

**ALMA MATER STUDIORUM – UNIVERSITA' DI
BOLOGNA**

SCUOLA DI ECONOMIA, MANAGEMENT E STATISTICA
Corso di Laurea Magistrale in Statistica, Economia e Impresa

**L'IMPATTO DELLA PERFORMANCE SUI CONTRATTI PROFESSIONALI:
APPLICAZIONE DEL MODELLO DI SELEZIONE DI HECKMAN ALLA MAJOR LEAGUE
BASEBALL**

Presentata da:

Luca Accorsi:

matricola: 0000771026

Relatore:

Prof. Mario Mazzocchi

APPELLO Dicembre 2017
ANNO ACCADEMICO 2016 / 2017

INDICE

INDICE	1
INTRODUZIONE	4
1. CONTESTUALIZZAZIONE ECONOMICA.....	6
1.1 Iter di formazione del giocatore di baseball	7
1.1.3 Probationary Period	8
1.1.4 Approdo in Major League	9
1.2 Giocatore professionista di baseball come lavoratore	10
1.2.1 Origine degli Sport Professionistici.....	10
1.2.2 Giocatore come lavoratore subordinato.....	11
1.2.3 Giocatore come lavoratore specializzato	12
1.2.4 Giocatore come lavoratore con vita lavorativa breve	12
1.2.5 Disparità del potere contrattuale.....	13
1.2.5.1 Modalità di redistribuzione del potere contrattuale	15
2. COSTRUZIONE DEL DATASET	18
2.1 Lahman Baseball Database.....	19
2.1.1 Sean Lahman e il “Giornalismo di Precisione”	19
2.1.1.1 Giornalismo di precisione.....	19
2.1.1.2 Campi di pertinenza.....	19
2.1.1.3 Sean Lahman	20
2.1.2 Descrizione del database	21
2.2 Dal Lahman Database al nostro dataset.....	22
2.2.3 Le variabili.....	22
2.2.3.1 Variabili relative alle apparizioni in battuta	22
2.2.3.2 Variabili relative alle apparizioni in difesa.....	22
2.2.3.3 Variabili descrittive del giocatore.....	23
2.2.3.4 Variabili relative alla squadra del giocatore	23
2.2.3.5 Variabili indicanti statistiche della battuta	23
2.2.3.6 Variabili relative alla carriera in Major League dei giocatori	26
3. METODOLOGIA.....	27
3.1 Uno sguardo alla letteratura esistente.....	28
3.2 Formulazione del modello	30
3.2.1 Stima del modello sulle determinanti del salario.....	30
3.2.2 Cosa si intende con “auto-selezione del campione”	31
3.2.3 Il modello di selezione di Heckman	34

3.3 Applicazione del modello al caso in esame	37
4. PREMESSA ALLE STIME	38
4.1 Premessa relativa ai dati destinati all'analisi	39
4.1.1 Perché i lanciatori sono esclusi dal modello	39
4.2 Come abbiamo utilizzato la variabile del salario.....	40
4.2.1 Rappresentazione grafica del problema dei salari nulli).....	40
4.2.2 Presenza di un salario minimo	41
4.2.3 Deflazionamento dei salari nominali	43
4.2.4 Dal salario effettivo, alla variazione dal minimo salariale	44
5. STIME E RISULTATI.....	46
5.1 La relazione temporale tra salari e prestazioni	47
5.2 Modello di Selezione di Heckman – Step I	49
5.2.1 La variabile dipendente	50
5.2.2 Le variabili indipendenti.....	50
5.2.2.1 Variabili che descrivono il giocatore.....	50
5.2.2.2 Variabili che descrivono la società.....	52
5.2.2.3 Variabili relative al rapporto giocatore-società	52
5.2.3 Le stime del Probit.....	53
5.2.3.1 Interpretazione dei coefficienti	54
5.2.3.2 Regressori con coefficienti negativi	54
5.2.3.3 Regressori con coefficienti positivi	55
5.4 Modello di Selezione di Heckman – Step II.....	56
5.4.1 Il calcolo del Mill's Ratio.....	56
5.4.2 La variabile dipendente	57
5.4.3 I regressori in comune col Probit.....	57
5.4.4 I regressori specifici della regressione.....	57
5.4.5 Le stime della regressione	58
5.4.5.1 Interpretazione dei coefficienti relativi all'esperienza e alle performance	59
5.4.5.2 Interpretazione dei coefficienti delle squadre di Major League	61
5.4.5.3 Interpretazione dei coefficienti relativi ai diversi ruoli	62
5.4.5.4 Interpretazione del parametro del Mill's Ratio.....	63
CONCLUSIONI	64
APPENDICE	66
BIBLIOGRAFIA	74
SITOGRAFIA	75

INTRODUZIONE

In questa tesi si vuole analizzare la relazione tra performance e salario per i giocatori professionisti della Major League Baseball. Per le caratteristiche del data-set si è resa necessaria l'applicazione del modello di selezione di Heckman.

In primis verrà redatta una parte di contestualizzazione economica, nella quale inquadriamo il giocatore professionista di baseball all'interno del mondo del lavoro specializzato. All'interno di questa sezione, da un lato descriveremo nel dettaglio le caratteristiche generiche dei lavoratori che presenta il giocatore di baseball, dall'altro faremo particolare attenzione alla regolamentazione alla base della Major League Baseball per descrivere con precisione il contratto di Major League, e quindi le caratteristiche del rapporto lavorativo che a firma del contratto instaura tra il giocatore e la società proprietaria.

Nel secondo capitolo ci concentreremo sulla creazione del data-set che verrà poi utilizzato per le analisi. Introducendo il concetto di sabermetrica, dedicheremo particolare attenzione alla costruzione di indicatori e di statistiche di performance, a partire dai dati relativi alle prestazioni dei giocatori, contenuti all'interno del Lahman Baseball Database.

Nel capitolo dedicato alla metodologia, effettueremo prima di tutto un excursus sulla letteratura esistente riguardo alla ricerca sul baseball, andando a citare diversi studi e pubblicazioni per riuscire a riassumere al meglio i possibili campi di studio, e soffermandoci in particolare sugli argomenti trattati all'interno della contestualizzazione economica. Conseguentemente illustreremo la formulazione teorica del modello, analizzando nel dettaglio il problema di selezione del campione e spiegando in che modo affrontare il problema attraverso la stima del modello di selezione di Heckman. Infine applichiamo la formulazione teorica al nostro caso specifico, andando a spiegare in cosa consiste il problema di selezione del campione all'interno della Major League Baseball.

Nella quarta sezione del trattato si introdurrà il problema della stima, e in tale contesto si discuteranno alcune correzioni necessarie all'utilizzo del data-set per i fini delle nostre stime. Dedicheremo particolare attenzione alla variabile salario, con lo scopo di trovare il modo migliore di rappresentarne il valore, che ci permetterà di interpretare in modo più immediato e corretto le nostre stime.

Nel quinto capitolo della tesi il modello di Heckman viene specificato e stimato per l'applicazione ai dati sui salari dei giocatori della Major League Baseball. In questo capitolo vengono discussi i risultati ottenuti.

Infine nelle conclusioni si analizzerà la rispondenza dei risultati alle aspettative. L'obiettivo è di verificare se l'applicazione del modello di selezione di Heckman sia un metodo plausibile per stimare in modo realistico l'impatto delle performance sul contratto di Major League Baseball.

CAPITOLO 1

CONTESTUALIZZAZIONE ECONOMICA

1.1 Iter di formazione del giocatore di baseball

L'obiettivo di questo primo capitolo è quello di effettuare una contestualizzazione economica del giocatore di baseball e del mercato del lavoro delle franchigie di Major League, andando in primo luogo a descrivere l'iter necessario per raggiungere la massima serie, in secondo luogo a trovare quali siano le caratteristiche del giocatore in quanto lavoratore, ed infine ad enfatizzare i dettagli salienti del contratto e del rapporto lavorativo tra giocatore e società.

1.1.2 Concezione dello sport negli Stati Uniti d'America

Negli Stati Uniti lo sport ricopre un'importanza assoluta, non solo rappresenta uno dei fulcri della cultura americana, ma è anche considerato un elemento imprescindibile come parte integrante dell'educazione scolastica. Infatti, da un lato sono le famiglie a trasmettere la passione per lo sport ai bambini fin da piccoli, dall'altro è compito della scuola iniziare a farli giocare in vere e proprie squadre. Con l'avanzare dell'età dei bambini avviene quindi un passaggio di testimone tra le famiglie e le scuole, dove le ultime si assumono la responsabilità della loro educazione sportiva, supportando e favorendo l'agonismo.

La formazione culturale e quella sportiva avanzano quindi di pari passo durante tutta la carriera scolastica, fino addirittura al College. Per riuscire ad entrare nei College più prestigiosi degli Stati Uniti gli studenti devono soddisfare determinati standard di rendimento scolastico ed inoltre superare severe prove di ammissione che testano la loro preparazione culturale. Una volta riconosciuto il merito dello studente (tramite l'ammissione), viene anche però riconosciuto il merito sportivo, che viene premiato tramite l'elargizione di generose borse di studio, grazie alle quali gli studenti possono permettersi di pagare anni di studi altrimenti molto costosi.

Essendo il baseball uno degli sport nazionali americani di maggior rilevanza, quello che avviene dalla Primary School al College è un elevato processo di selezione tra i giocatori, che permette di formare squadre di campioni e allo stesso tempo di dar vita a competizioni di livello nazionale, che riempiono gli stadi ed hanno anche un forte seguito mediatico.

Inoltre, come conseguenza dell'elevata visibilità del campionato dei College, il pubblico non è formato soltanto da appassionati e tifosi, bensì comprende anche un elevato numero di scout (osservatori) mandati dalle società di baseball professionistico per osservare i giocatori. Per fare un'analogia tra il baseball e il mondo del lavoro, possiamo identificare le migliaia di dati e statistiche risultanti dall'attenta osservazione degli scout, con i curriculum dei singoli giocatori: una volta analizzati e confrontati i vari curriculum, la società è quindi pronta a selezionare i candidati al posto di lavoro.

1.1.3 Probationary Period

Agli inizi di giugno di ogni anno la Major League Baseball organizza il “First Year Player Draft”. Si tratta di un evento durante il quale le 30 società di baseball si spartiscono le “nuove promesse” uscenti dai college americani. Solitamente verrebbe da pensare ad un’asta al rialzo che veda il migliore offerente aggiudicarsi il giocatore selezionato, ma così non è: ogni squadra ha un turno ben preciso per scegliere semplicemente un giocatore ed aggiudicarselo, e l’ordine di scelta segue la classifica della stagione precedente partendo dal basso (così che la squadra che abbia registrato il peggior record di vittorie/sconfitte sarà la prima a scegliere durante il draft dell’anno successivo).

Una volta selezionato, il giocatore intraprende un cosiddetto *Probationary Period* (periodo di prova) durante il quale la società proprietaria lo posiziona all’interno di una squadra delle leghe minori con la quale parteciperà al relativo campionato (Bollinger & Hotchkiss 2003: 924-926; 929). Continuando l’analogia precedentemente proposta, se pensiamo all’appartenere alla squadra di Major League come al posto di lavoro, di conseguenza il *probationary period* presenta elementi caratteristici di due momenti diversi di una generica carriera.

Durante questo periodo il giocatore si ritrova a sviluppare le proprie abilità tecniche e ad aumentare la propria esperienza, affrontando campionati e livelli di competizione differenti da una serie all’altra. Se poniamo il nostro sistema di riferimento all’interno della società intera (squadra di Major League e squadre di Minor Leagues affiliate), possiamo concludere che si tratta di una fase della carriera in cui il giocatore accresce e rafforza le proprie capacità, percependo un livello di stipendio minore di quello di Major League e proporzionale al livello della serie, e per questo motivo possiamo interpretarlo come un vero e proprio *Apprendistato*.

Spostiamo invece ora il sistema di riferimento alla sola squadra di Major League. La società monitora costantemente il comportamento e le capacità del giocatore, e decide di “assumerlo” (in Major League appunto) soltanto nel momento in cui vede soddisfatti determinati criteri e requisiti. In questo caso quindi, la società non sta valutando il giocatore sul posto di lavoro, ma guarda e valuta dall’esterno le sue capacità. Per questo motivo possiamo interpretare il monitoraggio della società, come una lettura molto dettagliata del curriculum del giocatore. Inoltre, non percependo il giocatore alcuno stipendi di Major League, possiamo interpretare l’intero *probationary period* (il quale può culminare con la firma di un contratto) con un colloquio presso la squadra di Major League.

Che lo si voglia interpretare in un modo o nell’altro, la durata complessiva lo renderebbe uno degli apprendistati (o colloqui) più lunghi del mondo del lavoro. Si parla infatti di un massimo di 5 anni, alla scadenza dei quali la società proprietaria si ritrova costretta a promuovere il giocatore nella massima serie, o in alternativa a rilasciarlo (sarà poi una delle altre 29 società a portare il

giocatore in Major League). Dopo 5 lunghi anni il giocatore finalmente si ritrova nel roster della prima squadra¹, ma ancora una volta non è sicuro di firmare un contratto. Andiamo a vedere perché.

1.1.4 Approdo in Major League

Generalmente il termine “roster” è utilizzato in molti sport per indicare la lista ufficiale stagionale dei giocatori di una squadra. Per quanto riguarda la Major League Baseball, la regolamentazione della lega prevede però l’esistenza di due diversi roster: il “Roster Attivo”, o *25-man Roster*, (Glossary MLB 2017a) ed il “Roster Allargato”, anche chiamato *40-man Roster* (Glossary MLB 2017b).

Il primo è la lista completa dei giocatori “attivi” della squadra, ovvero di quei soli giocatori che hanno le carte in regola per giocare in Major League durante la stagione di riferimento. Secondo le regole della MLB il Roster Attivo deve avere un minimo di 24 giocatori e un massimo di 25 durante la Regular Season, fatta eccezione per i *doubleheaders* (casi in cui si giochino due partite nello stesso giorno) nei quali le società hanno la possibilità di convocare un ulteriore giocatore, che mantiene il diritto di essere il 26esimo giocatore attivo solamente in quel giorno.²

Il secondo comprende tutti i giocatori che sono potenzialmente convocabili nel roster attivo in qualsiasi momento della stagione. In esso sono compresi tutti i giocatori del roster attivo più un massimo di altri 15 giocatori, per rientrare nel limite di 40 giocatori. I giocatori “non-attivi” sono una combinazione di giocatori con infortuni dal breve recupero (7-10 giorni), giocatori ai quali la società concede il permesso di allontanarsi dalla squadra per problemi familiari, e giocatori che militano nelle Minor Leagues. Inoltre il 1 Settembre di ogni stagione (fino alla fine della stessa) le società espandono i loro roster attivi dal *25-man* al *40-man*, che per questo motivo viene anche definito roster “allargato”.

Tornando alla conclusione del paragrafo (1.1.3), possiamo ora spiegare nel dettaglio che il giocatore (al più tardi alla fine del quinto anno nelle Minors) viene “promosso” all’interno del *40-man Roster*. Ottenuta questa promozione, rischia tuttavia di aspettare ancora fino a 3 anni, prima di ottenere il contratto di Major League: quando un giocatore viene inserito nel *40-man Roster*, la società detiene su di lui 3 opzioni, ovvero ha il potere di trasferire il giocatore avanti e indietro tra le *Minors* e la Major League per tutta la durata di 3 diverse intere stagioni.³ Quello che cambia rispetto

¹ Ricordiamo che 5 anni rappresenta il caso estremo, ma potrebbe anche accadere prima.

² Solitamente le scelte di società e allenatori ricadono su un Roster Attivo composto da 5 lanciatori partenti, 7 rilievi, 2 ricevitori, 6 interni e 7 esterni; eventuali correzioni rispetto a questo schema sono da attribuire a scelte strategiche e alle opportunità di effettuare cambi di ruolo, in base ai giocatori che si hanno a disposizione.

³ Le stagioni di cui stiamo parlando non devono essere necessariamente le prime 3, ma le opzioni restano valide (se non vengono utilizzate prima) per tutta la carriera del giocatore. Spesso le società le sfruttano all’inizio della carriera, per ritardare la firma del contratto del giocatore e assicurarselo per un ulteriore anno.

a prima è che militando in questo roster i giocatori hanno l'opportunità di iniziare ad accumulare esperienza nella massima serie.

Quello che permette loro di raggiungere la firma del contratto in anticipo rispetto ai 3 anni, è l'uscita dallo stato di *Rookie*. Con questo termine (in italiano *recluta*) si indicano generalmente i giocatori alle prime armi: nella Major League Baseball un giocatore rientra in questa categoria finchè, nelle stagioni precedenti: (a) effettua meno di 130 turni di battuta (giocatore di campo) o lancia meno di 50 riprese (lanciatore) nella massima lega; oppure (b) finchè non registra 45 presenze in un roster attivo. Superata la soglia acquisiscono infine il diritto di firmare il contratto di Major League.

1.2 Giocatore professionista di baseball come lavoratore

Giunti alla firma del contratto, tra il giocatore di baseball professionista e la società di Major League che lo assume, vige un vero e proprio rapporto lavorativo. Per questo motivo andremo ora a fare un'analisi approfondita della figura del giocatore professionista in quanto lavoratore. Facciamo però prima un passo indietro e, dopo aver parlato dello sport come parte integrante della formazione scolastica, introduciamo il concetto di professionismo nello sport.

1.2.1 Origine degli Sport Professionistici

Dicesi sport professionistico quell'attività sportiva dove l'atleta riceve un pagamento per le proprie performance, e questo è ciò che lo differenzia dal dilettantismo. Rispetto al passato però, oggi gli atleti ricevono dei veri e propri stipendi.

Possiamo ricondurre la nascita del professionismo sportivo moderno agli eventi relativi al secondo Dopoguerra: il miglioramento del tenore di vita, che ha garantito via via più tempo libero alle persone, e l'evoluzione tecnologica, che ha permesso un forte sviluppo dei mass media, hanno portato alla presenza di un pubblico sempre più ampio. La conseguenza di tutto ciò è che la crescita di attenzione e di interesse da parte della gente ha portato le organizzazioni e le società sportive ad avere a che fare con quantità di entrate sempre maggiori, risultato che a sua volta ha permesso a sempre più atleti di fare del proprio sport la loro principale carriera.

1.2.2 Giocatore come lavoratore subordinato

Avendo introdotto l'idea di stipendi e carriere lavorative nello sport, non resta che iniziare a parlare del contratto. All'interno del Collective Bargaining Agreement (2012)⁴ si leggono i dettagli del *Major League Uniform Player's Contract*, il generico contratto utilizzato per assumere i giocatori: suddiviso in 12 sezioni, esso tratta in ciascuna di queste, tutti i diversi aspetti dell'accordo (dall'assunzione del giocatore, alla terminazione o al rinnovo del contratto; dal servizio che il giocatore dovrà prestare, al pagamento che riceverà per farlo; e così via). Presentiamo in appendice (...) una copia autentica del contratto per la consultazione (Gaines 2011), mentre di seguito andiamo ad analizzare alcune delle sezioni di questo contratto.

Nella prima sezione, *Employment*, si legge che "Il Club assume il giocatore (il quale accetta) per rappresentare la squadra di baseball durante gli anni definiti nel contratto" e questo accordo di rappresentanza fa riferimento a qualsiasi partita o manifestazione a cui la società deciderà di partecipare (Collective Bargaining Agreement 2012:277). Nella seconda, *Payment*, la società stipulante si impegna a pagare la somma specificata semi-mensilmente per tutta la durata della stagione, e, nel caso il giocatore rimanga solo temporaneamente, il pagamento avverrà sulla base del numero di giorni in cui egli ha prestato servizio (Collective Bargaining Agreement 2012:277-279). Cosa significa quanto appena detto? Società e giocatore concordano nel contratto l'ammontare del salario annuale; se però il giocatore non gioca la totalità della stagione, ma soltanto una porzione della stessa, allora riceverà di conseguenza soltanto una percentuale del salario concordato, proporzionalmente al numero di giorni di servizio.

Secondo quanto letto nelle due sezioni del contratto citate, possiamo giungere alla conclusione che il giocatore professionista di baseball della Major League Americana è completamente identificabile come un generico lavoratore subordinato. Troviamo inoltre al riguardo una perfetta corrispondenza con il diritto italiano, il quale (citando la legge n.91 del 23/03/1981, denominata "Norme in materia di rapporti tra società e sportivi professionisti") afferma che "La prestazione a titolo oneroso dell'atleta costituisce oggetto di contratto di lavoro subordinato".

⁴Il Collective Bargaining Agreement (o Basic Agreement) è il contratto collettivo, in vigore tra la Major League Baseball Players Association e la Major League Baseball, che contiene la regolamentazione relativa al rapporto di lavoro tra società e giocatori e che descrive la struttura finanziaria del gioco. La prima versione risale al 1968 (primo sport professionistico in cui è stato istituito questo tipo di accordo) e da quel momento è sempre stato sottoposto ad un aggiornamento quinquennale, in modo da risultare sempre attuale ed in grado di assicurare i diritti necessari ad ogni nuova generazione di giocatori.

1.2.3 Giocatore come lavoratore specializzato

Proseguendo nella lettura del contratto di Major League arriviamo alla quarta sezione, dove, alla voce *Ability*, leggiamo: “Il giocatore dichiara, e condivide con la società, di possedere *skills* eccezionali ed uniche in quanto giocatore di baseball e che il servizio che dovrà esercitare è di carattere speciale, inusuale e straordinario, a tal punto che è impossibile rimediare e compensare a un danno secondo la legge, e che quindi la rottura del contratto da parte del giocatore causerebbe alla società danni irreparabili. Pertanto il giocatore accetta che il club sia intitolato a determinati poteri, atti alla prevenzione di questa rottura, uno su tutti l’aver l’esclusività del servizio del giocatore su qualunque altra persona o società durante la durata del contratto” (Collective Bargaining Agreement 2012: 280).

Queste “*skills* eccezionali” di cui parla il contratto, sono senza dubbio abilità tecniche (a volte innate, ma spesso derivanti da anni di allenamento) che un giocatore deve necessariamente avere e dimostrare per diventare professionista. Allo stesso tempo però, una volta raggiunto il traguardo della professionalità, il giocatore è metaforicamente soltanto al primo giorno di lavoro. Infatti, una volta timbrato il cartellino, la prima preoccupazione (ma anche il primo obiettivo) diventa il mantenersi all’altezza della categoria e della concorrenza.

