iniciales. Análogamente a otros análisis de robustez, se muestran los resultados usando estimadores *pooling*, *within*, efectos aleatorios (RE) y primeras diferencias (FD). El análisis sobre cómo se determinó el número óptimo de componentes se muestran el siguiente enlace [Componentes máximo - PCA](#).

J.1. Bateadores

Cuadro 36: Bateadores regulares: Modelos con PCA

	<i>Cambio en el poder de negociación: Y_{it}</i>			
	Pooling	Within	RE	FD
Edad _t	−0.006** (0.003)	−0.004 (0.006)	−0.005** (0.003)	−0.011*** (0.002)
Años contrato _t	−0.001 (0.004)	−0.032** (0.012)	−0.003 (0.004)	−0.045*** (0.009)
Equipo _t	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.002*** (0.001)
PCA _{1t}	0.00002 (0.00003)	−0.00000 (0.00004)	0.00001 (0.00003)	0.00002 (0.00001)
PCA _{1t−1}	−0.00000 (0.00002)	−0.00000 (0.00004)	−0.00000 (0.00002)	−0.00000 (0.00002)
Agente _t	0.157* (0.081)		0.148* (0.083)	

Note:

*p<0.1; **p<0.05; ***p<0.01

J.2. Lanzadores iniciales

Cuadro 37: Lanzadores Iniciales: Modelos con PCA

<i>Cambio en el poder de negociación: Y_{it}</i>				
	Pooling	Within	RE	FD
Edad _t	−0.008** (0.004)	−0.030** (0.015)	−0.010** (0.005)	−0.017* (0.009)
Años contrato _t	−0.006 (0.007)	−0.025 (0.019)	−0.006 (0.007)	−0.029*** (0.009)
Equipo _t	0.003* (0.002)	0.004 (0.002)	0.003* (0.001)	0.003*** (0.001)
PCA _{1t}	−0.002 (0.006)	−0.013 (0.008)	−0.003 (0.006)	−0.001 (0.003)
PCA _{2t}	−0.0001 (0.0001)	−0.00001 (0.0001)	−0.0001 (0.0001)	−0.0001*** (0.00003)
PCA _{1t−1}	0.00001 (0.00001)	−0.00001** (0.00000)	0.00000 (0.00000)	−0.00001** (0.00000)
PCA _{2t−1}	−0.00000 (0.00005)	0.00001 (0.0001)	−0.00001 (0.00004)	−0.0001 (0.00004)
Agente _t	0.242* (0.142)		0.310* (0.173)	

Note:

*p<0.1; **p<0.05; ***p<0.01

K. COVID-19

Las pruebas de Hausman para cambios estructurales de los modelos estimados de manera conjunta se muestran en las siguientes tablas para cada posición. Ambas estimaciones conjuntas tienen las mismas variables, solo que en una estimación se omitieron las observaciones del 2020.

K.1. Bateadores

Modelo	χ^2	df	p-value
Pooling	3.9513	5	0.5565
Within	3.0371	4	0.5516
Efectos aleatorios	1.392	4	0.8456
Primeras-diferencias	15.506	19	0.6899

Cuadro 38: Hausman Test Results for COVID-19 and MLB Hitters

K.2. Lanzadores iniciales

Modelo	χ^2	df	p-value
Pooling	6.6745	10	0.7558
Within	2.5947	8	0.9572
Efectos aleatorios	6.2746	10	0.7917
Primeras-diferencias	12.337	30	0.9982

Cuadro 39: Hausman Test Results for COVID-19 and MLB Starting Pitchers

L. Diferencias-en-Diferencias

En este último apartado mostramos solo las estimaciones para el modelo de diferencias-en-diferencias para el caso de la variable Y_i y los salarios normalizados ω_i . Las gráficas de las tendencias para el grupo con y sin tratamiento se encuentran en el siguiente enlace: [Análisis DID](#).

L.1. Bateadores

<i>Cambio en el poder de negociación: Y_{it}</i>	
Tratamiento	0.058* (0.032)
Periodo agente	-0.064*** (0.014)
Años contrato _t	0.003 (0.004)
Edad _t	-0.010*** (0.002)
Equipo _t	0.0005 (0.001)
ATE	-0.018 (0.034)
<i>Nota:</i> *p<0.1; **p<0.05; ***p<0.01	

<i>Salario regular: ω_{it}</i>	
Tratamiento	0.058* (0.032)
Periodo agente	-0.064*** (0.014)
Años contrato _t	0.003 (0.004)
Edad _t	-0.010*** (0.002)
Equipo _t	0.0005 (0.001)
ATE	-0.018 (0.034)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

L.2. Lanzadores iniciales

<i>Cambio en el poder de negociación: Y_{it}</i>	
Tratamiento	0.048 (0.035)
Periodo _t agente _t	0.021 (0.019)
Años contrato _t	0.065*** (0.007)
Edad _t	0.023*** (0.004)
Equipo _t	-0.0004 (0.001)
ATE	-0.063* (0.038)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