Dopo essere passato da attività d’intrattenimento a puro agonismo, lo sport diventa quindi ora anche priorità lavorativa: di conseguenza l’impegno richiesto aumenta considerevolmente e l’approccio dell’atleta all’attività sportiva si traduce nel dedicare all’allenamento tutto il tempo necessario. Per riuscire a mantenersi continuamente competitivi, i giocatori dovranno quindi costantemente perfezionare le proprie abilità, migliorare la propria condizione fisica ed incrementare la propria esperienza: e questo ci permette di inserirli nella categoria dei lavoratori fortemente specializzati.

1.2.4 Giocatore come lavoratore con vita lavorativa breve

Il terzo aspetto che andiamo a considerare è la durata della carriera. In quanto fattore comune a molti sport professionistici (ma anche per esempio ai broker finanziari, ai piloti d’aereo e ai creativi), anche il baseball presenta una lunghezza media di vita lavorativa molto bassa, pari a 5.6 anni. Considerando che solamente poco più di 300 giocatori dal 1871 ad oggi sono arrivati a giocare tra le 18 e le 27 stagioni, e che in media ogni anno si parla di 900-1000 giocatori in tutta la Major League Baseball, possiamo affermare e confermare che la stragrande maggioranza degli atleti percorre carriere brevi, se non brevissime.

Uno studio effettuato, da William Witnauer della State University of New York a Buffalo e da Richard Rogers e Jarron Saint Onge della University of Colorado, ha inoltre sollevato due ulteriori questioni relative all'aspettativa della durata della carriera (Roberts 2007)⁵.

Il primo risultato è relativo alla probabilità che la carriera del giocatore termini già alla fine della prima stagione in Major League, ma la cosa interessante è vedere come questa probabilità cambi notevolmente nei soggetti studiati. Il risultato dell'analisi è che un giocatore che debutta a 20 anni terminerà la propria carriera a fine stagione soltanto nel 10% dei casi, però se l'esordio avviene anche solo un anno più tardi (a 21) la probabilità aumenta già al 13%, e arriva addirittura al 36% per chi esordisce a 28 anni (età media dei giocatori in Major League). Si può quindi concludere che, come poteva essere facilmente deducibile, più è giovane il giocatore al momento del debutto più ciò sarà vantaggioso per la durata della sua carriera lavorativa.

L'altro aspetto degno di nota è che con l'evolversi della carriera i giocatori sviluppano prospettive di vita lavorativa diverse. Mentre nella prima parte dello studio si erano focalizzati sul primo anno in Major League, in questa seconda fase l'attenzione si sposta su come varia la probabilità di terminare la carriera a fine stagione, all'aumentare degli anni giocati. Considerando che nel 20% dei casi un atleta raggiunge la fine della propria carriera già al termine della prima stagione, ma che questa percentuale diminuisce fino all'11% durante il secondo anno (e cala ancora ogni anno che passa), la conclusione è che mentre un rookie può aspettarsi di giocare 5.6 anni⁶, un giocatore già alla terza stagione può aspettarsi di giocare altri 6 anni.

Il risultato di questa seconda parte dello studio sembra però descrivere una realtà ben diversa da una categoria di lavoro dalla carriera breve. La lunghezza media pari a 5.6 farebbe infatti pensare che all'aumentare delle stagioni giocate il giocatore veda sempre più vicina la fine della propria carriera; ma i dati dimostrano che così non è. Andiamo di seguito a spiegare quale sia l'altro fattore che incide sull'aspettativa e sulla prospettiva della vita lavorativa dei giocatori di baseball di Major League.

1.2.5 Disparità del potere contrattuale

Facendo un passo indietro, riteniamo importante far notare quanto sia degna di nota ed insolita la disparità del potere contrattuale tra le parti di questo specifico contratto. Diventa quindi doveroso effettuare un approfondimento mirato per spiegare nel dettaglio il rapporto tra giocatori e società di Major League Baseball.

⁵ Abbiamo fatto riferimento ai risultati dello studio descritti da Sam Roberts nel suo articolo pubblicato nel 2007 sul New York Times.

⁶ Dato confermato anche da Sandler (2012).

Solitamente negli sport, quando società ed atleta firmano un contratto, quest'ultimo mantiene la sua validità fino al termine della durata specificata, ma allo stesso tempo ci sono due motivi per cui potrebbe terminare anticipatamente. La causa più estrema è sicuramente la rescissione dovuta ad inadempienze per mano di una delle due parti, ovvero, nel momento in cui società o giocatore non rispettano i patti concordati al momento della firma, la controparte ha il diritto di rescindere il contratto. Il secondo motivo è invece rappresentato dal trasferimento del giocatore, che può avvenire a titolo definitivo (le prestazioni del giocatore vengono cedute ad un'altra società) oppure temporaneamente sotto forma di prestito, ma che in entrambi i casi richiede il consenso del giocatore stesso per quanto riguarda la destinazione.

La particolarità del contratto di MLB è che, al momento della firma, il giocatore non solo dichiara e sottoscrive di giocare solamente per la squadra stipulante (come visto in precedenza), ma si dichiara consapevole e d'accordo con quanto scritto nella sesta sezione. Alla voce *Assignment* si legge che la società possiede il diritto e il potere di cedere ad un'altra società le prestazioni del giocatore senza il suo consenso (Collective Bargaining Agreement 2012: 281).. Inoltre la società ha addirittura il potere di togliere il giocatore dal 40-man Roster e di rimandarlo nelle Minors, effettuando un cosiddetto *Outright Assignment*: il giocatore viene trasferito "a titolo definitivo" nella squadra di Minor League, e la prima volta che ciò avviene egli non ha possibilità di esprimersi, ed è costretto a scendere di categoria (Collective Bargaining Agreement 2012: 92; 103 ss.).

Ma perché un giocatore, sapendo di possedere abilità uniche e di essere altamente specializzato (come abbiamo affermato in precedenza), dovrebbe accettare di diventare potenzialmente semplice merce di scambio per il datore di lavoro? E perché, mentre per il giocatore è così difficile raggiungere il roster della massima serie, per la società risulta così facile riportarli nelle leghe minori? Le domande sembrano lecite, ma vedremo di seguito che la disparità di potere contrattuale che traspare da queste clausole, è in verità solo apparente, o quantomeno soltanto iniziale.

Il motivo per il quale la società presenta tutto questo potere nei confronti dei suoi giocatori è probabilmente riconducibile all'enorme numero di candidati presenti sul mercato tra cui essa può scegliere: da un lato il giocatore accetta "ad occhi chiusi" il contratto che gli viene proposto in quanto vede realizzarsi un sogno, dall'altro la società deve tutelarsi in diversi modi per aver scommesso su quel determinato giocatore, a discapito della miriade di alternative scartate. A questo punto entra però in gioco la regolamentazione della lega, la quale prevede una redistribuzione del potere contrattuale, andando in un certo senso a "premiare" quei giocatori che si dimostrano longevi: ecco che troviamo quindi, in questi premi, la causa della relazione positiva tra longevità del

giocatore e l'aspettativa sulla durata residua della sua carriera, cioè giocatori più longevi hanno una maggiore probabilità di mantenere più a lungo il contratto da Major League.

Passiamo di conseguenza a spiegare le 4 modalità secondo le quali può avvenire questa redistribuzione: modalità diverse tra loro, che seguono tuttavia la comune logica di annullare sempre più la disparità all'aumentare delle stagioni giocate.

1.2.5.1 Modalità di redistribuzione del potere contrattuale

Un primo intervento in questa direzione, seppur non precisamente regolamentato, è la possibilità che il giocatore (o il suo agente) riescano, trovando l'accordo con la società, ad inserire una **No-TradeClause**, una clausola che rende necessario il consenso del giocatore per un'eventuale cessione dello stesso ad un'altra società. Spesso le clausole di questo tipo sono limitate: a volte il trasferimento può essere limitato soltanto a determinati periodi durante la stagione, altre volte la limitazione può far riferimento alla destinazione (ad esempio il giocatore dichiara le potenziali società ritenute gradite o restringe il campo del trasferimento ad una regione geografica in particolare)(Collective Bargaining Agreement 2012: 78; 87)..

Esiste una vera e propria norma della regolamentazione di Major League che permette metaforicamente ai giocatori di "alzare la voce" nel rapporto con la società proprietaria. Stiamo parlando della **10&5 Rule**, la quale sancisce che un giocatore che sia stato per 10 intere stagioni in un roster attivo, e che sia rimasto nella stessa squadra negli ultimi 5 anni, non potrà più essere usato come merce di scambio senza la sua approvazione. Facciamo però notare ancora una volta che la durata media di una carriera è di 5.6 anni, pertanto l'agevolazione garantita da questa norma è un privilegio riservato alla minoranza di giocatori che supera i 10 anni nella massima serie.

Per parlare delle ultime due modalità, distogliamo l'attenzione dalle operazioni di scambio di giocatori tra un club e l'altro, e ci concentriamo invece sul rinnovo del contratto all'interno della stessa società. Iniziamo spiegando che ogni qual volta una società voglia rinnovare il contratto ad uno dei suoi giocatori, essa è tenuta a presentare l'offerta di rinnovo non oltre il 12 Dicembre. Inoltre, come avviene generalmente nella regolamentazione dei contratti, anche il modo in cui la Major League regola il rinnovo contrattuale presenta restrizioni per la società a favore di una maggiore tutela del giocatore (parte debole del contratto): la società che esprima la volontà di rinnovare il contratto a uno dei suoi giocatori, è obbligata a presentare un'offerta che non può essere inferiore all'80% del salario dell'anno precedente, né inferiore al 70% del salario di due anni prima.

Nel caso in cui un giocatore non ritenga opportuna l'offerta di rinnovo proposta dalla società, si può aprire allora un caso di **Salary Arbitration** (Collective Bargaining Agreement 2012: 17-22).. Nello specifico, entro un mese dalla proposta ricevuta, il giocatore può depositare i documenti per l'arbitrato, e da lì a 3 giorni giocatore e società devono presentare ciascuno una correzione alla proposta; se non si arriva ad una convergenza tramite le due correzioni, allora la questione viene presentata davanti al collegio arbitrale, il quale al termine dell'udienza assegna al giocatore un contratto di un anno ad uno dei due stipendi proposti.⁷Tuttavia, ancora una volta, questo discorso non vale per tutti, anzi sono soltanto due le categorie di giocatori che possono esercitare questo diritto:

- La prima di queste è formata da quei giocatori che presentano più di 3 anni di servizio all'attivo, ma meno di 6, nella Major League Baseball
 - La seconda è la categoria dei "Super 2": giocatori con meno esperienza alle spalle che vengono però "premiati" per il loro rendimento. Sono infatti giocatori con più di 2 anni di servizio (ma meno di 3), che devono però classificarsi tra i migliori di tutti gli eligibili all'arbitrato⁸.

L'ultimo strumento a favore dei giocatori, ma il primo in quanto a potere contrattuale fornito, è l'assumere il titolo di **Free Agent**, qualifica che li rende liberi di trattare con qualsiasi club essi vogliano (Collective Bargaining Agreement 2012: 86-94).. Ma ancora una volta questa qualifica è riservata a pochi:

- Un giocatore che abbia almeno 6 anni di servizio nella Major League Baseball e che sia privo di contratto per la stagione successiva.
 - Un giocatore che ha giocato per 3 stagioni in Major League e che subisce un *outright assignment*, o un qualsiasi giocatore che lo subisca per la seconda volta nella sua carriera, ha il diritto di:
 - 1) Rifiutare la decisione della società, rescindere il contratto con la stessa, e diventare immediatamente un free agent.
 - 2) Accettare la decisione, andare nella squadra di Minor League, e dichiararsi free agent tra la fine della stagione e il 15 Ottobre, sempre che entro tale data non sia stato nuovamente inserito in un roster di Major League.

⁷ Il contratto derivante dall'*Arbitration* sarà un contratto non garantito, pertanto il giocatore, se escluso dalla squadra entro il 16esimo giorno prima dell'inizio della stagione sarà soggetto solamente a 30 giorni di pagamento per buonauscita, se invece l'esclusione avviene nei 15 giorni che restano prima dell'inizio della stagione.

⁸ Fino al 2013, parlando dei migliori giocatori tra gli eligibili all'arbitrato, si faceva riferimento ai giocatori che occupavano il primo 13% della graduatoria. Dal 2013 la soglia si è allargata e ora si parla del primo 22%.

- Nel caso, il giocatore in esame, fosse precedentemente entrato a far parte della categoria dei “Super 2”, soltanto la prima delle due ipotesi sarebbe una via percorribile per la free agency.

CAPITOLO 2

COSTRUZIONE DEL DATASET

2.1 Lahman Baseball Database

2.1.1 Sean Lahman e il “Giornalismo di Precisione”

2.1.1.1 Giornalismo di precisione

Sean Lahman è un autore e giornalista Statunitense, nato il 9 Giugno 1968, che ha incentrato la propria carriera sul “Giornalismo di precisione”.

Il giornalismo di precisione, o data journalism, è la nuova frontiera del giornalismo, che vede quest’ultimo implementato di una serie di inchieste e di approfondimenti, realizzati tramite strumenti matematici, statistici e relativi alle scienze sociali e comportamentali (Thibodeaux 2011)..

In quanto nuova e moderna frontiera, fonda la propria esistenza sul web, traendone un doppio beneficio: da un lato tramite strumenti multimediali raggiunge un nuovo livello di interattività con i lettori, dall’altro il fenomeno dell’Open Data permette ai giornalisti di attingere ad un enorme quantità di informazioni dalle molteplici banche dati online (Thibodeaux 2011).

2.1.1.2 Campi di pertinenza

- *Computer assisted reporting [CAR] e data-driven journalism*: è la stesura di articoli che preveda l’utilizzo, e l’analisi tramite software, di grandissime banche dati. Il successo di computer, software e Internet ha sostanzialmente cambiato il modo di agire dei giornalisti: la routine prevede la raccolta delle informazioni all’interno di database, l’analisi dei dati pubblici tramite fogli di lavoro e programmi statistici, lo studio delle variazioni politiche e demografiche tramite sistemi di *mapping* informativo, la conduzione di interviste tramite e-mail, e la ricerca del background per gli articoli sul web (Cox 2000; Lorenz 2010).
- *Database journalism*: è un sistema di organizzazione dell’informazione che raccoglie articoli, notizie e informazioni, traducendoli in frammenti di dati, per poi riordinarli all’interno di database, che si differenzia dalla struttura “story-centric” di raccolta tradizionale (Holovaty 2006).
- *Data visualization*: è lo studio e la creazione di rappresentazioni visive dei dati, che risultino chiare ed efficienti. A causa del suo utilizzo di *infographics* per aiutare l’utente nella lettura e nell’analisi dei dati, è considerata da molte discipline l’evoluzione moderna della *comunicazione visiva*, la quale cerca di trasmettere messaggi (appunto comunicare) tramite l’uso di grafica, pittura, scrittura e video (Friendly 2008).
- *Infographics*: sono uno strumento del data visualization, e sono propriamente grafici finalizzati alla rappresentazione visiva dei dati, che vantano di chiarezza e velocità di

lettura grazie per esempio a grafiche che stimolino il sistema visivo umano e lo portino a riconoscere facilmente pattern e trend dei dati (Smiciklas 2012).

- **Interactive visualization:** è un ramo della visualizzazione grafica, in informatica, che permette (tramite una serie di strumenti) all'utente di analizzare i dati attraverso un'interazione diretta con una rappresentazione visiva degli stessi. Essa ricopre sia lo studio di come l'uomo interagisce coi computer, sia la ricerca e l'applicazione di metodi e strumenti grafici che facilitino l'interazione.
- **Serious games:** si tratta di videogiochi il cui primo obiettivo non è l'intrattenimento ma il valore pedagogico. Sono videogiochi definiti appunto "seri", poiché sono utilizzati in campi quali l'educazione, l'esplorazione scientifica, la progettazione civile (e altri ancora), condividendo diversi aspetti con le attività di simulazione svolte dalle stesse istituzioni (Djaouti, Alvarez, Jessel 2011).

2.1.1.3 Sean Lahman

Attualmente Lahman sviluppa database interattivi e articoli data-driven per il Rochester Democrat and Chronicle, e per lo stesso quotidiano dal 2011 scrive anche settimanalmente una colonna su tecnologie emergenti ed innovazione. Inoltre, in quanto uno dei principali promotori, fa spesso apparizioni pubbliche per parlare di *database journalism*, *data mining* e *database open source* (About Sean 2017).

Ciò che però l'ha reso famoso è il Lahman Baseball Database, a lui intitolato, ovvero una raccolta di statistiche riguardanti ciascuna squadra e giocatore nella storia della Major League Baseball americana. Iniziando nel 1995, è riuscito nell'impresa di costruire un database completamente gratuito, disponibile su internet ed interamente scaricabile, ed è così stato uno dei principali motori a dare il via ad una nuova era nella ricerca sul baseball, permettendo a chiunque di raggiungere ed utilizzare i dati grezzi. Oltre a promuovere la ricerca, il lavoro di Lahman ha permesso a vari videogiochi di simulazione di ricreare stagioni storiche sulla base di fatti e risultati effettivi.

Negli stessi anni Lahman diede anche vita alla prima enciclopedia del baseball online, sul suo sito Baseball Archive, che cambiò successivamente nome in Baseball-Reference.com. Attualmente invece lavora con la Società per la Ricerca sul Baseball Americano per coordinare i vari progetti di raccolta dati, tra cui si impegna fortemente nella costruzione di un database dettagliato ed esaustivo delle Minor Leagues, dal momento che la documentazione statistica su di esse è estremamente povera.

Il suo impegno nel documentare le statistiche della storia degli sport si è poi spinto anche oltre al baseball. Dalla fine degli Anni '90 ha infatti contribuito alla creazione e pubblicazione di enciclopedie dedicate anche a football professionistico, basket professionistico, e tennis. Inoltre nel 2008, il suo libro “*The Pro Football Historical Abstract*” ha ricevuto il Premio Nelson Ross, che viene assegnato ogni anno dalla Pro Football Researchers Association per “risultati eccezionali nella ricerca e nella storiografia del Football Professionistico” (About Sean 2017).

2.1.2 Descrizione del database

Il Lahman Baseball Database⁹ nella sua totalità è comprensivo di 24 singoli dataset che possono essere divisi e riassunti in 6 sotto-categorie (Seanlahman.com 2017).

Informazioni personali dei giocatori: tre diverse tabelle che contengono informazioni anagrafiche dei giocatori di Major League, informazioni riguardo alle scuole dove hanno studiato, e riguardo ai college dove hanno giocato in precedenza (Master¹⁰; Schools; CollegePlaying).

Dati e performance dei giocatori in Major League: un dataset contenente i salari dei giocatori che, anno per anno, rientrano nel 25-man roster (Salaries¹⁰), quattro tabelle che riportano i ruoli ricoperti in campo dai giocatori e le loro statistiche difensive (Appearances¹⁰; Fielding; FieldingOF; FieldingPost), due dataset riguardanti i lanciatori (Pitching; PitchingPost), e altri due dove troviamo le statistiche relative alla battuta (Batting¹⁰; BattingPost).

Riepilogo di Playoff e All Star Game: un dataset che registra nomi e record delle squadre vincitrici delle varie fasi della Postseason (SeriesPost), e un altro che fornisce dati relativi a convocazioni, presenze e ruoli ricoperti dai giocatori nell'All Star Game, partita tra l'élite delle due divisioni (National e American League) che avviene a metà stagione (AllstarFull).

Premi e riconoscimenti: quattro tabelle che racchiudono i premi stagionali e i voti relativi (AwardsManagers; AwardsPlayers; AwardsShareManagers; AwardsSharePlayers), e un'ulteriore tabella contenente i dati relativi alle votazioni per l'ingresso nella Hall Of Fame della MLB (HallOfFame).

Allenatori: due dataset con dati, statistiche e risultati dei vari manager (Managers; ManagersHalf).

Squadre: tre tabelle descrittive caratteristiche, dati e statistiche delle squadre (Teams¹⁰; TeamsFranchises; TeamHalf).

⁹ Insieme al database abbiamo scaricato anche un file che spiegava il contenuto dello stesso (Read Me 2016).

¹⁰ I cinque dataset sottolineati nel testo, sono quelli che, tra tutti quelli disponibili, sono stati utilizzati ed analizzati nel nostro studio.

2.2 Dal Lahman Database al nostro dataset

Il lavoro svolto da Sean Lahman ha reso raggiungibile a tutti un database completo di numerose informazioni utili. Per i nostri obiettivi, abbiamo selezionato quei dataset e quei dati che sarebbero stati utili al nostro studio per creare un nuovo dataset che contenesse tutte le informazioni necessarie a valutare gli effetti della performance sportiva sul salario.

Dopo aver selezionato 5 dei 24 dataset, tramite diversi merge (alcuni secondo i nomi dei giocatori, altri secondo quelli delle squadre e altri semplicemente secondo l'anno) siamo arrivati a creare la nostra tabella dati. Un totale di 36.038 osservazioni, ognuna delle quali rappresenta un diverso giocatore della Major League Baseball in un determinato anno tra il 1985 e il 2015.

2.2.3 Le variabili

Per quanto riguarda le colonne della mia tabella valori, il dataset completo contiene 82 variabili, alcune delle quali verranno usate direttamente nella stima del modello, altre che sono state punto di partenza per la creazione di altre utili variabili, e altre ancora che sono state inserite nel dataset in modo da renderne più facile e lineare la lettura. Lo stesso dataset è stato infine organizzato in modo che le variabili si suddividessero in 6 sottogruppi, sempre per facilitarne l'utilizzo e la lettura.

2.2.3.1 Variabili relative alle apparizioni in battuta

In questo gruppo troviamo le variabili *G* e *AB* (*numeriche*) che sono rispettivamente il numero di partite giocate e il numero di turni di battuta effettuati, 12 variabili (*numeriche*) che tengono traccia di tutti gli esiti dei turni di battuta del giocatore in quel determinato anno, e 3 variabili (*numeriche*) che parlano invece delle statistiche del giocatore in quanto "corridore" sulle basi.

2.2.3.2 Variabili relative alle apparizioni in difesa

Le prime 4 variabili (*numeriche*) sono riassuntive delle presenze del giocatore, indicando rispettivamente il numero di partite giocate, il numero di partite in cui il giocatore è subentrato dalla panchina, il numero di partite giocate in difesa e quello delle partite giocate solamente in battuta. Le rimanenti 12 variabili (*numeriche*) tengono il conteggio del numero di partite giocate nei 9 ruoli difensivi, tenendo inoltre in considerazione il ruolo del DH (battitore designato, che effettua durante tutta la partita i turni di battuta al posto del lanciatore), quello del PH (*Pinch Hitter*, giocatore che subentra in battuta al posto di un compagno durante una partita), e quello del PR (*Pinch Runner*, giocatore che sostituisce un compagno nella corsa sulle basi, sempre durante una partita).

Utilizzando 10 di queste 12 variabili¹¹, abbiamo quindi creato la variabile *pos* (*stringa*) che ci indica il ruolo (in inglese *position*) ricoperto maggiormente dal giocatore durante la stagione di riferimento.

2.2.3.3 Variabili descrittive del giocatore

Dal momento che i giocatori vengono identificati in modo unilaterale tramite un codice alfanumerico che troviamo nella variabile *ID*, tra le variabili descrittive sono state inserite *nameFirst* e *nameLast* (*stringhe*) in modo da riuscire ad identificarne il nome completo in qualsiasi momento risultasse necessario. Altre due variabili anagrafiche del giocatore sono *birthYear* (*stringa*) e *birthCountry* (*stringa*), che indicano rispettivamente l'anno di nascita e lo Stato natale. Sono state infine inserite in questo gruppo anche le variabili *throws* e *bats* (*stringhe*) che indicano se il giocatore sia destro o mancino in difesa e in attacco, con l'aggiunta per quanto riguarda l'attacco dell'opzione ambidestro (o più tecnicamente *switch-hitter*).

2.2.3.4 Variabili relative alla squadra del giocatore

Le prime 3 sono variabili (*stringhe*) che identificano rispettivamente la squadra del giocatore, la lega in cui gioca la squadra (American o National League), e la division (East o West, dal 1994 anche Center). Successivamente abbiamo 4 variabili dummy che riassumono l'esito del campionato, indicando rispettivamente se la squadra ha vinto la division, la wild card¹², la lega e le World Series. Infine 2 variabili (*numeriche*), la prima, *Rank*, indicante la posizione raggiunta in classifica, l'altra, *post*, una dummy costruita in modo che sia 1 quando la squadra partecipa alla post-season (ovvero quando *WCWin* e *DivWin* sono uguali a 1).

2.2.3.5 Variabili indicanti statistiche della battuta

A partire dai dati riguardanti le apparizioni in battuta, siamo passati alla costruzione di statistiche che avessero un maggior valore esplicativo delle performance dei giocatori.

La prima di cui andiamo a parlare è la **media battuta** (*bavg* nel nostro dataset):

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

¹¹ Abbiamo escluso il Pinch Hitter ed il Pinch Runner dai ruoli da tenere in considerazione, in quanto non rappresentano ruoli principali del gioco, ma soltanto ruoli derivanti da scelte strategiche della singola partita e quindi non identificativi del giocatore.

¹² Generalmente nello sport, la wild card è il permesso di partecipazione ad una determinata gara o competizione accordato a squadre o atleti che non si sono qualificate. Nella MLB, la competizione a cui essa si riferisce è la fase finale della stagione (*post-season*), alla quale per regola sono ammesse le squadre vincitrici della propria *division* (in tutto 6). Vengono quindi messe in palio 2 wild card ogni anno, una per ciascuna lega, in modo da arrivare ad 8 squadre qualificate alla post-season: in ogni lega, le due squadre che presentano il miglior record vittorie/sconfitte (escluse le 3 vincitrici delle division) si affrontano in una gara secca, nella quale chi vince accede alle fasi finali.

È una statistica derivante dal semplice rapporto tra il numero di battute valide (H) e il numero di turni alla battuta (AB), che indica pertanto la percentuale di volte in cui un battitore effettua una battuta che gli permette di arrivare salvo sulle basi.

Un'altra statistica della battuta è l'**On-base Percentage** (obp), che indica, come si può ben capire dal nome, la percentuale di arrivi in base del giocatore:

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

A differenza della media battuta, essa prende in considerazione (oltre alle battute valide) anche le basi per ball (BB) e le volte che il battitore guadagna la base dopo esser stato colpito dal lanciatore (HBP , dall'inglese *Hit By Pitch*). Inoltre al denominatore del rapporto, al numero di turni alla battuta vengono sommati il numero di basi ball, colpiti, e volate di sacrificio¹³ (SF), tutti e 3 esiti che vengono esclusi dal conteggio dei turni di battuta.¹⁴

Una statistica che invece tiene conto della potenza e della “produttività” del battitore (in quanto a numero di basi raggiunte) è la **Slugging Percentage** ($slug$), calcolata tramite il rapporto tra il numero totale di basi raggiunte e il numero di turni alla battuta:

$$slug = \frac{(1B) + (2 \times 2B) + (3 \times 3B) + (4 \times HR)}{AB}$$

Quello che fa il numeratore di questo rapporto è dare un peso diverso alle diverse battute valide in termini di basi raggiunte: 1 al *singolo*, 2 al *doppio*, 3 al *triplo*, e 4 al *fuoricampo* (considerando che 1, 2, 3 e 4 sono il relativo numero di basi). Nel dataset avevamo già le variabili relative a doppi (B), tripli(K), e fuoricampi(HR), mentre ci mancava quella relativa ai singoli, che abbiamo pertanto calcolato per esclusione dal numero totale di valide: $H - (B + K + HR)$.

Abbiamo inserito anche una variabile che, rispetto alle 3 appena presentate (statistiche basilari della battuta nel gioco del baseball), ha una storia più recente. Si sente parlare per la prima volta nel 1984 della **On-base plus slugging** (OPS) nel libro “The Hidden Game of Baseball” (Thorn, Palmer 1984); successivamente il New York Times iniziò a scrivere dei leader in questa statistica settimanalmente nella colonna chiamata “By the Numbers”, e continuò a farlo per quattro anni; poi via via sempre più giornalisti si sono appassionati a questo dato, che ha quindi assunto sempre più popolarità, finché nel 2004 è stato inserito tra le statistiche presenti sulle carte da collezione *Topps* dei giocatori di baseball (Schwartz 2004). Per quanto riguarda il calcolo della statistica, essa è semplicemente la somma tra le due precedenti, come lascia ben intendere il suo nome:

¹³ In inglese *Sacrifice Fly*, la volata di sacrificio è un particolare esito del turno di battuta, nel quale il battitore viene eliminato al volo, ma a seguito di questa eliminazione il corridore che occupava la terza base riesce a segnare un punto.

¹⁴ Mentre la base per ball ed il colpito sono esclusi dal conteggio in quanto non rappresentano propriamente un merito del battitore, la volata di sacrificio è esclusa perché al contrario penalizzerebbe la statistica del battitore (perché verrebbe classificato come un eliminato, quando invece la sua battuta ha permesso la segnatura del punto).

$$OPS = obp + slug$$

Convenzionalmente il modo in cui viene rappresentato il dato di queste statistiche è sempre lo stesso: il risultato dei vari rapporti viene espresso utilizzando 3 cifre decimali precedute da un punto (senza la 0 iniziale) e viene letto come se fosse un numero intero (ad esempio, un giocatore che presenti una *bavg* pari a .500, si dirà in gergo che “batte 500”). Ci sono però alcune differenze per quanto riguarda la distribuzione dei possibili valori e dei valori medi raggiunti dai giocatori in Major League:

- Iniziando dal confronto tra media battuta e on-base percentage, possiamo dire che entrambe hanno lo stesso range di valori (.000 - 1.000, da *zero* a *mille*), ma la seconda, considerando un numero maggiore di esiti positivi del turno di battuta, presenta mediamente valori più alti. I leader assoluti riguardo ad una singola stagione sono infatti rispettivamente Hugh Duuffy (1894) con una media battuta pari a .439 (Baseball Reference 2017a) e Barry Bonds (2004) con una on-base percentage pari a .609 (Baseball Reference 2017b).
- Passando alla slugging percentage, troviamo il primo cambio nel range dei valori. Rispetto alla media battuta dove tutte le battute valide erano considerate allo stesso modo, in questa statistica ogni valida ha un peso diverso (da 1 a 4), pertanto i valori possibili variano da *zero* a *quattromila* (.000 - 4.000). Troviamo ancora una volta Barry Bonds come leader assoluto, con la slugging percentage di .863 ottenuta nel 2001 con un totale di 411 basi raggiunte in 476 turni di battuta (Baseball Reference 2017c).
- Infine la on-base plus slugging, essendo la somma delle due vedrà il suo limite massimo alzarsi ancora fino a 5.000. Per rendere più facile la lettura di questa statistica faremo riferimento alla classificazione proposta da Bill James nel suo saggio intitolato “The 96 Families of Hitters”, la quale redistribuisce i giocatori tra 7 categorie (James 2009: 24 ss.):

Category	Classification	OPS Range
A	<i>Great</i>	.900 – 5.000
B	<i>Very Good</i>	.834 - .899
C	<i>Above Average</i>	.767 - .833
D	<i>Average</i>	.700 - .766
E	<i>Below Average</i>	.634 - .699
F	<i>Poor</i>	.567 - .633
G	<i>Very Poor</i>	.000 - .566

Successivamente abbiamo inoltre creato: 3 variabili che calcolassero rispettivamente la differenza tra *obp*, *slug*, e *OPS* con la media battuta (*avg*) (per provare ad isolare rispettivamente l'abilità nell'arrivare in base, la potenza, e la somma delle due, dalla semplice media battuta), e altre cinque variabili che mi permettessero di avere la performance media per partita, dividendo le variabili originarie per il numero di partite giocate dal giocatore. Infine ho creato la variabile *sbp* (Stolen Bases Percentage), effettuando il rapporto tra le basi rubate (*SB*) e il numero di eliminazioni subite su rubata (*CS*). Essendo tutte le variabili create, medie o rapporti di variabili numeriche, sono di conseguenza anch'esse tutte *variabili numeriche*.

2.2.3.6 Variabili relative alla carriera in Major League dei giocatori

In questo gruppo abbiamo le 6 prime variabili che identificano il giocatore: *ID* (codice del giocatore), *yearID* (anno di riferimento), *team* (squadra), *ruolo*, *age* (età)¹⁵ e *bat* (come bats, ma "encode").

Troviamo inoltre una serie di variabili riguardanti lo stipendio del giocatore, *salary*, *salario_reale*, *sal_min*, *sal_min_reale* e *differenza_sal_minimo*, per ognuna delle quali abbiamo anche creato la relativa variabile contenente il logaritmo naturale (aggiungendo *ln* davanti al nome della variabile di riferimento). Le variabili *salario_reale* e *sal_min_reale* sono state create utilizzando i valori del CPI (Indice dei Prezzi al Consumo) statunitense, grazie ai quali siamo stati anche in grado di creare la nostra variabile numerica *CPI* (McMahon 2017); i due salari sono quindi stati deflazionati e portati al valore reale in dollari 2015 moltiplicandoli per il rapporto tra il CPI del 2015 e quello dell'anno in oggetto, così da attualizzare i valori al 2015. Inoltre inseriamo tra le variabili relative al salario anche *first*, una dummy che è uguale a 1 solamente nell'anno del primo stipendio del giocatore, e *roster*, un'altra variabile dummy che è uguale a 1 se lo stipendio è diverso da 0 (ovvero se il giocatore fa parte del *25-man Roster*).

Infine abbiamo 4 variabili che ci parlano della carriera del giocatore: *stint* (1,2,...) indica per quante squadre ha giocato durante la stagione di riferimento il giocatore osservato; *anni* (0,1,...) indica quanti anni di gioco in MLB ha alle spalle il giocatore; *Gtot* è il totale cumulato delle partite giocate dal giocatore in MLB; infine *debutto* indica l'anno in cui il giocatore ha debuttato nella massima serie.

¹⁵ Utile per calcolare l'età del giocatore (variabile che mancava nel database), è stata la variabile *birthYear*, menzionata nel gruppo delle variabili descrittive: sottraendo semplicemente l'anno di nascita del giocatore all'anno di riferimento dell'osservazione, siamo riusciti a trovare l'età del giocatore in quella stagione.

CAPITOLO 3

METODOLOGIA

3.1 Uno sguardo alla letteratura esistente

Nel 2003, Michael Lewis pubblicò il libro *Moneyball: The Art of Winning an Unfair Game*, il quale diede il via ad un'immediata ed inesorabile mutazione della concezione economica e finanziaria del gioco del baseball (Lewis 2003). Quella raccontata nel libro è la storia di Billie Beene, general manager degli Oakland Athletics, e della sua rivoluzionaria strategia di scelta dei giocatori, basata esclusivamente sull'utilizzo di statistiche sulle performance. L'obiettivo di Beene era quello di mettere insieme un roster di giocatori che potesse vincere la maggior parte delle partite, minimizzando completamente il costo per costruirlo. La strategia che intraprese fu quella di affidarsi completamente alla *sabermetrica*¹⁶, ed i risultati ottenuti nella stagione del 2002 gli diedero assolutamente ragione.

Il successo della strategia, e la pubblicazione del libro a un anno di distanza, hanno quindi radicalmente modificato le politiche societarie dei club di Major League. Questi ultimi si ritrovano oggi, da un lato a lavorare tutti nella stessa direzione per riuscire (attraverso l'uso e la creazione di sempre più statistiche) a riassumere in un'unica misura la produttività dei giocatori, dall'altro a perfezionare una strategia specifica e differente (ognuno dagli altri) per mantenere un vantaggio competitivo. La coesistenza di queste due facce della strategia è ciò che ha permesso il continuo progresso del movimento iniziato da Beene.

Anche all'esterno della Major League, la ricerca sul baseball si è quindi basata sulla ricerca della formula perfetta per il calcolo del valore dei giocatori. Si contano infatti numerose pubblicazioni che vanno a studiare e ad analizzare il salario: alcune studiano la regressione del salario su tutti i fattori che lo influenzano, altre invece si interessano simmetricamente sull'impatto che ha lo stipendio sulle performance.

. Per esempio lavori come quello di Nicholas Dorsey in cui si cerca di trovare quali siano le statistiche di performance più giuste da utilizzare, anche differenziando da un ruolo all'altro (Dorsey, 2015). Oppure, per quanto riguarda gli studi "simmetrici", troviamo un esempio nella ricerca effettuata da Werner e Mero, nella quale gli autori si interrogavano sugli effetti che le condizioni di equità interna, esterna e tra dipendenti¹⁷ potessero avere sulle performance dei giocatori nella Major League Baseball (Werner, Mero 1999).

¹⁶ La *sabermetrica* è la scienza che analizza empiricamente il gioco del baseball, nello specifico utilizzando numerose statistiche che misurano le performance di gara dei giocatori. Ciò che fanno i *sabermetrici* è estrapolare dati da quello che accade in tempo reale sul campo di gioco durante le partite, per poi raccogliarli ed analizzarli. Il nome della scienza fu coniato da Bill James, uno dei pionieri e dei principali esponenti della stessa, e deriva dall'acronimo SABR che sta per Society for American Baseball Research (fondata nel 1971).

¹⁷ Per *equità* si intende il confronto equo tra diversi salari:

- quella *esterna* confronta lavoratori con la stessa occupazione, ma diverso datore di lavoro
- quella *interna* confronta lavoratori con diverse mansioni, ma stesso datore di lavoro
- quella *tra dipendenti* confronta lavoratori che svolgono le stesse mansioni nella stessa azienda.

Generalmente possiamo affermare che i metodi e le logiche secondo cui le società di Major League pagano i loro giocatori rappresentano un argomento di forte interesse, ed inoltre molti degli studi pubblicati presentano l'intenzione di svelare nuovi e possibili fattori da prendere in considerazione, stimandone l'impatto sugli stipendi. Di seguito citiamo quindi alcuni di questi studi che si basano sugli stessi argomenti trattati nel nostro lavoro durante la contestualizzazione economica.

Link e Yosifov nel loro studio si sono interrogati sull'importanza della sicurezza del posto di lavoro (Link, Yosifov 2012). In altre parole si sono chiesti se i giocatori della Major League Baseball fossero disposti a rinunciare a compensazioni monetarie più alte (come retribuzione per le proprie performance), in cambio di garanzie sul lavoro. Quello che hanno trovato è che effettivamente i giocatori sono propensi ad accettare salari annuali minori per riuscire ad ottenere un contratto la cui validità si estenda per un numero maggiore di stagioni; propensione che senza dubbio è riconducibile alla breve vita lavorativa che caratterizza i giocatori professionisti di baseball.

Come diretta conseguenza della breve durata della carriera media, abbiamo quindi parlato del problema della disparità del potere contrattuale e dei metodi di redistribuzione applicati. Quello che fa Tyler Wasserman al riguardo, nel suo *Determinants of Major League Baseball Player Salaries*, è un'analisi approfondita sul ruolo dell'agente dei giocatori (Wasserman 2013). Inserendo una dozzina di agenti all'interno della regressione del salario, l'obiettivo di Wasserman è cogliere l'importanza e l'efficacia della gestione delle trattative (effettuata da un agente piuttosto che da un altro), in base all'impatto del relativo regressore sul salario.

Un ulteriore studio che si fonda sugli stessi temi discussi precedentemente in questo trattato è quello svolto da Bollinger e Hotchkiss, nel quale gli autori applicano alla Major League Baseball il modello teorizzato da Edward Lazear riguardo ai *risky workers*¹⁸ (Bollinger, Hotchkiss 2003; Lazear 1998). Infatti alcune delle condizioni necessarie per garantire la validità del modello di Lazear sono: (1) l'esistenza di premi monetari per la longevità dimostrata dai lavoratori, (2) il passaggio per un probationary period durante il quale il datore di lavoro acquisisce informazioni sulla produttività degli stessi, e (3) il vantaggio competitivo delle società, garantito da un'elevata quantità di informazioni proprietarie relative ai propri giocatori e dalla limitata mobilità degli stessi.

Ovviamente esistono centinaia di autori che invece decidono di trattare argomenti diversi, come ad esempio Palmer e King (Palmer, King 2006): essi hanno testato e scoperto che, nel baseball moderno, il modo in cui i giocatori vengono ricompensati per il loro contributo in campo non è più affetto da discriminazioni razziali (cosa invece comune in passato). Tuttavia, nonostante

¹⁸ Con il termine "risky workers" Lazear fa riferimento a quei lavoratori che presentano una produttività incerta, o ad alta variabilità.

l'argomento alla base dello studio sia diverso dai nostri, nella formulazione della regressione compaiono comunque alcuni dei dati contenuti nei nostri dataset, come la *slugging percentage*, gli *AB* (At Bat, ovvero turni di battuta) e gli anni di *esperienza*.

3.2 Formulazione del modello

3.2.1 Stima del modello sulle determinanti del salario

Definito l'obiettivo di studiare l'impatto dei diversi fattori che influenzano il salario di un giocatore professionista di baseball ed osservata con attenzione la letteratura esistente, la scelta del modello da utilizzare ricade quindi in un primo momento su una regressione lineare multipla, nella quale possiamo appunto inserire diversi regressori e che allo stesso tempo ci permette di isolare i singoli effetti che ciascuno di essi produce.

A seguire descriviamo allora l'impostazione teorica della regressione, la quale vedrà il salario come variabile dipendente (y), e tutta una serie di altre variabili, che utilizzeremo per la spiegazione della y , come variabili indipendenti (o regressori):

	$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$	(1)
	$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i$	(2)

Mentre l'equazione (1) è la generica formula della regressione lineare multipla, l'equazione (2) rappresenta la realizzazione della (1) specifica dell'individuo i . In essa abbiamo utilizzato la forma compatta, $\mathbf{x}_i \boldsymbol{\beta}$ (simboli che rappresentano rispettivamente il vettore dei regressori ed il vettore dei relativi parametri), come abbreviazione di $\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$. Facendo attenzione a questa formula capiamo inoltre che gli effetti dei regressori (β) che troveremo, saranno gli stessi per ciascun individuo, e ci diranno quindi in quale misura inciderà mediamente il singolo regressore sulla definizione del salario.

Prima di considerare il metodo di stima, è necessario considerare una fondamentale caratteristica dei dati utilizzati per questo studio, in particolare la variabile dipendente relativa al salario. Nello specifico, l'ammontare del salario percepito in ciascuna stagione sportiva, non è disponibile per tutti i giocatori.

```

use C:\Users\Luca\Desktop\TESI\Salaries.dta

merge m:m playerID yearID using "C:\Users\Luca\Desktop\TESI\Batting.dta"

Result                                     # of obs.
-----
not matched                                12,677
  from master                               792  (_merge==1)
  from using                               11,885 (_merge==2)
matched                                    27,246 (_merge==3)

```

Figura 3.1: Comando di Stata per il merge e relativo risultato

La prima operazione per la preparazione dei dati è stata quella di effettuare il merge tra il dataset contenente i salari dei giocatori e quello contenente le statistiche di battuta. Come si può ben vedere nella Figura 3.1, il risultato dell'operazione si è concretizzato in 27.246 match perfetti su 39.923 (circa il 68%), mentre in 12.677 casi mancavano i valori da uno dei due dataset. Per quanto riguarda le 792 osservazioni, nelle quali avevamo solo ed esclusivamente il valore del salario, ma non le statistiche relative alla performance sportiva, si è deciso di eliminare il record, in quanto inutilizzabile per la stima del modello. In ogni caso, questa porzione di osservazioni rappresenta una piccola parte del nostro campione.

Necessitano invece particolare attenzione le rimanenti osservazioni mancanti, che rappresentano circa il 30% del campione. Queste 11.885 osservazioni, a differenza di quelle eliminate, non sono da ricondurre ad errori occorsi durante la raccolta e la trascrizione dei dati, oppure a situazioni particolari (ad esempio infortunio o squalifica del giocatore a contratto firmato), ma sono piuttosto il segnale di un potenziale caso di auto-selezione del campione.

3.2.2 Cosa si intende con “auto-selezione del campione”

L'inferenza statistica prevede che il campione di dati sia estratto in maniera casuale dalla popolazione d'interesse, in modo da garantire la rappresentatività statistica rispetto alle caratteristiche della popolazione. Se questa condizione non viene soddisfatta, c'è il rischio che le stime campionarie siano distorte, ossia che il loro valore atteso sia sistematicamente diverso da quello della popolazione oggetto di interesse.