	<i>Salario regular: ω_{it}</i>
Tratamiento	0.048 (0.035)
Periodo agente	0.021 (0.019)
Años contrato _t	0.065*** (0.007)
Edad _t	0.023*** (0.004)
Equipo _t	−0.0004 (0.001)
ATE	−0.063* (0.038)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Referencias

- Baseball Rules* (2023). <https://www.rulesofsport.com/sports/baseball.html>.
- Blum, R. (2019). How mlb's luxury tax works, <https://www.nytimes.com/2019/03/20/sports/baseball/mlb-luxury-tax-explainer.html>.
- Brown, J. (2009). Understanding the mlb's free agent compensation, <https://bleacherreport.com/articles/296012-understanding-free-agent-compensation>.
- Brown, M. (2022). How major league baseball could crack \$11 billion in revenues in 2022, <https://www.forbes.com/sites/maurybrown/2022/04/07/how-major-league-baseball-could-crack-11-billion-in-revenues-in-2022/?sh=3ac967ed7f63>.
- Congdon-Hohman, J. M. and Lanning, J. A. (2018). Beyond moneyball: Changing compensation in mlb, *Journal of Sports Economics* **19**(7): 1046–1061.
- Connor Heaton, P. M. (2022). Using machine learning to describe how players impact the game in the mlb, *Sports Analytics Conference*.
- ContractsCounsel (2023). Sports agency contract, <https://www.contractsounsel.com/t/us/sports-agency-contract>.
- Dickson, P. (1999). *The New Dickson Baseball Dictionary: A Cyclopedic Reference to More Than 7,000 Words, Names, Phrases, and Slang Expressions that Define the Game, Its Heritage, Culture, and Variations*, Houghton Mifflin Harcourt.
- DiGiannatale et al, Curiel-Cabral I., B. G. (2023). The dynamics of bargaining power in a principal-agent model, *División de Economía. Centro de Investigación y Docencia Económicas (CIDE)* p. 630.

- Enders, C. K. (2022). *Applied missing data analysis*, Guilford Publications.
- Ferwana, I. and Varshney, L. R. (2021). The impact of covid-19 lockdowns on mental health patient populations: Evidence from medical claims data, *medRxiv*.
- Free Agency Definition* (2023). <https://www.mlb.com/glossary/transactions/free-agency>.
- Gardner, R. (1995). Games for business and economics.
- Heyman, J. (2023). The comprehensive guide to mlb contract terminology, <https://www.fancredsports.com/articles/jon-heyman/the-comprehensive-guide-to-mlb-contract-terminology>.
- Hoffman, M. G. (2014). *Analysis of salary for Major League Baseball players*, PhD thesis, North Dakota State University.
- Holmstrom, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives, *Econometrica: Journal of the Econometric Society* pp. 303–328.
- Janvrin, R. (2022). Nfl vs mlb: Revenue, salaries, viewership, attendance and ratings, <https://www.wsn.com/nfl/nfl-vs-mlb/>.
- Mas-Colell, A., Whinston, M. D., Green, J. R. et al. (1995). *Microeconomic theory*, Vol. 1, Oxford university press New York.
- MLB (2023). Baseball basics: On the field, <https://www.mlb.com/official-information/basics/field>.
- Navin, J. C. and Sullivan, T. S. (2002). The relationship between bargaining power and salary: Evidence from the national hockey league, *Technical report*, Working Paper.
- Official Playing Rules Committee, Amateur Baseball Advisory Member, C. (2021). Official baseball rules edition.
- Ohrnberger, J., Fichera, E. and Sutton, M. (2017). The relationship between physical and mental health: A mediation analysis, *Social science & medicine* **195**: 42–49.
- Paine, N. (2023). Is a mlb star more valuable than an nba or nfl star?, <https://fivethirtyeight.com/features/is-a-mlb-star-more-valuable-than-an-nba-or-nfl-star/>.
- Status, B. (2021). Top 10 most watched sports in the world today, <https://www.ballerstatus.com/2021/01/06/top-10-most-watched-sports-in-the-world-today/>.
- Thorpe, D. (2010). An introduction to sports law and contracts, *Australian and New Zealand Sports Law Journal* **5**(2).
URL: <http://classic.austlii.edu.au/au/journals/ANZSportsLawJl/2010/2.pdf>

- Todd, J. (2017). Why do mlb free agents sign longer contracts than those in other sports?, <https://www.mlbtraderumors.com/2017/01/why-do-mlb-free-agents-sign-longer-contracts-than-those-in-other-sports.html>.
- Treatment, R. (2022). Mlb revenue split versus nfl and nba, <https://www.royalsreview.com/2022/11/4/23438378/mlb-revenue-split-versus-nfl-and-nba>.
- Trendafilov, R., Markovski, S. and Ivanov, S. (2017). The relationship between bargaining power and salary: Evidence from the national hockey league, *Economics and Management* **13**: 21–28.
- Yoon, S.-J. (2019). What can we obtain from mental health care? the dynamics of physical and mental health, *International Journal of Environmental Research and Public Health* **16**(17): 3098.