Nel caso specifico, il nostro dataset è potenzialmente un censimento di tutti i giocatori con almeno una presenza in Major League, per cui si pone il problema delle caratteristiche delle osservazioni per le quali il salario non è disponibile. Se queste osservazioni sono “missing at random”, cioè non presentano differenze sistematiche dalle osservazioni per cui invece è disponibile il dato sul salario, allora si potrebbe considerare il dataset residuo come un campione

casuale della popolazione di riferimento. In realtà, questa selezione è tutt'altro che casuale, ossia c'è un motivo ben preciso per cui determinate osservazioni hanno un valore mancante per il salario. Il processo di auto-selezione verrà discusso più nel dettaglio nelle pagine seguenti, ma in sostanza i giocatori senza il dato sul salario sono quelli che pur giocando in Major League non hanno firmato un contratto per quella categoria. L'immediata conseguenza è che questi giocatori avranno con tutta probabilità caratteristiche ben diverse da quei giocatori che invece sono in possesso di contratto, ad esempio è verosimile che abbiano un'età inferiore e/o performance che ancora non hanno condotto le società ad offrire loro un posto più stabile.

In presenza di un processo di selezione, le stime campionarie del modello basate sul metodo dei minimi quadrati saranno distorte e inconsistenti. Esistono però modelli che permettono di risolvere o almeno mitigare il problema della selezione. Il problema dei dati mancanti in maniera non casuale può essere inquadrato in tre casi diversi.

Nel primo di questi, l'assenza di un dato corrisponde al valore zero, e tale valore assume un chiaro significato economico. Questo sarebbe il caso se il salario mancante fosse in realtà un salario nullo. In questa circostanza, la variabile dipendente può assumere il valore 0 o valori strettamente positivi, ma in entrambi i casi tale valore è spiegabile da una decisione economica. Sempre nell'ipotesi che il salario mancante fosse zero, sarebbe una situazione in cui il giocatore accetta di giocare gratis in Major League. Siamo nel caso del *corner solution model*, che assume una completa osservazione della variabile dipendente all'interno del campione, ed interpreta quindi gli "zeri" come effettive soluzioni a problemi di ottimizzazione economica degli individui (Woolridge 2015: 596-604). Nel caso specifico, il valore $y = 0$ significherebbe che il giocatore trova nel giocare gratuitamente la massimizzazione della propria utilità. Nel caso di questa interpretazione economica, diventa appropriato l'utilizzo di un modello Tobit. Secondo tale modello, si assume che la variabile dipendente segua una distribuzione teorica normale, dove viene però osservata solo la porzione con valori strettamente positivi. Gli "zeri" in questione sono valori che, se fossero osservabili, sarebbero valori negativi o nulli che completerebbero la distribuzione normale. In altre parole, il "salario di riserva" è nullo o negativo, ma non esistendo economicamente salari negativi, l'esito dell'accordo contrattuale sarebbe un salario nullo. Se da un lato l'assunzione di una distribuzione gaussiana consente di ovviare al problema della concentrazione degli zeri, è altrettanto evidente che l'interpretazione economica nel caso specifico sarebbe fallace.

Di seguito è rappresentata la specificazione del modello e la rappresentazione grafica della distribuzione del modello Tobit (in figura 3.2):

$$\begin{cases} y^* = \mathbf{x}\boldsymbol{\beta} + u, u|x \sim N(0, \sigma^2) \\ y = \max(0, y^*) \end{cases}$$

dove y^* è una variabile latente che soddisfa le classiche assunzioni del modello lineare. La nostra variabile dipendente osservata, y , sarà allora uguale a quella latente quando $y^* > 0$, e uguale a zero quando $y^* = 0$.

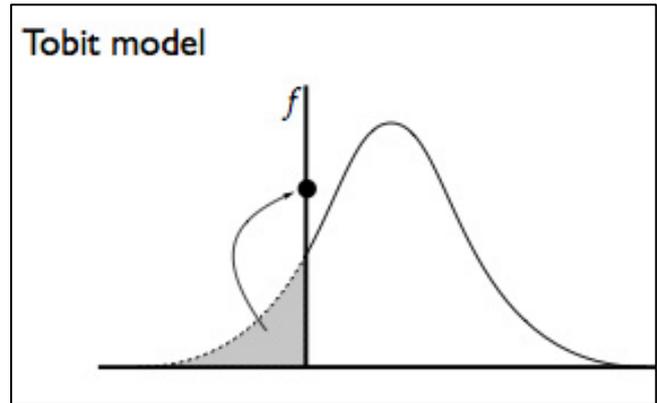


Figura 3.2: Grafico del modello Tobit

Una seconda interpretazione è quella della *variabile dipendente censurata* (Woolridge 2015: 609-613). Questo schema interpretativo si basa su una lettura statistica rispetto a quella economica precedente. In pratica con *censura* si intende che la variabile dipendente è osservabile soltanto sopra o sotto una determinata soglia (in questo caso sopra lo zero) e pertanto tutti quei valori esclusi dalla soglia vengono “compressi” al valore della soglia stessa. Non si richiede che i valori non osservati abbiano un significato economico, ma è sostanzialmente un problema di misura e osservabilità. In questo caso l’alta frequenza di zeri sarebbe quindi associata a tutte quelle osservazioni che presentavano una $y^* < 0$, e che sono quindi state riassunte in quel punto. Se dal punto di vista tecnico non esiste una chiara differenza con il modello Tobit, che comunque tratta tali valori come negativi, può esserci una chiave interpretativa sostanzialmente diversa, che può essere spiegata facendo riferimento alla popolazione di interesse. Nel caso del Tobit, il ricercatore non ha particolare interesse a vedere come la performance impatta su variazioni nel salario di riserva quando questo si trova al di sotto dello zero, in quanto tali performance non hanno alcun impatto economico reale. In altre parole, l’attenzione del modello Tobit è principalmente sulla relazione tra le performance e i valori strettamente positivi del salario, ossia sulla distribuzione condizionata della variabile dipendente. Se invece ci si riferisce ad una “censura”, i dati negativi sono interpretati come altrettanto informativi rispetto a quelli positivi, e nei modelli di regressione censurata l’attenzione è generalmente sull’intera distribuzione. Si tratta comunque di una distinzione semantica, perché è sempre possibile passare da una distribuzione all’altra. Non solo, ma il modello Tobit può essere descritto statisticamente come un caso specifico di regressione censurata con censura per valori nulli o negativi.

Detto che per una censura strettamente a zero i due modelli praticamente coincidono, le differenze interpretative possono diventare rilevanti anche tecnicamente nel caso in cui il punto di censura non fosse esattamente lo zero. Ad esempio, se il data-set non registrasse salari inferiori a

2000 dollari all'anno, si otterrebbero stime diverse considerando gli zeri come corner solution o considerando un censoring che potenzialmente ammette anche valori piccoli ma positivi.

In realtà, quello che accade ai nostri dati non sembra essere riconducibile a nessuno dei due casi descritti. Nel caso specifico, l'inquadramento più appropriato sembra quello dei dati mancanti che non corrispondono necessariamente a salari nulli o negativi, mentre in entrambi i casi discussi in precedenza il salario osservato era effettivamente zero. I dodicimila dati mancanti nella variabile del salario del nostro data-set hanno infatti la spiegazione legata al processo di rilevazione del dato anticipata in precedenza, cioè tali giocatori percepiscono un salario in quanto contrattualizzati nelle leghe minori, partecipano a partite di Major League in una sorta di tirocinio o apprendistato che, in caso di prestazioni positive, potrebbe portarli ad un contratto di Major League. Per la natura del dataset, i salari legati a contratti di *Minor League* (ovverosia l'apprendistato svolto principalmente nelle leghe minori, ma anche parzialmente nella lega maggiore) non sono osservati.

Il terzo modo di inquadrare il problema dei dati mancanti verrà trattato nel prossimo paragrafo.

3.2.3 Il modello di selezione di Heckman

James Joseph Heckman, nato il 19 Aprile 1944, è uno dei più influenti economisti al mondo che attualmente lavora alla University of Chicago (Heckmanequation.org 2017), dove ricopre i ruoli di (1) Henry Schultz Distinguished Service Professor of Economics nel dipartimento di economia (Uchicago.edu 2017a), (2) professore di legge all'interno della Law School (Uchicago.edu 2017b), e (3) direttore del Center for the Economics of Human Development (Uchicago.edu 2017c). Famoso per essere uno dei pionieri nei campi dell'econometria e della microeconomia, vince nel 2000 (insieme a Daniel McFadden¹⁹) il Premio Nobel per l'Economia per lo sviluppo della teoria e dei metodi per l'analisi dei campioni selettivi.

Gli studi effettuati dall'economista tra il 1976 ed il 1979 sulla distorsione derivante da selezione, lo hanno portato allo sviluppo di metodi statistici mirati alla soluzione di questo problema, tra cui l'introduzione della *Heckman correction* (correzione di Heckman) a lui intitolata.

In diversi dei propri studi, Heckman analizza l'offerta di lavoro femminile, facendo riferimento ai dati derivati dal National Longitudinal Survey (indagine nazionale su dati panel) del 1967, destinato alle donne tra i 30 ed i 44 anni (Heckman 1977). Il problema, come nel nostro caso,

¹⁹ McFadden vince il Premio Nobel per l'economia per il suo sviluppo di teoria e metodi per l'analisi dei modelli di scelta discreta.

era che nonostante il sondaggio fosse stato compilato attentamente dalle rispondenti, in alcuni casi non era presente alcun valore riguardo al salario percepito. Il motivo alla base della mancata osservazione del salario viene associato da Heckman all'esito di un'ulteriore variabile delle donne intervistate: la loro partecipazione o non partecipazione al mercato del lavoro.

La soluzione pensata e teorizzata dall'economista fu quella di esplicitare un'equazione di selezione del campione (4), da affiancare al modello di regressione del salario (3):

	$y = \mathbf{x}\boldsymbol{\beta} + u, \quad E(u \mathbf{x}) = 0$	(3)
	$s = 1[\mathbf{z}\boldsymbol{\gamma} + v \geq 0],$	(4)

dove $s = 1$ se osserviamo la y e zero altrimenti, e dove $\mathbf{x}\boldsymbol{\beta} = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k$ e $\mathbf{z}\boldsymbol{\gamma} = \gamma_0 + \gamma_1z_1 + \dots + \gamma_mz_m$.

Vediamo quindi come la (3) rappresenti una semplice regressione multipla, la quale, se prendessimo in considerazione un campione casuale, ci permetterebbe di stimare consistentemente (tramite il metodo dei minimi quadrati) il vettore dei parametri $\boldsymbol{\beta}$, e quindi l'influenza sul salario di ciascun fattore considerato. L'equazione di selezione (4) ci dice invece che la probabilità che un'osservazione rientri nel campione dipende da una serie di variabili osservate, z_h , e da una componente d'errore non osservata v . Diventa inoltre opportuno fare due assunzioni riguardo l'equazione di selezione.

La prima riguarda il vettore di variabili \mathbf{z} , e la condizione necessaria è che non ci sia perfetta sovrapposizione tra il vettore dei regressori del salario e \mathbf{z} . E' cioè necessario che almeno un elemento di \mathbf{z} , non sia incluso in \mathbf{x} : si richiede di fatto che \mathbf{z} sia esogeno rispetto alla (3), e quindi $E(u|\mathbf{x}, \mathbf{z}) = 0$. In pratica le variabili in \mathbf{z} agiscono come variabili strumentali. La seconda si riferisce invece al termine d'errore v , che ipotizzeremo essere (a) indipendente dai valori contenuti in \mathbf{z} (e quindi anche da \mathbf{x}), (b) distribuito secondo una normale standard, e (c) correlato con il termine d'errore della regressione, u .

Una volta esplicitate le equazioni, è ora il momento di andare a vedere come esse interagiscono tra loro, e soprattutto che effetto ha il processo di selezione delle osservazioni sul valore atteso del salario. Calcoliamo allora il valore atteso del salario y condizionatamente a \mathbf{z} e v :

	$E(y \mathbf{z}, v) = \mathbf{x}\boldsymbol{\beta} + E(u \mathbf{z}, v) = \mathbf{x}\boldsymbol{\beta} + E(u v) = \mathbf{x}\boldsymbol{\beta} + \rho v,$	(5)
--	---	-----

Si noti come il risultato finale sia stato raggiunto mettendo in pratica le assunzioni fatte precedentemente: l'assunzione sul vettore \mathbf{z} ha permesso che il valore atteso della componente stocastica u rimanesse condizionato solamente a v , quelle invece fatte sul termine d'errore ci hanno permesso di trovare la relazione $E(u|v) = \rho v$ (dove ρ è un generico parametro che tiene conto della correlazione tra u e v).

Arrivati a questo risultato, non è ancora possibile identificare v , ma si può utilizzare l'equazione (5) per calcolare un ulteriore valore atteso, $E(y|\mathbf{z}, s)$, che restringeremo al caso $s = 1$. Ovvero:

$$E(y|\mathbf{z}, s = 1) = \mathbf{x}\boldsymbol{\beta} + \rho E(v|\mathbf{z}, s = 1).$$

Se ci concentriamo sull'ultimo termine dell'equazione, esso è il valore atteso di una variabile distribuita secondo una normale standard²⁰, condizionatamente al fatto che essa sia maggiore di una determinata quantità.²¹ Allora, secondo le seguenti proprietà della distribuzione normale

	$z \sim N(0,1) \Rightarrow E(z z > c) = \frac{\phi(c)}{1 - \Phi(c)}$	
	$\phi(c) = \phi(-c)$	
	$1 - \Phi(c) = \Phi(-c),$	

dove ϕ e Φ sono rispettivamente la funzione di densità e la funzione di distribuzione della normale il valore atteso di cui stiamo parlando sarà uguale al rapporto tra la funzione di densità e la funzione di distribuzione della normale, calcolate nel punto $\mathbf{z}\boldsymbol{\gamma}$:

	$E(v \mathbf{z}, s = 1) = E(v \mathbf{z}, v > -\mathbf{z}\boldsymbol{\gamma}) = \frac{\phi(-\mathbf{z}\boldsymbol{\gamma})}{1 - \Phi(-\mathbf{z}\boldsymbol{\gamma})} = \frac{\phi(\mathbf{z}\boldsymbol{\gamma})}{\Phi(\mathbf{z}\boldsymbol{\gamma})}.$	
--	---	--

Il rapporto trovato è il cosiddetto “*inverse Mills ratio*”, ed Heckman lo riassume utilizzando la lettera greca λ (per questo motivo ci si riferisce ad esso anche chiamandolo “*lambda di Heckman*”). Arriviamo così ad un importante risultato:

	$E(y \mathbf{z}, s = 1) = \mathbf{x}\boldsymbol{\beta} + \rho\lambda(\mathbf{z}\boldsymbol{\gamma}).$	(6)
--	---	-----

L'equazione (6) ci mostra che il valore atteso di y , dati \mathbf{z} e l'osservabilità di y , è uguale a $\mathbf{x}\boldsymbol{\beta}$, più un termine addizionale che dipende dall'*inverse Mills ratio* calcolato nel punto $\mathbf{z}\boldsymbol{\gamma}$. Ricordando che il nostro obiettivo è quello di stimare il vettore dei parametri $\boldsymbol{\beta}$, l'equazione in realtà ci dice anche che la stima corretta e consistente diventa possibile se, oltre ad usare soltanto il campione selezionato, includiamo il termine $\lambda(\mathbf{z}\boldsymbol{\gamma})$ come ulteriore regressore del modello.

Arrivati a questo punto, è necessario stimare $\boldsymbol{\gamma}$, perché non conoscendone il valore non possiamo di conseguenza valutare il termine $\lambda(\mathbf{z}_i\boldsymbol{\gamma})$ per ciascuna i . Avendo il processo di selezione delle osservazioni un esito binario, Heckman decide di stimare i parametri attraverso il modello Probit di risposta binaria:

$$P(s = 1|\mathbf{z}) = \Phi(\mathbf{z}\boldsymbol{\gamma}).$$

²⁰ Nota che v si distribuisce come una normale standard secondo l'assunzione (b) descritta nella formulazione dell'equazione (5).

²¹ Se facciamo riferimento alla (4), la condizione $s = 1$ si verifica quando è vera la disuguaglianza $\mathbf{z}\boldsymbol{\gamma} + v \geq 0$, la quale può essere appunto scritta, isolando il termine d'errore, come $v \geq -\mathbf{z}\boldsymbol{\gamma}$.

La stima del Probit avviene tramite il metodo della massima verosimiglianza, nel quale i parametri stimati sono quelli che massimizzano la funzione di log-verosimiglianza, formulata come segue:

$$\ln \mathcal{L}(\boldsymbol{\gamma}) = \sum_{i=1}^n (s_i \ln \Phi(\mathbf{z}'_i \boldsymbol{\gamma}) + (1 - s_i) \ln(1 - \Phi(\mathbf{z}'_i \boldsymbol{\gamma}))).$$

Una volta ottenute queste stime siamo allora in grado di valutare il *Mills ratio* per ciascuna osservazione e quindi di avere un valore noto per tutti i regressori dell'equazione (6). Possiamo pertanto proseguire con la stima della regressione multipla, la quale a questo punto avrà come oggetto non solo i β_j , ma anche il parametro ρ relativo al Mills ratio (o correzione di Heckman).

3.3 Applicazione del modello al caso in esame

Considerata la non rilevanza del corner solution model, e della regressione censurata, la scelta della soluzione al problema degli “zeri” per il nostro dataset è ricaduta sul *modello di selezione di Heckman* appena spiegato.

Nello specifico, avremo quindi che il caso $s = 1$ (che implica l'appartenenza al campione) indicherà l'appartenenza del giocatore al 25-man roster di una società di Major League, o in altre parole l'aver firmato un contratto di Major League. Quando invece $s = 0$, vorrà dire che il giocatore in esame non è nel roster attivo di quell'anno, ma che allo stesso tempo egli ha disputato partite in Major League percependo però un salario di Minor League (che non viene rilevato).

Gli approcci alternativi non sono stati considerati in quanto:

- a) Il corner solution model non può assolutamente rappresentare la nostra scelta ideale, in quanto applicarlo vorrebbe dire che tutti gli 11.885 giocatori che presentano un valore mancante nel salario non ricevano effettivamente alcun salario; ma ovviamente ciò andrebbe assolutamente in contrasto sia con la definizione di giocatore professionista di baseball, sia con la contestualizzazione economica effettuata sul mercato del lavoro della Major League Baseball.
- b) La regressione censurata potrebbe rappresentare la nostra scelta, in quanto i giocatori per i quali si verifica la mancata osservazione del salario percepiscono nella realtà salari “non nulli”, ma nel dataset osserviamo degli “zeri” perché essi non possiedono un contratto di Major League. La cosa che però ci fa scartare la scelta di questa soluzione è l'ipotesi alla base del modello, secondo la quale i giocatori che presentano il valore del salario appartengono alla stessa medesima popolazione di quelli che presentano un valore mancante; ci sembra infatti un'assunzione troppo forte, quella di racchiudere giocatori delle Minors con quelli di Major League.

CAPITOLO 4

PREMESSA ALLE STIME

4.1 Premessa relativa ai dati destinati all'analisi

Riteniamo opportuno fare una premessa riguardo ai dati che useremo per la stima del nostro modello di selezione di Heckman applicato alla Major League Baseball. In primo luogo lavoreremo su un numero minore di osservazioni rispetto a quanto detto durante la descrizione del nostro data-set (vedi paragrafo 2.2), in quanto sono esclusi dalle analisi il ruolo del lanciatore, per concentrarci invece sui giocatori di campo e sulle performance di battuta. In secondo luogo risulta necessario fare una serie di considerazioni riguardo al dato del salario, e spiegheremo come infine abbiamo scelto di utilizzare questa variabile.

4.1.1 Perché i lanciatori sono esclusi dal modello

Per quanto riguarda l'esclusione dei lanciatori, un primo motivo di questa scelta è che i lanciatori nel baseball vengono valutati su statistiche di performance molto diverse da quelle di tutti gli altri giocatori, e possiamo quindi considerarli come appartenenti ad una popolazione differente da quella degli altri ruoli presenti nel data-set.

Inoltre, all'interno della categoria dei lanciatori troviamo tre sotto-ruoli, ognuno con le proprie caratteristiche e che presentano tra loro una notevole differenza per quanto riguarda la durata media della performance. Essi si differenziano tra *partenti*, *rilievi*, e *closer*: i primi (come lascia intendere il nome) sono quelli designati ad iniziare la partita sul monte di lancio e destinati a rimanerci più a lungo possibile, i rilievi invece sono quelli che subentrano al lanciatore partente e pertanto giocano principalmente nella parte centrale della partita, infine il closer è quel tipo di lanciatore a cui ci si affida per la chiusura della partita (da qui il nome). Includendo semplicemente il ruolo del "lanciatore" generico diventerebbe quindi difficile interpretare le sue stime proprio a causa di questa suddivisione, al contrario sarebbe probabilmente più opportuno effettuare un'analisi a loro specifica, ma la nostra attenzione si è concentrata sulle performance di battuta e sui giocatori di campo.

Infine, oltre al fatto di rappresentare una categoria a parte rispetto agli altri ruoli, spesso i dati e le misure tramite i quali essi vengono valutati sono fortemente dipendenti dal rendimento di anche tutto il resto della squadra. Più precisamente, scorrendo le statistiche specifiche dei lanciatori, si leggono misure come il rapporto tra partite vinte e perse dagli stessi lanciatori²² o come il numero di punti subiti dalla squadra durante la loro performance. Diventa pertanto difficile riuscire ad isolare il contributo ed il merito specifico del lanciatore, in quanto le statistiche a cui si fa

²² Un lanciatore viene considerato il *vincente* o il *perdente* della partita, se nel momento in cui esce dal campo (che sia per una sostituzione o per la fine della partita) la squadra stava rispettivamente vincendo o perdendo, ed le sorti della partita non cambiano da quel momento alla fine della stessa.

riferimento si basano spesso sull'esito della partita, che naturalmente dipende dalle performance di tutti i componenti della squadra.

In conclusione, in seguito all'eliminazione di tutte le osservazioni che rappresentavano i lanciatori, il data-set che useremo nell'analisi sarà composto da 18,525 osservazioni.

4.2 Come abbiamo utilizzato la variabile del salario

Il salario dei giocatori di Major League Baseball è la variabile dipendente del nostro modello. Riteniamo pertanto necessario effettuare alcune considerazioni al riguardo, per spiegare le scelte che abbiamo effettuato rispetto alla modalità di rappresentazione del dato del salario.

4.2.1 Rappresentazione grafica del problema dei salari nulli

Iniziamo andando a vedere come si distribuisce il dato grezzo del salario attraverso la costruzione di un istogramma (grafico costruito tramite il software Stata, di cui mostriamo anche il comando):

```
. histogram salary, frequency normal scheme(sj)
(bin=42, start=0, width=785714.29)
```

Figura 4.3: Comando di Stata per la creazione di un istogramma

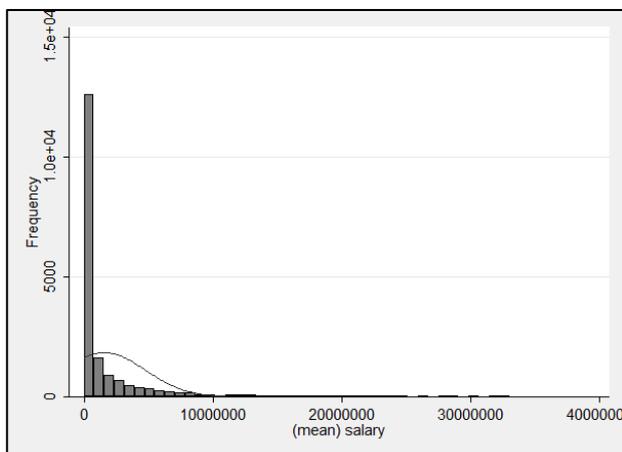


Figura 4.2: Istogramma di frequenza del salario

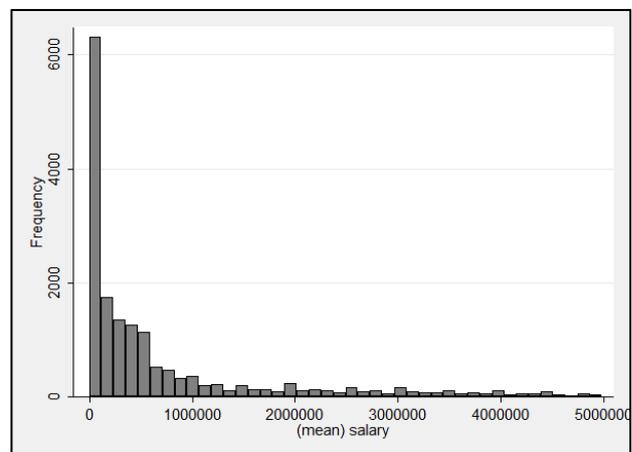


Figura 4.3: Dettaglio del grafico nella 4.2, ristretto ai salari minori di 5 milioni di dollari

Guardando la figura 4.2, la prima cosa che attira l'attenzione è senza dubbio l'altissima frequenza di osservazioni in corrispondenza del valore nullo del salario. Possiamo quindi affermare che la costruzione di questo istogramma ci ha restituito la rappresentazione grafica del problema di mancata osservazione del salario per quei giocatori che non possiedono il contratto di Major League (problema di cui si è discusso nel paragrafo 3.2.2).

Tuttavia la rappresentazione non corrisponde pienamente con quanto descritto in precedenza. In seguito all'eliminazione dei lanciatori dal data-set, il numero di osservazioni che presentavano un salario nullo erano infatti diventate 5271 (dalle quasi dodicimila iniziali), e

pertanto la frequenza della prima colonna dell'istogramma che supera le diecimila osservazioni, sembra rappresentare un dato diverso da quello reale.

In realtà restringendo il campo di costruzione dell'istogramma alle sole osservazioni che presentino un valore del salario minore di 5 milioni di dollari (figura 4.3), si osserva come il primo istogramma accorpasse sia valori nulli che valori vicini al salario minimo e maggiori di zero.²³

4.2.2 Presenza di un salario minimo

La seconda cosa che traspare guardando entrambe le figure (4.2 e 4.3), è la marcata asimmetria positiva presente nella distribuzione dei valori del salario. Questa asimmetria è causata dalla presenza di outliers corrispondenti a valori molto alti del salario (ovvero nella coda destra della distribuzione) e da un troncamento a sinistra.

La spiegazione di questo fenomeno è da ricondurre alla regolamentazione della lega e dalle norme relative al rapporto di lavoro contenute nel Collective Bargaining Agreement (2012), che prevedono la presenza di un salario minimo che le società sono costrette ad elargire ai giocatori che firmano il contratto, ma allo stesso tempo non vi è regolamentazione relativa ad alcun salario massimo. La conseguenza è che i giocatori “peggiori” non possono vedersi abbassare lo stipendio al di sotto di una certa soglia, mentre i “migliori” potrebbero potenzialmente ottenere un salario illimitato.

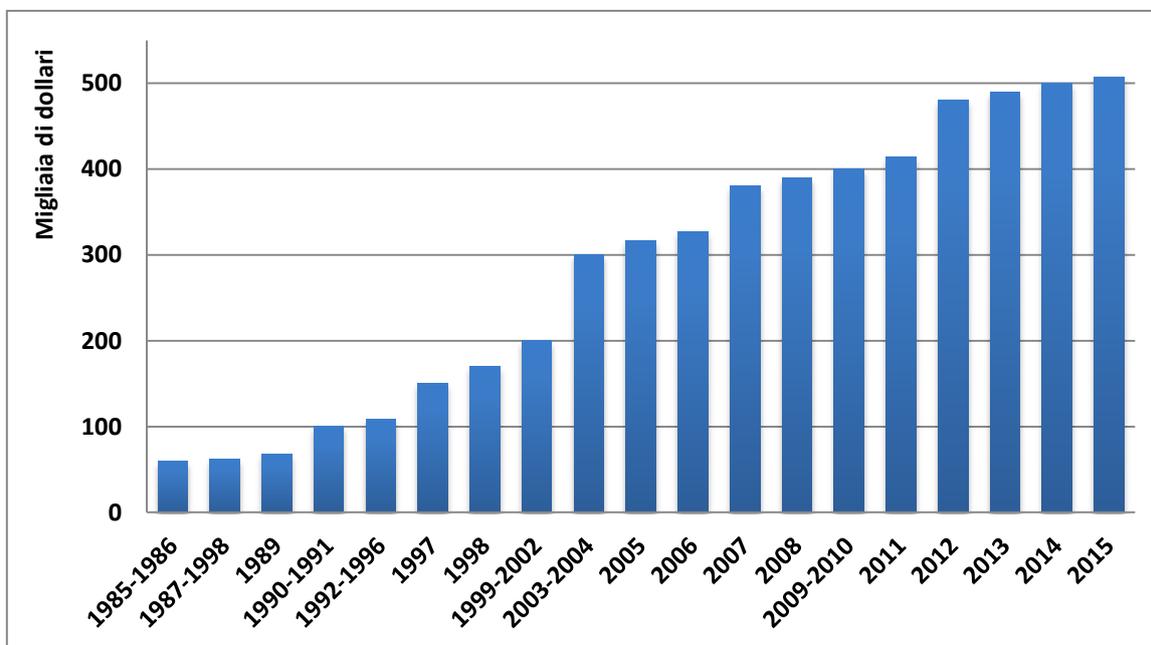


Figura 4.4: Istogramma del minimo salariale

²³ In realtà il fenomeno del raggruppamento è rappresentato, seppur in minima parte, anche nella figura 4.2, mentre per farlo sparire sarebbe necessario restringere ancora il campo ai salari minori di 1 milione di dollari. La scelta di lasciare comunque la versione a 5 milioni, è stata condizionata dal fatto che effettuando l'ulteriore restrizione non era più possibile cogliere, guardando il grafico, l'andamento asimmetrico del salario.

Nella figura 4.4 sono riportate le soglie annuali del salario minimo nominale (cioè senza considerare l'inflazione) imposte dalla lega, relative agli anni dal 1985 al 2015 (gli anni presi in considerazione nel nostro data-set): sull'asse delle ordinate troviamo il valore in migliaia di dollari americani, mentre lungo l'asse delle ascisse troviamo gli anni di validità del relativo minimo salariale. Tramite l'utilizzo dei valori di queste soglie, siamo stati in grado di creare la variabile *sal_min* da inserire all'interno del data-set. L'andamento crescente del minimo salariale nel tempo dipende ovviamente anche dall'adeguamento all'inflazione, e sarà necessario deflazionare le variabili relative a salari e salari minimi per permettere confronti intertemporali..

Creando successivamente una nuova variabile contenente il logaritmo naturale del salario per ogni giocatore, abbiamo quindi costruito nuovamente l'istogramma. Il vantaggio di ricorrere ad una trasformata logaritmica è quello di considerare la variazione tra salari in termini percentuali e non assoluti, di conseguenza riducendo (e stabilizzando) la varianza tra i salari:

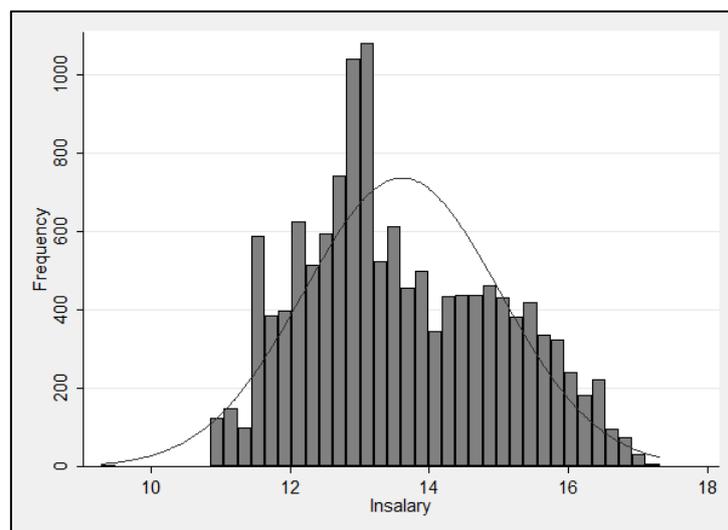


Figura 4.5: Istogramma di frequenza del logaritmo naturale del salario

Osservando il grafico in figura 4.5 notiamo che la distribuzione del logaritmo naturale del salario dei giocatori presenta dei picchi nella parte sinistra della campana, e sembrano essere proprio quei picchi di frequenza a creare un'asimmetria rispetto alla normale. Essendo posizionati nella parte sinistra della distribuzione, ovvero quella che presenta i valori minori della variabile rappresentata (e quindi di conseguenza anche del salario semplice), ci sorge il dubbio che l'alta frequenza registrata nei picchi potrebbe combaciare coi valori dei salari minimi.

Per andare a verificare questa nostra ipotesi facciamo riferimento ad un altro comando del nostro software di riferimento, *summarize*, il quale, applicato ad una delle variabili, ne restituisce 5 statistiche descrittive basilari (ovvero il numero di osservazioni, la media, la deviazione standard, il valore minimo ed il valore massimo):

```
. summarize lnsal_min
```

Variable	Obs	Mean	Std. Dev.	Min	Max
lnsal_min	18525	12.18325	.7100294	11.0021	13.13725

Figura 4.6: Istogramma di frequenza del logaritmo naturale del salario

I nostri sospetti si rivelano presto fondati. Il valore minimo (pari a 11.0021) ed il valore massimo (pari a 13.12725) del logaritmo naturale del salario minimo riportati nella figura 4.5, ci indicano infatti che i salari minimi dei vari anni si concentrano in quel range di valori, il quale, a conferma della nostra ipotesi, è lo stesso identico range di valori della variabile d'interesse che racchiude tutti i picchi di frequenza della distribuzione.

4.2.3 Deflazionamento dei salari nominali

Abbiamo detto più volte in precedenza che il nostro data-set contiene i salari di tutti i giocatori sotto contratto in Major League dal 1985 al 2015, e nell'arco di 30 anni le variazioni nel livello dei prezzi hanno agito in maniera consistente, per cui un'analisi intertemporale richiede un aggiustamento rispetto all'inflazione

La teoria finanziaria di base afferma che il valore dei soldi dipende dal momento in cui essi vengono ricevuti: “un dollaro oggi vale di più di un dollaro domani” (Brealy et al. 2014) Ciò che diventa necessario è effettuare un'operazione di deflazionamento e a questo scopo faremo riferimento al *CPI* (Indice dei Prezzi al Consumo) statunitense.

L'indice dei prezzi al consumo è una misura statistica che si ottiene attraverso la media dei prezzi di beni e servizi contenuti in un paniere teorico, formato sulla base delle abitudini di acquisto del consumatore medio e basato su prezzi rilevati presso i punti vendita. In altre parole è una misura del “costo della vita medio” nel paese considerato. Mettendo in relazione i CPI dei diversi anni si riesce quindi a misurare l'aumento del livello generale dei prezzi, e quindi l'inflazione dei prezzi al consumo.

Dopo aver creato la nostra variabile *CPI*, utilizzando i dati ufficiali sui CPI statunitensi, abbiamo proseguito con il processo di deflazionamento dei salari, con l'obiettivo di attualizzare tutti i valori al livello generale dei prezzi del 2015 (l'anno più recente del data-set) in modo da riuscire ad interpretare meglio le stime che otterremo in seguito:

$$sal_reale_{anno} = \frac{salary_{anno}}{\left(\frac{CPI_{anno}}{CPI_{2015}}\right)}$$

4.2.4 Dal salario effettivo, alla variazione dal minimo salariale

Se facciamo un piccolo passo indietro e torniamo per un attimo sul grafico in figura 4.5, diamo una migliore interpretazione dei picchi di frequenza che compaiono nella distribuzione e proviamo ad applicare una soluzione al problema. Dovendo nel nostro studio confrontare giocatori che hanno militato in Major League in anni ed epoche diverse, e quindi anche i loro rispettivi salari, la presenza di diversi salari minimi potrebbe portare a conclusioni e stime poco realistiche.

Nel dettaglio, un giocatore che per diversi anni non presenta statistiche di performance e prestazioni degne di nota, agli occhi della società, non appare meritevole di un aumento del salario; allo stesso tempo però, come abbiamo detto, la società non può ridurre gli stipendi oltre a specifiche percentuali e anzi, nel caso del salario minimo, non ha proprio modo di applicare riduzioni (stiamo quindi escludendo dal discorso la possibilità di escludere il giocatore dal roster).

La diretta conseguenza è che questo tipo di giocatore si ritrova a percepire per più anni consecutivi il minimo salariale imposto dalla regolamentazione. Il problema sta nel fatto che, a causa dell'aumento del livello generale dei prezzi, la soglia del salario minimo cresce col passare degli anni, e quindi mentre nella realtà dei fatti il giocatore continua per più anni a ricevere il minimo salariale, agli occhi del nostro modello lo stesso giocatore sembrerebbe ricevere invece un aumento di stipendio.

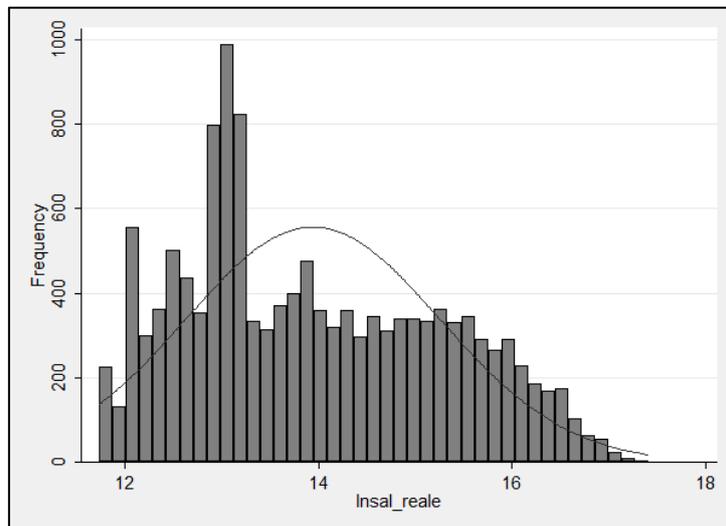


Figura 4.7: Istogramma di frequenza del logaritmo naturale del salario reale

La semplice derivazione del salario reale e quindi il tenere in considerazione il processo d'inflazione degli anni, non sembra sufficiente per risolvere il problema. Come si può ben vedere nella figura 4.7, costruendo nuovamente l'istogramma, utilizzando questa volta il logaritmo naturale del salario reale, i picchi di frequenza causati dai diversi salari minimi sono ancora presenti in maniera altrettanto evidente.

Per questo motivo abbiamo quindi applicato il deflazionamento anche al dato del salario minimo, utilizzando la stessa formula proposta precedentemente. Questo perché, per rimediare alla presenza dei picchi di frequenza nella distribuzione del salario, la soluzione ottimale è trasferire la nostra attenzione dal valore assoluto del salario alla distanza del salario rispetto al salario minimo dello stesso anno.

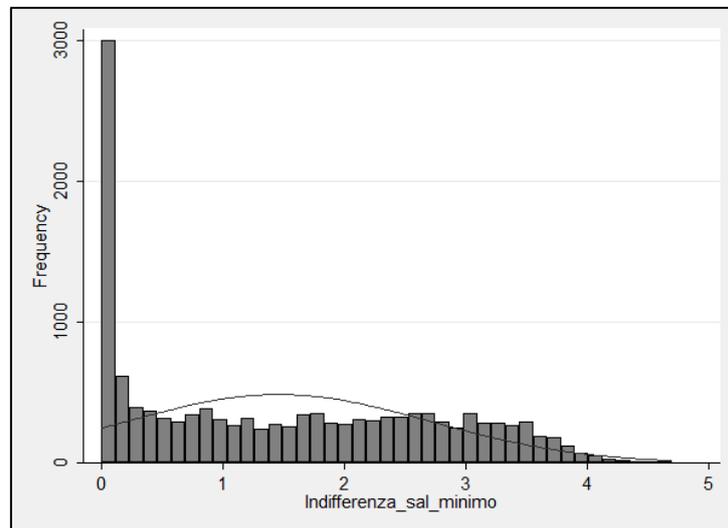


Figura 4.8: Istogramma di frequenza della differenza tra i logaritmi del salario reale e del salario minimo reale

Dopo aver creato le nostre variabili (contententi rispettivamente il salario reale, il salario minimo reale, il logaritmo naturale di entrambi, e la differenza tra i logaritmi), abbiamo infine costruito ancora una volta l'istogramma, per andare a guardare il comportamento della differenza tra i logaritmi del salario reale e del salario minimo reale. Quanto mostrato in figura 4.8 rispecchia a pieno le nostre previsioni.

Infatti rispetto all'istogramma del logaritmo naturale del salario, effettuando la differenza con il logaritmo del salario minimo, risulta molto chiaro guardando il grafico come i diversi picchi presenti prima, siano qui scomparsi: l'ovvia conseguenza è che il picco di frequenza sia, in quest'ultimo istogramma, tutto concentrato sulla prima colonna, in quanto essa racchiude tutti i giocatori che percepiscono il salario minimo in tutti i 31 anni compresi nello studio. Per quanto riguarda il resto dei valori, questi ricordano una distribuzione uniforme.

CAPITOLO 5

STIME E RISULTATI

Il nostro modello consiste due step (uno consecutivo all'altro), un Probit che cattura il processo di selezione delle osservazioni (processo che si sintetizza con la firma di un contratto di Major League), e una regressione lineare multipla del salario dei giocatori su un gruppo di regressori che comprenderà statistiche performative, informazioni del giocatore e della squadra, e ovviamente anche il fattore di correzione (*lambda di Heckman*) derivante dalla stima dell'equazione di selezione.

5.1 La relazione temporale tra salari e prestazioni

Considerando che il processo di selezione delle osservazioni corrisponde nella realtà all'ingresso dei giocatori all'interno del 25-man roster, e che ovviamente la scelta delle società viene fortemente influenzata dalle prestazioni passate dei giocatori, riteniamo che sia quindi opportuno utilizzare le statistiche di performance come regressori, facendo riferimento al loro valore relativo alla stagione precedente. Per permettere l'utilizzo di variabili ritardate relative allo stesso giocatore, è utile fare riferimento ad una struttura panel (anche se la stima del modello non lo rende necessario).

```
. xtset ID yearID
      panel variable:  ID (unbalanced)
      time variable:  yearID, 1985 to 2015, but with gaps
                    delta: 1 unit
```

Figura 5.1: Comando di Stata per impostare i dati in formato panel

Utilizzando il comando *xtset* di Stata, il software ci restituisce l'output mostrato nella figura 5.1. Scrivendo “*xtset ID yearID*” si crea un panel formato dai singoli giocatori contenuti nel data-set (ricordiamo che *ID* è il codice identificativo dei giocatori), che si evolve nel tempo secondo la variabile *yearID*, contenente l'anno di riferimento: ovviamente il panel risulta non bilanciato (la nota “*but with gaps*” si riferisce appunto ai giocatori e non agli anni) in quanto ogni anno si manifesta una notevole rotazione dei giocatori nell'arco dei 30 all'interno della Major League (giocatori che di conseguenza entrano od escono dal nostro panel).

Per analizzare le dinamiche della rotazione all'interno del panel, dobbiamo fare una distinzione tra:

- giocatori *nuovi*, ovvero i giocatori che l'anno precedente non giocavano in Major League, e che quindi non comparivano nel panel;
- giocatori *vecchi*, ovvero che erano già presenti all'interno del panel;
- giocatori *che rimangono* all'interno del panel a fine stagione e quindi che vengono confermati in Major League per l'anno successivo;
- giocatori *che escono* dal panel, o meglio che a fine stagione non vengono confermati.

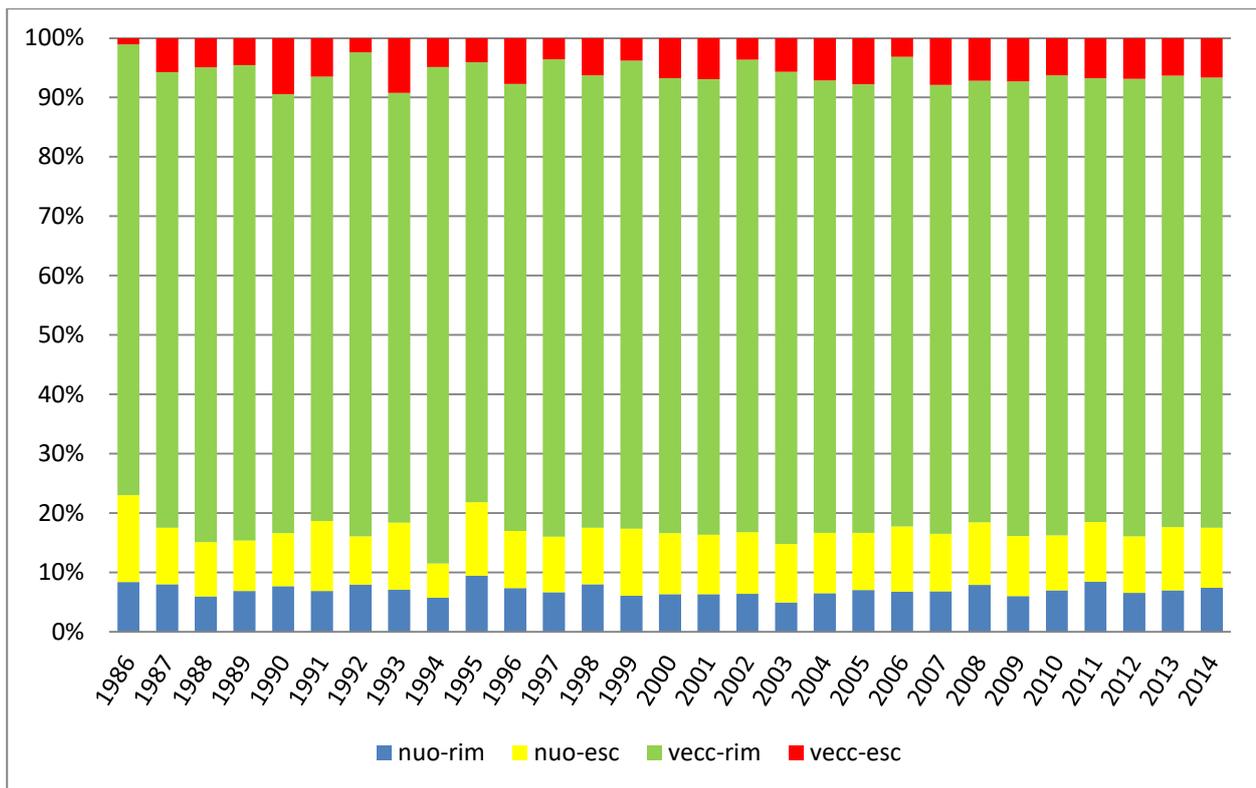


Figura 5.2: Istogramma con colonne in pila relativo alla rotazione degli individui del panel

Il grafico presentato in figura 5.2 è un istogramma *con colonne in pila*. Nel dettaglio, ciascuna colonna rappresenta il totale di giocatori che in quel determinato anno fanno parte del nostro panel, ma questo totale è suddiviso in modo ben evidente in quattro parti (in base alla percentuale sul totale). Nella parte bassa di ciascuna colonna sono rappresentati i giocatori che abbiamo definito *nuovi* (colori blu e giallo), mentre nella parte superiore compaiono i giocatori *vecchi* (colori verde e rosso). Scendendo in ciascuno dei due gruppi così formati, possiamo dire che il colore più in basso rappresenta i giocatori che alla fine della stagione rimarranno all'interno del panel (blu e verde), mentre gli altri sono di conseguenza quelli che non verranno confermati in Major League.

Guardando il grafico è facile notare che la maggioranza dei giocatori all'interno del panel ogni anno è rappresentata da quelli *vecchi – che rimangono*: questo primo dato osservato, sembra quindi confermare quanto detto nel Capitolo 1, riguardo alle regole che permettono ai giocatori che sono già da alcuni anni in Major League di aumentare la probabilità di mantenere il contratto di Major League a fine stagione.

Presentiamo inoltre quello che possiamo definire il processo di rotazione medio del nostro dataset nel corso di un anno (vedi Figura 5.3). Mediamente il panel contiene ogni anno 1,163

giocatori, di cui soltanto il 17% circa²⁴ sono giocatori *nuovi*, e di questi ultimi soltanto il 41% circa²⁵ viene confermato alla fine dell'anno (contro il 59%²⁶, formato da quelli che vengono “bocciati”). Per quanto riguarda invece i giocatori *vecchi*, come era già evidente dal grafico, la stragrande maggioranza viene confermata a fine anno (e probabilmente si aspetta anche già che succeda), mentre l'8%²⁷ uscirà dal panel.

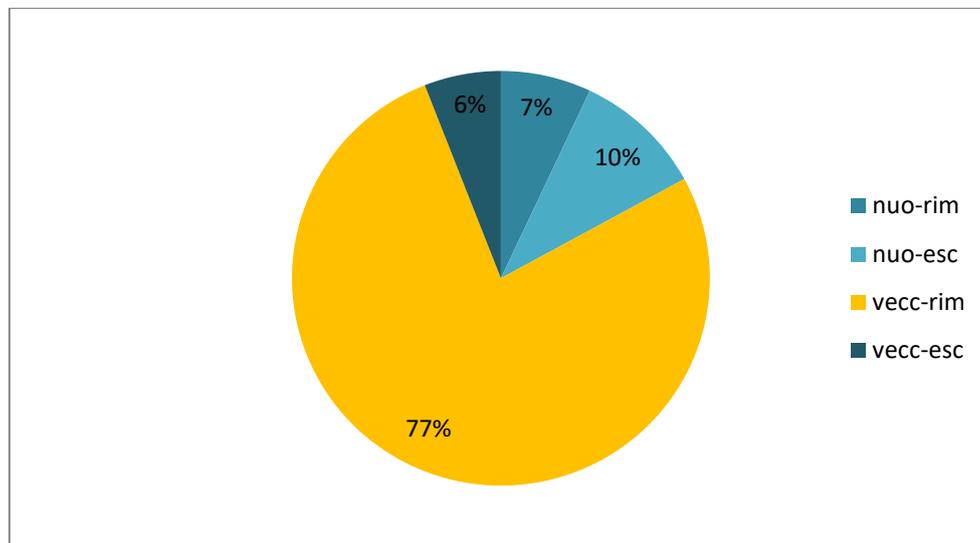


Figura 5.3: Grafico a torta che descrive il processo di rotazione medio del panel

Infine, osservando i colori con cui è stato costruito il grafico a torta in figura, possiamo concludere dicendo che i giocatori *vecchi* – *che rimangono* rappresentano con il loro 77% il nocciolo duro del panel, e di conseguenza il rimanente 23% (formato dall'insieme degli altri 3 sottogruppi) rappresenta la porzione di giocatori che rende il nostro panel non bilanciato.

5.2 Modello di Selezione di Heckman – Step I

Prima di procedere alla stima del modello e all'interpretazione dei coefficienti stimati, è necessaria una descrizione delle variabili dipendenti ed esplicative selezionate per il nostro modello di selezione.

²⁴ Ottenuto sommando i nuovi che rimangono ed i nuovi che escono: $7\% + 10\% = 17\%$.

²⁵ Ottenuto calcolando la percentuale di nuovi che rimangono: $7\% / 17\% = 41,1\%$.

²⁶ Ottenuto semplicemente per esclusione: $100\% - 41\% = 58,9\%$

²⁷ Ottenuto calcolando il rapporto tra i vecchi che escono ed il totale dei vecchi: $7\% / (7\% + 77\%) = 8,3\%$

5.2.1 La variabile dipendente

Avendo ampiamente spiegato che la correzione di Heckman si basa sull'identificazione del processo di selezione delle osservazioni, quindi il possesso di un contratto di Major League da parte dei giocatori, abbiamo creato una variabile binaria per catturare questo processo.

Come dichiarato in precedenza, i dati permettono di individuare i giocatori che detengono un contratto di Major League (e quindi distinguerli da quelli che non ce l'hanno) semplicemente considerando il valore della variabile del salario. Nello specifico, i valori della variabile *salary* diversi da zero indicano l'entità del salario percepito dai giocatori in quella determinata stagione, secondo quanto specificato dal contratto che essi hanno firmato con le varie società di Major League. Al contrario, un valore nullo in questa variabile non riflette l'assenza di stipendio, ma il dato non è registrato nel data-set che non riporta informazioni su contrattualizzazioni con parametri da Minor Leagues.

Pertanto la mancata osservazione della variabile *salary* rappresenta il processo di selezione negativo. Si può quindi creare la nostra variabile dipendente s , che sarà quindi uguale a 0 quando $salary = 0$ e uguale a 1 altrimenti (ovvero quando il giocatore presenta un salario non nullo).

5.2.2 Le variabili indipendenti

Diventa ora rilevante identificare le variabili che guidano il processo di selezione. Abbiamo preso in considerazione diverse categorie di fattori che possano spiegare e rappresentare al meglio le condizioni di contorno del processo di scelta dei giocatori (effettuato dalle società di Major League). Nel dettaglio, verranno considerati sia informazioni sul giocatore, sia informazioni sulla società, sia informazioni riguardo al rapporto tra le due parti.

Per quanto riguarda il giocatore riteniamo doveroso introdurre variabili che ne descrivano le prestazioni e performance, e variabili che invece parlino della sua esperienza all'interno della Major League Baseball. Quello che invece utilizziamo della società di appartenenza sono dati sui risultati ottenuti nella stagione precedente dalla propria squadra di Major League. Infine inseriamo anche due variabili che descrivono il rapporto tra giocatore e società.

5.2.2.1 Variabili che descrivono il giocatore

Iniziamo la descrizione delle variabili inserite nel modello, partendo da quelle che descrivono l'esperienza in Major League del giocatore: *anni* ed *age*.

La prima è una variabile che abbiamo creato in fase di costruzione del data-set che indica, per ciascuna osservazione, il numero di stagioni passate in cui il giocatore ha registrato presenze nella massima serie: decidiamo di inserirla per provare a cogliere il processo precedentemente

spiegato, che in un certo senso “assicura” la permanenza in Major League ai giocatori con più stagioni alle spalle.

La seconda è un semplice dato anagrafico del giocatore, che ne indica appunto l'età: considerando quanto visto nello studio effettuato da Witnauer e Rogers (vedi paragrafo 1.2.3), ci aspettiamo di trovare che la giovane età rappresenti un elemento a favore dei giocatori durante il processo di selezione.

Per quanto riguarda le prestazioni dei giocatori, specifichiamo che i dati che abbiamo utilizzato durante la nostra analisi fanno esclusivamente riferimento alle performance dei giocatori rilevate durante la *regular season*.²⁸ Il motivo alla base di questa scelta risiede nel fatto che ai playoff accedono soltanto 8 squadre su 30, e il meccanismo dei playoff prevede progressive eliminazioni, quindi anche le squadre che conquistano la post-season giocano un numero di partite diverso tra loro. Le variabili di performance che assumiamo influenzare il processo di selezione sono tutte inserite con un ritardo di un anno.²⁹

Abbiamo deciso di utilizzare la statistica *OPS_avg* in quanto racchiude in essa tutte le altre statistiche relative alla battuta (vedi paragrafo 2.2.3.5). Di base la *On-base Plus Slugging* contiene al suo interno sia l'indicatore di potenza della battuta (rappresentato dalla *Slugging Percentage*), che l'indicatore dell'abilità di raggiungere la prima base al termine del turno di battuta (rappresentato dalla *On-Base Percentage*, la quale a questo scopo considera anche le *basi ball* oltre alle battute valide). Inoltre, effettuando la differenza tra questa variabile e il valore della media battuta, riusciamo a considerare anche quest'ultima all'interno della variabile che ne risulta (ovvero proprio *OPS_avg*).

Dalle variabili relative alle apparizioni in battuta (vedi paragrafo 2.2.3.1), abbiamo inserito di conseguenza quelle che rappresentavano gli esiti del turno di battuta non considerati da *OPS_avg*. Nel dettaglio, la nostra scelta è ricaduta sulle seguenti variabili:

- *G* - numero di partite giocate durante la stagione;
- *rbi_g* - rappresenta un indice della produttività del giocatore, in quanto indica quanti punti della squadra sono mediamente prodotti dai turni di battuta del giocatore in una partita;

²⁸ La *regular season* (in italiano stagione regolare) è la fase principale e più lunga del campionato, durante la quale le 30 squadre di Major League si scontrano per un totale di 162 partite annuali (per squadra) ed in base all'esito delle partite si stila una classifica delle squadre. Solitamente questa comincia con il cosiddetto *Opening Day* nella prima domenica di Aprile, e si conclude nel primo weekend di Ottobre.

²⁹ L'utilizzo dei *lagged values* (valori ritardati), ci permette quindi di riprodurre la condizione realistica in cui la società che deve scegliere il giocatore ne valuta le performance della stagione precedente.

- *so_g* – è un indice che spiega in un certo senso il mancato rendimento del battitore, in quanto calcola il numero di volte che questo viene eliminato al piatto (ovvero quando conclude il turno di battuta senza effettivamente produrre alcuna battuta)³⁰;
- *sbp* – rappresenta un indice di abilità come corridori sulle basi, in quanto è precisamente il rapporto tra le *basi rubate*³¹ e tentativi effettuati e rappresenta quindi il tasso di successo.

Quello che ci aspettiamo dalle stime è pertanto un parametro positivo per tutte le variabili sulle performance spiegate, fatta eccezione per *so_g* che invece rappresenta un indice di mancato rendimento.

5.2.2.2 Variabili che descrivono la società

Con le variabili che indicheremo di seguito, proviamo a studiare quanto incidano i risultati ottenuti dalla società l'anno prima, sul suo orientamento a selezionare nuovi giocatori. Includiamo pertanto nel modello le variabili *rank* e *champion*:

- la prima indica la posizione raggiunta in classifica dalla squadra al termine della stagione precedente, all'interno della propria divisione³²;
- la seconda è semplicemente una dummy che è uguale a 1 quando la squadra ha vinto il campionato (vincendo le *World Series*) dell'anno precedente.

Una previsione che facciamo a proposito di queste due variabili è semplicemente che presentino parametri stimati di segno opposto, in quanto la squadra vincitrice del campionato avrà concluso la *regular season* nelle prime due posizioni, mentre il parametro di *rank* ci indicherà come incide sul processo di selezione il passaggio da una posizione in classifica a quella subito sotto. Una volta letti i risultati delle stime saremo quindi in grado di capire se avrà effetto positivo una posizione alta o bassa in classifica.

5.2.2.3 Variabili relative al rapporto giocatore-società

L'ultimo gruppo di variabili indipendenti è quello delle variabili relative al rapporto lavorativo.

³⁰ Non abbiamo di proposito considerato le eliminazioni sulle basi in quanto esse sarebbero fortemente dipendenti dal rendimento difensivo della squadra avversaria.

³¹ Il battitore, una volta raggiunta una base, deve convenzionalmente aspettare la battuta di un compagno per avanzare alla base successiva. Esiste però una giocata, chiamata appunto *rubata*, tramite la quale il giocatore può tentare di conquistare la base successiva senza aspettare la battuta successiva.

³² Ricordiamo che American League e National League sono entrambe suddivise nelle divisioni Ovest Centro ed Est, e la classifica a cui facciamo riferimento è quella interna della divisione (da cui poi si determinano le squadre che accedono ai playoff).

Le variabili che abbiamo deciso di inserire all'interno del processo di selezione dei giocatori sono il valore del salario percepito l'anno precedente, e il valore ritardato di una dummy che indica se il giocatore fa parte del *25-man roster* o meno.

Lo scopo di questa scelta è andare a vedere se e quanto incide l'entità del contratto di un giocatore sulla sua conferma all'inizio della stagione successiva: ipotizziamo per entrambe le variabili una relazione positiva con la probabilità che la variabile dipendente sia uguale a 1.

5.2.3 Le stime del Probit

Una volta elencate e decise le variabili da inserire all'interno del nostro modello Probit siamo quindi giunti alla seguente specificazione:

$$s = \gamma_0 + \gamma_1(l.sal_reale) + \gamma_2(l.s) + \gamma_3(anni) + \gamma_4(age) + \gamma_5(l.OP_S_avg) + \gamma_6(l.rbi_g) + \gamma_7(l.so_g) + \gamma_8(l.sbp) + \gamma_9(l.G) + \gamma_{10}(rank) + \gamma_{11}(champion) + v.$$

Le stime sono riportate di seguito.

```
. probit s l.sal_reale anni age l.OP_S_avg l.rbi_g l.so_g l.sbp l.G rank champion

Iteration 0:   log likelihood = -6356.4269
Iteration 1:   log likelihood = -4677.7882
Iteration 2:   log likelihood = -4536.4758
Iteration 3:   log likelihood = -4520.876
Iteration 4:   log likelihood = -4520.4456
Iteration 5:   log likelihood = -4520.4456

Probit regression               Number of obs   =       14346
                               LR chi2(10)           =       3671.96
                               Prob > chi2            =       0.0000
Log likelihood = -4520.4456     Pseudo R2       =       0.2888
```

Figura 5.4: Output della stima del modello Probit

In figura 5.4 troviamo la prima parte dell'output del software, e vediamo che il modello presenta una log-verosimiglianza pari a -4,520.4456 ed uno *Pseudo R²* circa pari a 0.30. Considerando che questo indice di bontà varia tra i valori 0 e 0.6, il valore ottenuto tramite la stima indica che il nostro modello è riuscito a cogliere quasi il 50% della variabilità della variabile dipendente *s*. Ciò vuol dire che le variabili che abbiamo inserito nel modello riescono a cogliere il 50% delle cause che fanno variare la probabilità di essere o non essere inclusi nel 25-man Roster.

s	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
sal_reale					
L1.	1.34e-07	1.51e-08	8.86	0.000	1.04e-07 1.64e-07
anni	.0367726	.0085172	4.32	0.000	.0200792 .0534661
age	-.0427082	.0080784	-5.29	0.000	-.0585417 -.0268748
OPS_avg					
L1.	.5075277	.1182154	4.29	0.000	.2758298 .7392255
rbi_g					
L1.	1.261673	.1198878	10.52	0.000	1.026698 1.496649
so_g					
L1.	-.3080186	.0637572	-4.83	0.000	-.4329805 -.1830567
sbp					
L1.	.0750175	.0407869	1.84	0.066	-.0049235 .1549584
G					
L1.	.0117864	.0004367	26.99	0.000	.0109304 .0126424
rank	-.0182963	.0094038	-1.95	0.052	-.0367274 .0001349
champion	.2129989	.0960836	2.22	0.027	.0246785 .4013194
_cons	.6625929	.2136016	3.10	0.002	.2439415 1.081244

Figura 5.4: Output della stima del modello Probit

5.2.3.1 Interpretazione dei coefficienti

L'interpretazione dei coefficienti del modello probit non è immediata come ad esempio l'interpretazione dei coefficienti nella regressione lineare: l'aumento della probabilità che si verifichi l'evento $s = 1$ causato dall'aumento unitario di uno dei regressori dipende infatti sia dal proprio valore iniziale, sia dal valore degli altri regressori.

Tuttavia siamo in grado di effettuare un'interpretazione rispetto alla direzione dell'effetto delle variabili: un coefficiente positivo significa che l'aumento del regressore porta ad un aumento della probabilità prevista; un coefficiente negativo rappresenterà al contrario che all'aumentare del regressore corrisponde una diminuzione della probabilità. In poche parole, studiando il segno delle stime riusciamo a capire se l'aumento della variabile associata favorisce oppure ostacola l'ingresso del giocatore nel *25-man roster*.

5.2.3.2 Regressori con coefficienti negativi

Guardando velocemente le stime dei coefficienti ed i loro segni, vediamo che le uniche tre variabili che incidono negativamente sono *age*, *so_g* e *rank*.

Per quanto riguarda la prima, possiamo concludere che all'aumentare dell'età del giocatore si riducono le sue chance di entrare nel *25-man roster* della società di Major League Baseball; questa considerazione si allinea completamente con quanto descritto nel paragrafo 1.2.4.

Come abbiamo detto nella presentazione delle variabili indipendenti del modello Probit (vedi paragrafo 5.2.2.1), la variabile *so_g* era l'unica che rappresentasse un indicatore di prestazione negativa del giocatore, e anche il segno del parametro stimato conferma questa caratteristica.

Infine, non ci eravamo troppo sbilanciati sulla previsione di quello che avrebbe mostrato la stima del parametro della variabile *rank*. Avevamo semplicemente pensato che il suo segno potesse risultare opposto al segno della variabile *champion*, e così effettivamente è stato. In conclusione, il segno negativo di fianco al coefficiente ci indica che all'aumentare della variabile *rank* diminuisce la probabilità. Traducendolo in termini realistici, sembra che siano le squadre meglio classificate nella stagione precedente a garantire una maggior probabilità di promozione in Major League dei suoi giocatori.

5.2.3.3 Regressori con coefficienti positivi

Passando agli altri regressori, possiamo dire in partenza che presenteranno tutti una relazione positiva con la probabilità del giocatore di essere scelto.

Per quanto riguarda le variabili di performance della battuta vale ovviamente il discorso inverso rispetto a quello fatto per la variabile *so_g*, pertanto non ci soffermeremo su di esse e ci limitiamo solo a dire che migliore è il rendimento del giocatore nell'anno passato, più è probabile che gli venga proposto un contratto.

Allo stesso modo possiamo sorvolare l'interpretazione riguardo alla variabile *champion* in quanto ne abbiamo già parlato insieme a *rank*.

Meritano invece particolare attenzione le variabili *anni* e *l.sal_reale*:

- Il segno positivo del coefficiente della prima rappresenta e riassume con un semplice simbolo tutto il discorso trattato nei paragrafi 1.1.3 e 1.1.4: ricordiamo velocemente che un giocatore di Minor League rischia di impiegare fino a 8 anni per raggiungere il contratto di Major (fino a 5 anni di *probationary period*, più 3 anni di opzione). Pertanto sembra logico che all'aumentare degli anni giocati in Major League il giocatore si avvicini sempre più alla firma di un contratto.
- Il coefficiente del salario reale percepito dal giocatore nella stagione precedente, racchiude in sé due concetti precedentemente trattati. Il primo è che in generale, le società non sono propense a retrocedere i giocatori o a rescinderne i contratti

(sicuramente anche a causa della regolamentazione di lega e delle relative restrizioni), e ciò si traduce di conseguenza nella permanenza della stragrande maggioranza dei giocatori all'interno del panel tra un anno e il successivo. Il secondo è un significato un po' più sottinteso: i giocatori che percepiscono i salari più alti vengono quasi certamente confermati a fine stagione, e queste potrebbe far capire che esiste una logica di premio delle performance all'interno della società.

5.4 Modello di Selezione di Heckman – Step II

Dopo aver effettuato la stima del nostro modello Probit, che descrive il processo di selezione dei giocatori di Major League all'interno dei *25-man roster* delle varie società, possiamo utilizzare i risultati di questa stima per trovare il termine di correzione che ci serve per poter infine effettuare la regressione lineare del salario.

5.4.1 Il calcolo del Mill's Ratio

Proseguiamo pertanto l'analisi andando ad applicare al nostro caso ed ai nostri dati quanto detto nel paragrafo 3.2.3, riguardo all'estrazione del *Mill's Ratio*. La teoria dice che questo fattore di correzione è il rapporto tra la funzione di densità e la funzione di distribuzione cumulata della normale. Mostriamo di seguito (in figura 5.5) il comando utilizzato per la creazione delle variabili necessari alla formulazione di questo rapporto:

```
predict xb , xb|
gen double cdf=normal(xb)
gen double pdf=normalden(xb)
gen double IMR=pdf/cdf
```

Figura 5.5: Lista di comandi di Stata per ricavare l'Inverse Mill's Ratio

Tramite il comando *predict xb, xb* stiamo chiedendo al software di calcolare per ciascuna osservazione del nostro dataset il relativo $\mathbf{z}_i\boldsymbol{\gamma}$, ovvero l'argomento su cui valutare le due funzioni al numeratore ed al denominatore del *Mill's Ratio*. Nel dettaglio, Stata procederà nel calcolo della combinazione lineare dei regressori del modello Probit per ciascuna osservazione, utilizzando il valore specifico delle variabili e il le stime dei parametri ($\hat{\boldsymbol{\gamma}}$) appena ottenute.

Una volta creata questa nuova variabile è quindi possibile procedere con il calcolo delle funzioni del rapporto: la prima istruzione [*gen double cdf=normal(xb)*] restituisce il valore della funzione di distribuzione cumulata della normale, calcolata in $\mathbf{z}_i\boldsymbol{\gamma}$; la seconda [*gen double pdf=normalden(xb)*] calcola invece il valore della distribuzione di densità della normale, sempre calcolato nello stesso punto. Così facendo troviamo rispettivamente il denominatore ed il numeratore dell'*Inverse Mill's Ratio* per ciascuna delle nostre variabili.

Infine non resta che effettuare il rapporto, e arrivare quindi al nostro fattore di correzione (*IMR*) tramite l'ultima riga di comando presente in figura [*gen double IMR=pdf/cdf*].

5.4.2 La variabile dipendente

Ricordando le considerazioni effettuate nel paragrafo 4.2 riguardo al valore del salario, la scelta della miglior rappresentazione di questo valore è infine ricaduta sulla differenza tra i logaritmi naturali rispettivamente del salario reale e del salario minimo reale. La variabile dipendente della nostra regressione sarà pertanto *diff_log*: essendo essa la differenza tra due logaritmi, l'impatto dei nostri regressori (descritto dai relativi parametri stimati) sarà da interpretare in termini di variazione percentuale.

5.4.3 I regressori in comune col Probit

Per quanto riguarda le variabili indipendenti, ripetiamo che non vi deve essere perfetta sovrapposizione tra i regressori del salario e le variabili del Probit, pertanto avremo una parte di variabili specifiche della regressione e una parte di variabili in comune con il modello di selezione (escludendo quelle specifiche del Probit).

Dal vettore *z* delle variabili del modello di selezione abbiamo inserito nella regressione tutte le variabili di performance dei giocatori³³, rendendo pertanto le variabili relative alla società "specifiche" del modello di selezione. Inoltre, abbiamo mantenuto solamente la variabile *age* (inserendola poi in relazione quadratica), in quanto abbiamo considerato il valore della variabile *anni* un fattore molto importante per il processo di selezione dei giocatori (che quindi è rimasto una delle variabili specifiche), mentre nella stima della regressione perdeva significatività.

5.4.4 I regressori specifici della regressione

La prima delle variabili specifiche della regressione che abbiamo aggiunto è stata la variabile *team*, variabile che identifica la squadra per cui il giocatore accetta di giocare firmando il suo contratto. Tramite l'inserimento di questa variabile ci aspettiamo di cogliere gli effetti caratteristici delle società, uno su tutti il monte ingaggi stagionale, che è sicuramente quello che incide maggiormente nella determinazione del salario dei giocatori. Ci aspettiamo quindi di trovare che le stime dei parametri riescano ad individuare le società che solitamente registrano i monte ingaggi più alti della Major League.

La seconda è la variabile *ruolo*. Durante tutta l'analisi del nostro studio abbiamo sempre fatto solamente riferimento alle performance di battuta dei giocatori del data-set, per questo motivo

³³ Abbiamo in realtà sostituito la variabile *I.G* indicante la partite giocate dal giocatore durante l'anno prima, con la variabile *I.Gtot*, contenente invece il numero di partite giocate fino a quel momento nella sua carriera in Major League.

ci è sembrato opportuno inserire questa variabile, in modo da capire quali siano i ruoli mediamente più pagati e riuscire quindi a trarre conclusioni su quali siano i ruoli che richiedono livelli maggiori di specializzazione. Si noti che la variabile *ruolo* non era stata inclusa nel modello di selezione, in quanto abbiamo spiegato nell'introduzione ai roster (vedi paragrafo 1.1.4) che questi vengono costruiti allo stesso modo da tutte le società di Major League, pertanto il ruolo del giocatore non avrebbe rappresentato un fattore influenzante il raggiungimento del contratto.

Infine, per far sì che il modello stimato sia il modello di selezione di Heckman, l'ultima variabile specifica della regressione da inserire, è quindi rappresentata dall'*Inverse Mill's Ratio*. L'inserimento all'interno della regressione della cosiddetta "correzione di Heckman" è infatti il fulcro del modello, in quanto all'interno di questa variabile viene riassunto il processo di selezione che avviene riguardo alle osservazioni. Sarà il segno del suo coefficiente a dirci quanto e come incida sul valore atteso del salario, l'esclusione dal modello dei giocatori senza contratto.

5.4.5 Le stime della regressione

L'equazione del Modello di Selezione di Heckman applicato ai nostri dati, risulta, in seguito a tutte le considerazioni effettuate, come segue:

$$\mathbf{diff_log} = \beta_0 + \beta_1(l.\mathbf{diff_log}) + \gamma_2(\mathbf{age}) + \gamma_3(\mathbf{age})^2 + \gamma_4(l.\mathbf{OPS_avg}) + \gamma_5(l.\mathbf{rbi_g}) + \gamma_6(\mathbf{so_g}) + \gamma_7(\mathbf{sbp}) + \gamma_8(l.\mathbf{Gtot}) + \gamma_9(\mathbf{team}) + \gamma_{10}(\mathbf{ruolo}) + \rho(\mathbf{IMR}) + \mathbf{u}.$$

Procediamo allora con la stima sul software Stata, mostrando di seguito gli output.

. reg diff_log l.diff_log age c.age#c.age l.OPS_avg l.rbi_g l.so_g l.sbp l.Gtot i.team i.ruolo IMR			
Source	SS	df	MS
Model	11960.3206	46	260.006969
Residual	2501.45673	10125	.247057455
Total	14461.7773	10171	1.42186386

Number of obs =	10172
F(46, 10125) =	1052.41
Prob > F	= 0.0000
R-squared	= 0.8270
Adj R-squared	= 0.8262
Root MSE	= .49705

Figura 5.6: Output della stima della regressione

Le nostre stime (figura 5.6) hanno prodotto un R^2 corretto pari a 0.8270. Questo vuol dire che il modello da noi stimato è riuscito a cogliere l'82,7% della varianza del salario dei giocatori, quindi le variabili che abbiamo inserito come regressori hanno un impatto deciso sul valore del salario.

diff_log	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
diff_log L1.	.714938	.0066538	107.45	0.000	.7018952 .7279808
age	.2279867	.0165915	13.74	0.000	.1954641 .2605093
c.age#c.age	-.0038811	.0002635	-14.73	0.000	-.0043976 -.0033646
OPS_avg L1.	.5147642	.0803628	6.41	0.000	.3572373 .6722912
rbi_g L1.	.8107872	.0503545	16.10	0.000	.7120824 .9094919
so_g L1.	-.0322655	.0248389	-1.30	0.194	-.0809547 .0164238
sbp L1.	.0828683	.0165735	5.00	0.000	.050381 .1153555
Gtot L1.	.0000405	.0000186	2.18	0.029	4.13e-06 .000077

Figura 5.7: Output della stima della regressione

5.4.5.1 Interpretazione dei coefficienti relativi all'esperienza e alle performance

Iniziamo parlando delle stime del primo blocco di variabili, raffigurato nella figura 5.7.

Il valore ritardato della variabile dipendente presenta un coefficiente pari a circa 0.71. Tale coefficiente rappresenta una misura di “persistenza” del salario e questo tipo di modelli in economia sono detti “ad aggiustamento parziale”. In maniera semplificata, possiamo dire che il salario nella stagione corrente è spiegato per il 71% dal salario della stagione precedente, e per il restante 29% si aggiusta secondo le statistiche di performance.

È curioso quanto succede relativamente alle variabili relative alle statistiche di battuta. Ricordando che il valore dell'*OPS* e della media battuta si esprimono in millesimi³⁴ occorre dividere il valore della stima del parametro di *OPS_avg* per mille. La conseguenza è che la variabile che, secondo la teoria sabermetrica, racchiude tutte le informazioni sulla battuta purtroppo spiega soltanto una piccolissima parte del salario e quindi del valore del giocatore. Quello che invece sembra spiegare fortemente il valore del giocatore sono gli *rbi* (punti segnati grazie al battitore): possiamo allora interpretare questa statistica come indice di produttività del giocatore in termine di punti (e quindi volendo anche di vittorie), e dichiarare che la produttività dei suoi lavoratori rappresenta un fattore molto importante per la società di Major League Baseball.

³⁴ I valori di queste statistiche si scrivono in millesimi (ad esempio .300) ma si leggono come un numero intero (*trecento*). Vedi spiegazione al paragrafo 2.2.3.5.

Il regressore che abbiamo definito come indicatore di scarse performance non presenta un alto livello di significatività, probabilmente a causa del fatto che parte del suo effetto è racchiuso nella statistica *OPS*, in quanto una registra in un certo senso l'opposto dell'altra. In ogni caso l'interpretazione che possiamo dare alla stima si allinea comunque con la logica e con quanto previsto, in quanto il coefficiente presenta il segno negativo.

E' relativamente sorprendente il basso contenuto informativo della variabile che descrive l'abilità di corridore dei giocatori. Infatti solitamente è un requisito della minoranza dei battitori, quello di essere anche dei *rubatori di basi*, e per questo motivo pensavamo potesse essere un valore aggiunto importante. Probabilmente l'impatto dell'abilità nella corsa sulle basi è stato proprio "diluito" dal fatto che soltanto un piccolo sottoinsieme dei giocatori ottiene salari più alti grazie alle basi rubate.

Per quanto riguarda l'età possiamo invece dire che la stima conferma l'assunzione fatta riguardo alla relazione che esiste tra essa ed il salario dei giocatori. Come nella maggior parte degli sport, i giocatori raggiungessero il picco del proprio valore dopo alcuni anni (attorno a metà della propria carriera), nel momento in cui hanno già dalla loro anni di esperienza e allo stesso tempo sono ancora altamente performanti fisicamente: per questo motivo la relazione quadratica nel modello e il coefficiente negativo del termine quadratico.

Infine il numero di partite giocate dal giocatore in Major League negli anni passati, riescono a spiegare soltanto una piccola parte del salario. Un motivo è certamente che l'età del giocatore probabilmente spiega già al suo interno la maggior parte del livello di esperienza del giocatore.

team						
2	-.003708	.0435163	-0.09	0.932	-.0889983	.0815823
3	.0202734	.0376174	0.54	0.590	-.0534554	.0940022
4	.0169412	.0366851	0.46	0.644	-.0549603	.0888427
5	.0446194	.0365772	1.22	0.223	-.0270706	.1163094
9	-.0005657	.0370971	-0.02	0.988	-.0732747	.0721434
10	-.0484541	.0369975	-1.31	0.190	-.1209677	.0240596
11	-.1013606	.04053	-2.50	0.012	-.1807979	-.0219233
12	.0307136	.0372746	0.82	0.410	-.0423432	.1037704
14	.0012342	.0375036	0.03	0.974	-.0722715	.0747399
19	-.0608464	.0371116	-1.64	0.101	-.1335837	.011891
20	-.0339364	.0380662	-0.89	0.373	-.1085448	.040672
25	-.0736635	.0372577	-1.98	0.048	-.1466872	-.0006399
26	.0132204	.0367324	0.36	0.719	-.0587737	.0852145
27	-.012544	.0373273	-0.34	0.737	-.0857042	.0606162
29	.0053828	.0370616	0.15	0.885	-.0672567	.0780223
33	-.0269166	.0382595	-0.70	0.482	-.1019038	.0480706
34	.0437344	.0370134	1.18	0.237	-.0288105	.1162793
36	.0105987	.0377327	0.28	0.779	-.063356	.0845535
37	-.0350171	.0369412	-0.95	0.343	-.1074204	.0373863
38	-.1055779	.0404285	-2.61	0.009	-.1848163	-.0263394
39	-.0011	.0370196	-0.03	0.976	-.073657	.071457
40	.058539	.0363181	1.61	0.107	-.0126431	.1297211
41	.0073069	.0371339	0.20	0.844	-.0654741	.080088
42	-.0196902	.036527	-0.54	0.590	-.0912818	.0519014
43	-.0700716	.0372848	-1.88	0.060	-.1431484	.0030052
44	.003733	.0371641	0.10	0.920	-.0691073	.0765732
45	.0005932	.0369823	0.02	0.987	-.0718908	.0730773
46	-.1035648	.0440954	-2.35	0.019	-.1899901	-.0171395
47	-.0703423	.0380602	-1.85	0.065	-.1449388	.0042543

Figura 5.8: Output della stima della regressione

5.4.5.2 Interpretazione dei coefficienti delle squadre di Major League

In figura 5.8 troviamo i coefficienti delle squadre appartenenti alla Major League Baseball. Come anticipato nella formulazione del modello (paragrafo 5.4.4) potremmo andare a vedere se con l'inserimento della variabile team all'interno della regressione siamo riusciti a cogliere gli effetti del monte ingaggi. Conoscendo la squadra esclusa dal modello secondo la codifica d'angolo, possiamo capire in base ai segni dei coefficienti quali siano le squadre che sono disposte a pagare un salario più o meno elevato.

ruolo						
2	.0650928	.0214052	3.04	0.002	.0231394	.1070462
3	.0245496	.0206865	1.19	0.235	-.0159952	.0650945
4	.0760096	.0200284	3.80	0.000	.0367546	.1152646
5	.1097515	.0219315	5.00	0.000	.0667665	.1527365
6	-.0307757	.0282135	-1.09	0.275	-.0860733	.0245218
7	.0211845	.0201945	1.05	0.294	-.0183959	.0607649
9	.0269857	.0205378	1.31	0.189	-.0132676	.067239
10	.1215896	.0221236	5.50	0.000	.0782281	.1649512
IMR	-.6448853	.0384447	-16.77	0.000	-.7202355	-.5695352
_cons	-3.283535	.2637305	-12.45	0.000	-3.800437	-2.766633
sigma_u	0					
sigma_e	.50036236					
rho	0	(fraction of variance due to u_i)				

Figura 5.9: Output della stima della regressione

5.4.5.3 Interpretazione dei coefficienti relativi ai diversi ruoli

Proseguiamo l'interpretazione delle stime guardando i coefficienti della variabile *ruolo* rappresentati nella figura 5.9. Spieghiamo il software inserisce i valori della variabile *ruolo* tramite la creazione di dummy, che vengono poi inserite attraverso la codifica d'angolo. Nel dettaglio sarà escluso il ruolo del Prima Base dal modello, e di conseguenza i coefficienti degli altri ruoli andranno interpretati in rapporto al Prima Base: un coefficiente positivo indicherà che il ruolo di riferimento è mediamente più pagato del ruolo escluso.

Le stime sono molto sensate, rispetto alle caratteristiche di questo sport.

L'unico ruolo pagato meno del Prima Base risulta essere il *DH*, o battitore designato, ruolo che mediamente viene ricoperto da giocatori che sono bravi in battuta, ma non all'altezza della concorrenza in difesa: spesso il ruolo viene ricoperto infatti da giocatori più "anziani", che in difesa non rendono più come un tempo.

Inoltre le stime riescono ad individuare molto bene la *spina dorsale* della squadra. Con spina dorsale ci riferiamo ai ruoli che uniti raccolgono la maggior parte delle giocate difensive effettuate dalla squadra. Essi sono l'Interbase, l'Esterno Centro, il Ricevitore ed il Seconda Base, e i loro coefficienti sono infatti quelli più alti (rispettivamente 0.122, 0.109, 0.076 e 0.065).

Un valore meno scontato è il coefficiente del Terza Base che risulta alla pari di quello dei due esterni rimanenti, i quali invece tornano ad essere realistici, dicendo che tra Esterno Destro e Sinistro il secondo è il meno pagato. Naturalmente queste valutazioni sono fatte "a parità di performance offensive".

5.4.5.4 Interpretazione del parametro del Mill's Ratio

Abbiamo lasciato in fondo l'interpretazione del parametro dell'Inverse Mill's Ratio. Ricordiamo che l'Inverse Mill's Ratio è il rapporto tra (1) la covarianza tra l'errore del modello Probit e quello della regressione, e (2) la varianza dell'errore del modello probit.

Se il coefficiente risulta negativo, come nel nostro caso, vuol dire che la covarianza è negativa. In altre parole, omettendo i giocatori senza contratto si ottiene una distorsione verso l'alto dei salari: il valore atteso della variabile dipendente risulta sistematicamente troppo alto.

Il risultato ottenuto è plausibile e soddisfacente, in quanto se i giocatori al di fuori del 25-man Roster percepissero un salario di Major League, questo sarebbe logicamente inferiore rispetto a quello dei giocatori che invece ne fanno parte. Di conseguenza la loro inclusione nella stima del modello porterebbe ad un abbassamento sostanziale della media dei salari.

CONCLUSIONI

In questa tesi abbiamo analizzato l'effetto della performance "produttiva" sul salario di lavoratori altamente specializzati, utilizzando come caso studio i giocatori di baseball professionisti delle leghe maggiori statunitensi. Per giungere a stime consistenti, si è utilizzato il modello a due stadi di Heckman per la correzione dell'errore di selezione, in quanto il nostro data-set conteneva l'informazione sui salari solo per un sottoinsieme selezionato di giocatori, quelli inseriti nella liste con contratto da Major League.

La firma di un contratto all'interno di una realtà sportiva professionistica, come il baseball americano, richiede la capacità da parte delle società di riconoscere un determinato valore dello stipendio ai suoi giocatori, in base a quanto essi siano produttivi in campo.

Rispetto alla misura della performance, la sabermetrica, la scienza che analizza empiricamente il gioco del baseball, ha dato vita ad innumerevoli statistiche ed indicatori di performance, le quali, dalla sua nascita (all'inizio degli Anni '70) ad oggi, hanno permesso alla ricerca di effettuare studi sempre più rappresentativi della realtà.

Considerando che la firma di un contratto all'interno di una realtà sportiva professionistica, come il baseball, richiede la capacità da parte delle società di riconoscere un determinato valore dello stipendio ai suoi giocatori in base a quanto essi siano produttivi in campo, la maggior parte degli studi relativi al baseball si occupano proprio di utilizzare le statistiche sabermetriche di performance, per riuscire ad ottenere una misura realistica del valore dei giocatori.

Nel campionato di Major League Baseball, accade però che all'interno della stessa squadra compaiano una maggioranza di giocatori che hanno firmato un contratto di Major League, ed una minoranza di giocatori che, nonostante abbiano un contratto con una squadra di Minor League, hanno la possibilità di effettuare partite nella massima serie. Questo rappresenta un problema per la stima del salario dei giocatori, in quanto i giocatori senza salario di Major League verrebbero completamente esclusi.

Il nostro trattato era quindi partito con l'obiettivo di scoprire se l'esclusione dei giocatori che giocano nella massima serie senza percepire uno stipendio di Major League, potesse essere causa di distorsioni nelle stime.

A questo scopo abbiamo utilizzato il modello di selezione di Heckman. Tale modello è composto da due stadi. Nel primo viene modellizzata la probabilità di ottenere un contratto di Major League, e nel secondo stadio il salario è modellizzato in funzione delle statistiche di performance e di un fattore di correzione (chiamato *Inverse Mill's Ratio*) basato sulle stime di primo stadio. Grazie a questo fattore di correzione, le stime dei coefficienti forniscono un'indicazione consistente

sull'effetto della performance dei salari di tutti i giocatori, inclusi quelli per i quali non è disponibile il dato del salario.

I risultati ottenuti dalla stima del modello di selezione indica – come ci si poteva aspettare – che l'esclusione di questi giocatori si traduce in una sovrastima dei salari. Questo ci permette quindi di concludere che l'inclusione del fattore di correzione all'interno della regressione del salario porta ad una rappresentazione più realistica dell'impatto che le statistiche di performance hanno sul valore del giocatore. Infine, le stime ottenute sono coerenti con le attese e indicano che le performance della stagione precedente alla firma o al rinnovo del contratto hanno un peso importante sul salario negoziato, attorno al 30% del valore complessivo, mentre il 70% rimanente è basato sugli stipendi passati.

APPENDICE

MAJOR LEAGUE UNIFORM PLAYER'S CONTRACT

Parties

Between _____, herein called the Club,

and _____

of _____, herein called the Player.

Recital

The Club is, along with other Major League Clubs, signatory to the Major League Constitution and has subscribed to the Major League Rules.

Agreement

In consideration of the facts above recited and of the promises of each to the other, the parties agree as follows:

Employment

1. The Club hereby employs the Player to render, and the Player agrees to render, skilled services as a baseball player during the year(s) **See Special Covenants** including the Club's training season, the Club's exhibition games, the Club's playing season, the Division Series, the League Championship Series and the World Series (or any other official series in which the Club may participate and in any receipts of which the Player may be entitled to share).

Payment

2. For performance of the Player's services and promises hereunder the Club will pay the Player the sum of **See Special Covenants** in semi-monthly installments after the commencement of the championship season(s) covered by this contract except as the schedule of payments may be modified by a special covenant. Payment shall be made on the day the amount becomes due, regardless of whether the Club is "home" or "abroad." If a monthly rate of payment is stipulated above, it shall begin with the commencement of the championship season (or such subsequent date as the Player's services may commence) and end with the termination of the championship season and shall be payable in semi-monthly installments as above provided.

Nothing herein shall interfere with the right of the Club and the Player by special covenant herein to mutually agree upon a method of payment whereby part of the Player's salary for the above year can be deferred to subsequent years.

If the Player is in the service of the Club for part of the championship season only, he shall receive such proportion of the sum above mentioned, as the number of days of his actual employment in the championship season bears to the number of days in the championship season. Notwithstanding the rate of payment stipulated above, the minimum rate of payment to the Player for each day of service on a Major League Club shall be at the applicable rate set forth in Article VI(B)(1) of the Basic Agreement between the Thirty Major League Clubs and the Major League Baseball Players Association, effective December 20, 2006 ("Basic Agreement"). The minimum rate of payment for Minor League service for all Players (a) signing a second Major League contract (not covering the same season as any such Player's initial Major League contract) or a subsequent Major League contract, or (b) having at least one day of Major League service, shall be at the applicable rate set forth in Article VI(B)(2) of the Basic Agreement. The minimum rate of payment for Minor League service for all Players signing a first Major League contract who are not covered by Article VI(B)(2) of the Basic Agreement shall be at the applicable rate set forth in Article VI(B)(3) of the Basic Agreement.

Payment to the Player at the rate stipulated above shall be continued throughout any period in which a Player is required to attend a regularly scheduled military encampment of the Reserve of the Armed Forces or of the National Guard during the championship season.

Loyalty

3.(a) The Player agrees to perform his services hereunder diligently and faithfully, to keep himself in first-class physical condition and to obey the Club's training rules, and pledges himself to the American public and to the Club to conform to high standards of personal conduct, fair play and good sportsmanship.

Baseball Promotion

3.(b) In addition to his services in connection with the actual playing of baseball, the Player agrees to cooperate with the Club and participate in any and all reasonable promotional activities of the Club and Major League Baseball, which, in the opinion of the Club, will promote the welfare of the Club or professional baseball, and to observe and comply with all reasonable requirements of the Club respecting conduct and service of its team and its players, at all times whether on or off the field.

Pictures and Public Appearances

3.(c) The Player agrees that his picture may be taken for still photographs, motion pictures or television at such times as the Club may designate and agrees that all rights in such pictures shall belong to the Club and may be used by the Club for publicity purposes in any manner it desires. The Player further agrees that during the playing season he will not make public appearances, participate in radio or television programs or permit his picture to be taken or write or sponsor newspaper or magazine articles or sponsor commercial products without the written consent of the Club, which shall not be withheld except in the reasonable interests of the Club or professional baseball.

PLAYER REPRESENTATIONS

Ability

4.(a) The Player represents and agrees that he has exceptional and unique skill and ability as a baseball player; that his services to be rendered hereunder are of a special, unusual and extraordinary character which gives them peculiar value which cannot be reasonably or adequately compensated for in damages at law, and that the Player's breach of this contract will cause the Club great and irreparable injury and damage. The Player agrees that, in addition to other remedies, the Club shall be entitled to injunctive and other equitable relief to prevent a breach of this contract by the Player, including, among others, the right to enjoin the Player from playing baseball for any other person or organization during the term of his contract.

Condition

4.(b) The Player represents that he has no physical or mental defects known to him and unknown to the appropriate representative of the Club which would prevent or impair performance of his services.

Interest in Club

4.(c) The Player represents that he does not, directly or indirectly, own stock or have any financial interest in the ownership or earnings of any Major League Club, except as hereinafter expressly set forth, and covenants that he will not hereafter, while connected with any Major League Club, acquire or hold any such stock or interest except in accordance with Major League Rule 20(e).

Service

5.(a) The Player agrees that, while under contract, and prior to expiration of the Club's right to renew this contract, he will not play baseball otherwise than for the Club, except that the Player may participate in post-season games under the conditions prescribed in the Major League Rules. Major League Rule 18(b) is set forth herein.

Other Sports

5.(b) The Player and the Club recognize and agree that the Player's participation in certain other sports may impair or destroy his ability and skill as a baseball player. Accordingly, the Player agrees that he will not engage in professional boxing or wrestling; and that, except with the written consent of the Club, he will not engage in skiing, auto racing, motorcycle racing, sky diving, or in any game or exhibition of football, soccer, professional league basketball, ice hockey or other sport involving a substantial risk of personal injury.

Assignment

6.(a) The Player agrees that his contract may be assigned by the Club (and reassigned by any assignee Club) to any other Club in accordance with the Major League Rules. The Club and the Player may, without obtaining special approval, agree by special covenant to limit or eliminate the right of the Club to assign this contract.

Medical Information

6.(b) The Player agrees:

(1) that the Club's physician and any other physician or medical professional consulted by the Player pursuant to Regulation 2 of this contract or Article XIII(D) of the Basic Agreement may furnish to the Club all relevant medical information relating to the Player. Except as permitted by Article XIII(G) of the Basic Agreement, which is incorporated herein by reference, the Club is prohibited from re-disclosing any such information without the express written consent of the Player. The Club's physician shall be the custodian of the medical records furnished to a Club pursuant to this Paragraph 6(b). The Club's trainers shall have access to all such records provided to the Club.

(2) that, should the Club contemplate an assignment of this contract to another Club or Clubs, the Club's physician

may furnish to the physicians and officials of such other Club or Clubs all relevant medical information relating to the Player; provided, however, that said physicians and officials are prohibited from re-disclosing any such information without the express written consent of the Player. In addition, within thirty (30) days from the receipt of the Player's medical information, the physicians and officials of the Club which requested the medical information will return any and all documents received to the Player's Club, and will not keep copies of any documents it received or any other records indicating the substance of the medical information transmitted. If the Player's UPC is assigned before the information is returned in accordance with this subparagraph (2), the assignee Club may retain the information. A Player may, at the time that he is no longer under reserve to the Club or on December 1 or every other year, whichever is earlier, request that the Club notify him of the Clubs to which his medical information was provided pursuant to this Paragraph 6(b)(2).

No Salary Reduction

6.(c) The amount stated in paragraph 2 and in special covenants hereof which is payable to the Player for the period stated in paragraph 1 hereof shall not be diminished by any such assignment, except for failure to report as provided in the next subparagraph (d).

Reporting

6.(d) The Player shall report to the assignee Club promptly (as provided in the Regulations) upon receipt of written notice from the Club of the assignment of this contract. If the Player fails to so report, he shall not be entitled to any payment for the period from the date he receives written notice of assignment until he reports to the assignee Club.

Obligations of Assignor and Assignee Clubs

6.(e) Upon and after such assignment, all rights and obligations of the assignor Club hereunder shall become the rights and obligations of the assignee Club; provided, however, that

(1) The assignee Club shall be liable to the Player for payments accruing only from the date of assignment and shall not be liable (but the assignor Club shall remain liable) for payments accrued prior to that date.

(2) If at any time the assignee is a Major League Club, it shall be liable to pay the Player at the full rate stipulated in paragraph 2 hereof for the remainder of the period stated in paragraph 1 hereof and all prior assignors and assignees shall be relieved of liability for any payment for such period.

(3) Unless the assignor and assignee Clubs agree otherwise, if the assignee Club is a Minor League Baseball Club, the assignee Club shall be liable only to pay the Player at the rate usually paid by said assignee Club to other Players of similar skill and ability in its classification and the assignor Club shall be liable to pay the difference for the remainder of the period stated in paragraph 1 hereof between an amount computed at the rate stipulated in paragraph 2 hereof and the amount so payable by the assignee Club.

(4) If performance and/or award bonuses are included as Special Covenants hereunder and an assignment is made during the championship season, the responsibility for such bonuses shall be as follows:

(i) All performance and/or award bonuses earned prior to the assignment shall be the responsibility of the assignor Club;

(ii) The responsibility for any and all performance bonuses earned after the assignment shall be prorated between the assignor and assignee Clubs in proportion to the total number of relevant events attained during the season with each Club involved; and
(iii) The responsibility for any and all award bonuses earned after the assignment shall be the full and exclusive responsibility of the Club for whom the Player was performing services at the end of the championship season. For purposes of this paragraph, an award bonus for election or selection to the All-Star Game shall be deemed to be earned on the day of the announcement of the election or selection, an award bonus for performance over the championship season shall be deemed earned on the last day of the championship season and an award bonus for performance in the post-season shall be deemed earned on the day of the announcement of the award.

Moving Allowances

6.(f) The Player shall be entitled to moving allowances under the circumstances and in the amounts set forth in Articles VII(F) and VIII of the Basic Agreement.

"Club"

6.(g) All references in other paragraphs of this contract to "the Club" shall be deemed to mean and include any assignee of this contract.

TERMINATION

By Player

7.(a) The Player may terminate this contract, upon written notice to the Club, if the Club shall default in the payments to the Player provided for in paragraph 2 hereof or shall fail to perform any other obligation agreed to be performed by the Club hereunder and if the Club shall fail to remedy such default within ten (10) days after the receipt by the Club of written notice of such default. The Player may also terminate this contract as provided in subparagraph (d)(4) of this paragraph 7. (See Article XV(I) of the Basic Agreement.)

By Club

7.(b) The Club may terminate this contract upon written notice to the Player (but only after requesting and obtaining waivers of this contract from all other Major League Clubs) if the Player shall at any time:

- (1) fail, refuse or neglect to conform his personal conduct to the standards of good citizenship and good sportsmanship or to keep himself in first-class physical condition or to obey the Club's training rules; or
- (2) fail, in the opinion of the Club's management, to exhibit sufficient skill or competitive ability to qualify or continue as a member of the Club's team; or
- (3) fail, refuse or neglect to render his services hereunder or in any other manner materially breach this contract.

7.(c) If this contract is terminated by the Club, the Player shall be entitled to termination pay under the circumstances and in the amounts set forth in Article IX of the Basic Agreement. In addition, the Player shall be entitled to receive an amount equal to the reasonable traveling expenses of the Player, including first-class jet air fare and meals en route, to his home city.

Procedure

7.(d) If the Club proposes to terminate this contract in accordance with subparagraph (b) of this paragraph 7, the procedure shall be as follows:

(1) The Club shall request waivers from all other Major League Clubs. Such waivers shall be good for two (2) business days only. Such waiver request must state that it is for the purpose of terminating this contract and it may not be withdrawn.

(2) Upon receipt of waiver request, any other Major League Club may claim assignment of this contract at a waiver price of \$1.00, the priority of claims to be determined in accordance with the Major League Rules.

(3) If this contract is so claimed, the Club shall, promptly and before any assignment, notify the Player that it had requested waivers for the purpose of terminating this contract and that the contract had been claimed.

(4) Within five (5) days after receipt of notice of such claim, the Player shall be entitled, by written notice to the Club, to terminate this contract on the date of his notice of termination. If the Player fails to so notify the Club, this contract shall be assigned to the claiming Club.

(5) If the contract is not claimed, the Club shall promptly deliver written notice of termination to the Player at the expiration of the waiver period.

7.(e) Upon any termination of this contract by the Player, all obligations of both Parties hereunder shall cease on the date of termination, except the obligation of the Club to pay the Player's compensation to said date.

Regulations

8. The Player accepts as part of this contract the Regulations set forth herein.

Rules

9.(a) The Club and the Player agree to accept, abide by and comply with all provisions of the Major League Constitution, and the Major League Rules, or other rules or regulations in effect on the date of this Uniform Player's Contract, which are not inconsistent with the provisions of this contract or the provisions of any agreement between the Major League Clubs and the Major League Baseball Players Association, provided that the Club, together with the other Major League Clubs and Minor League Baseball, reserves the right to modify, supplement or repeal any provision of said Constitution, Major League Rules or other rules and regulations in a manner not inconsistent with this contract or the provisions of any then existing agreement between the Major League Clubs and the Major League Baseball Players Association.

Disputes

9.(b) All disputes between the Player and the Club which are covered by the Grievance Procedure as set forth in the Basic Agreement shall be resolved in accordance with such Grievance Procedure.

Publication

9.(c) The Club, the Vice President, On-Field Operations and the Commissioner, or any of them, may make public the findings, decision and record of any inquiry, investigation or hearing held or conducted, including in such record all evidence or information given, received, or obtained in connection therewith.

Renewal

10.(a) Unless the Player has exercised his right to become a free agent as set forth in the Basic Agreement, the Club may retain reservation rights over the Player by instructing the Office of the Commissioner to tender to the Player a contract for the term of the next year by including the Player on the Central Tender Letter that the Office of the Commissioner submits to the Players Association on or before December 12 (or if a Sunday, then on or before December 11) in the year of the last playing season covered by this contract. (See Article XX(A) of and Attachments 9 and 12 to the Basic Agreement.) If prior to the March 1 next succeeding said December 12, the Player and the Club have not agreed upon the terms of such contract, then on or before ten (10) days after said March 1, the Club shall have the right by written notice to the Player at his address

following his signature hereto, or if none be given, then at his last address of record with the Club, to renew this contract for the period of one year on the same terms, except that the amount payable to the Player shall be such as the Club shall fix in said notice; provided, however, that said amount, if fixed by a Major League Club, shall be in an amount payable at a rate not less than as specified in Article VI, Section D, of the Basic Agreement. Subject to the Player's rights as set forth in the Basic Agreement, the Club may renew this contract from year to year.

10.(b) The Club's right to renew this contract, as provided in subparagraph (a) of this paragraph 10, and the promise of the Player not to play otherwise than with the Club have been taken into consideration in determining the amount payable under paragraph 2 hereof.

Governmental Regulation-National Emergency

11. This contract is subject to federal or state legislation, regulations, executive or other official orders or other governmental action, now or hereafter in effect respecting military, naval, air or other governmental service, which may directly or indirectly affect the Player, Club or the League and subject also to the right of the Commissioner to suspend the operation of this contract during any national emergency during which Major League Baseball is not played.

Commissioner

12. The term "Commissioner" wherever used in this contract shall be deemed to mean the Commissioner designated under the Major League Constitution, or in the case of a vacancy in the office of Commissioner, the Executive Council or such other body or person or persons as shall be designated in the Major League Constitution to exercise the powers and duties of the Commissioner during such vacancy.

Supplemental Agreements

The Club and the Player covenant that this contract, the Basic Agreement, the Agreement Re Major League Baseball Players Benefit Plan effective April 1, 2007 and Major League Baseball's Joint Drug Prevention and Treatment Program and applicable supplements thereto fully set forth all understandings and agreements between them, and agree that no other understandings or agreements, whether heretofore or hereafter made, shall be valid, recognizable, or of any effect whatsoever, unless expressly set forth in a new or supplemental contract executed by the Player and the Club (acting by its President or such other officer as shall have been thereunto duly authorized by the President or Board of Directors as evidenced by a certificate filed of record with the Commissioner) and complying with the Major League Rules.

Special Covenants

I. Term

This Contract shall cover the following Championship Seasons:
2011

II. Compensation and Payment Schedule

Minor League Salary

In accordance with Paragraph 2 of the Uniform Player's Contract and subject to the Minimum Salary Requirements of the Basic Agreement, for each day of service while in the Minor Leagues, Player shall be paid at the following Annual rate(s):

Year	AAA	AA	A	Full Season A	Short A	Rookie Advanced	Rookie
2011	\$67,300	\$67,300	\$67,300	\$67,300	\$67,300	\$67,300	\$67,300

Major League Salary

In accordance with Paragraph 2 of the Uniform Player's Contract, for each day of service while in the Major Leagues, Club shall pay the Player a salary (less applicable withholding taxes and other normal payroll deductions and any other amounts required by law to be withheld), at a rate and for the season(s) indicated below in the following manner:

Year	Salary
2011	\$414,000

III. Additional Special Covenants

Payroll Deduction

The Club and Player agree as follows:

The Player authorizes the Club to make deductions from his salary to reimburse the Club for:

- (a) any hotel charge over and above the room charge that the Player neglects to pay when checking out of the hotel (Player is required to pay all personal items when checking out of hotel and will be assessed a \$100.00 administrative fee per hotel checkout for unpaid charges);
- (b) any monies advanced to the Player by the Club;
- (c) any other items, not specified herein which might be paid for by the Club on behalf of the Player.

Initial:

Player Club

REGULATIONS

1. The Club's playing season for each year covered by this contract and all renewals hereof shall be as fixed by the Office of the Commissioner.

2. The Player, when requested by the Club, must submit to a complete physical examination at the expense of the Club, and if necessary to treatment by a regular physician or dentist in good standing. Upon refusal of the Player to submit to a complete medical or dental examination, the Club may consider such refusal a violation of this regulation and may take such action as it deems advisable under Regulation 5 of this contract. Disability directly resulting from injury sustained in the course and within the scope of his employment under this contract shall not impair the right of the Player to receive his full salary for the period of such disability or for the season in which the injury was sustained (whichever period is shorter), together with the reasonable medical and hospital expenses incurred by reason of the injury and during the term of this contract or for a period of up to two years from the date of initial treatment for such injury, whichever period is longer, but only upon the express prerequisite conditions that (a) written notice of such injury, including the time, place, cause and nature of the injury, is served upon and received by the Club within twenty days of the sustaining of said injury and (b) the Club shall have the right to designate the doctors and hospitals furnishing such medical and hospital services. Failure to give such notice shall not impair the rights of the Player, as herein set forth, if the Club has actual knowledge of such injury. All workmen's compensation payments received by the Player as compensation for loss of income for a specific period during which the Club is paying him in full, shall be paid over by the Player to the Club. Any other disability may be ground for suspending or terminating this contract.

3. The Club will furnish the Player with two complete uniforms, exclusive of shoes, unless the Club requires the Player to wear nonstandard shoes in which case the Club will furnish the shoes. The uniforms will be surrendered by the Player to the Club at the end of the season or upon termination of this contract.

4. The Player shall be entitled to expense allowances under the circumstances and in the amounts set forth in Article VII of the Basic Agreement.

5. For violation by the Player of any regulation or other provision of this contract, the Club may impose a reasonable fine and deduct the amount thereof from the Player's salary or may suspend the Player without salary for a period not exceeding thirty days or both. Written notice of the fine or suspension or both and the reason therefor shall in every case be given to the Player and the Players Association. (See Article XII of the Basic Agreement.)

6. In order to enable the Player to fit himself for his duties under this contract, the Club may require the Player to report for practice at such places as the Club may designate and to participate in such exhibition contests as may be arranged by the Club, without any

other compensation than that herein elsewhere provided, for a period beginning not earlier than thirty-three (33) days prior to the start of the championship season; provided, however, that the Club may invite players to report at an earlier date on a voluntary basis in accordance with Article XIV of the Basic Agreement. The Club will pay the necessary traveling expenses, including the first-class jet air fare and meals en route of the Player from his home city to the training place of the Club, whether he be ordered to go there directly or by way of the home city of the Club. In the event of the failure of the Player to report for practice or to participate in the exhibition games, as required and provided for, he shall be required to get into playing condition to the satisfaction of the Club's team manager, and at the Player's own expense, before his salary shall commence.

7. In case of assignment of this contract, the Player shall report promptly to the assignee Club within 72 hours from the date he receives written notice from the Club of such assignment, if the Player is then not more than 1,600 miles by most direct available railroad route from the assignee Club, plus an additional 24 hours for each additional 800 miles.

Post-Season Exhibition Games. Major League Rule 18(b) provides:

(b) EXHIBITION GAMES. No player shall participate in any exhibition game during the period between the close of the Major League championship season and the following training season, except that, with the consent of the player's Club and permission of the Commissioner, a player may participate in exhibition games for a period of not less than 30 days, such period to be designated annually by the Commissioner. Players who participate in barnstorming during this period cannot engage in any Winter League activities.

Player conduct, on and off the field, in connection with such postseason exhibition games shall be subject to the discipline of the Commissioner. The Commissioner shall not approve of more than three players of any one Club on the same team. The Commissioner shall not approve of more than three players from the joint membership of the World Series participants playing in the same game.

No player shall participate in any exhibition game with or against any team which, during the current season or within one year, has had any ineligible player or which is or has been during the current season or within one year, managed and controlled by an ineligible player or by any person who has listed an ineligible player under an assumed name or who otherwise has violated, or attempted to violate, any exhibition game contract; or with or against any team which, during said season or within one year, has played against teams containing such ineligible players, or so managed or controlled. Any player who participates in such a game in violation of this Rule 18 shall be fined not less than \$50 nor more than \$500, except that in no event shall such fine be less than the consideration received by such player for participating in such game.

PRINTED IN U.S.A.

REVISED AS OF DECEMBER 2006

LIFE INSURANCE NOTICE AND CONSENT FORM

1. Your Club intends to insure your life under a life insurance policy (the "Policy"). The purpose of the Policy is to offset amounts that the Club may pay under your Uniform Player's Contract ("Player Contract") in the event of your death.
 2. The maximum amount of life insurance coverage for which you may be insured at the time the Policy is issued will be equal to the sum of:
 - The Major League salary provided under your Player Contract, up to one million dollars (\$1,000,000) *and* (if applicable)
 - Seventy-five percent (75%) of the amount by which the Major League salary provided under your Player Contract exceeds one million dollars (\$1,000,000.00).
- In no event will the amount for which your life is insured under the Policy ever exceed a maximum limit of thirty million dollars (\$30,000,000.00). The amount of coverage that your Club purchases under the Policy may be reduced under certain circumstances in order to reflect (if applicable) other insurance coverage on your life. The amount of insurance coverage that your Club purchases under the Policy will decrease over time as the amount owed under your Player Contract is paid to you.
3. Your Club may purchase insurance coverage on you under the Policy before you and your Club sign your Player Contract. Your Club will do so only when and if there is an agreement in principle with you as to the terms of your Player Contract and such terms have been reported to, and confirmed by, the Office of the Commissioner of Baseball and the Major League Baseball Players Association.
 4. The Policy and this consent form will be valid for the length of your Player Contract. Each time you enter into a new or revised Player Contract, a new Policy will be purchased and you will be asked to sign a new consent form.
 5. Your Club will be the sole beneficiary of any life insurance proceeds payable under the Policy in the event of your death.

Consent of Employee for Life Insurance Coverage

By signing below, I agree to, consent to, and understand the following:

- A. I may be insured under the Policy up to a maximum face amount equal to the sum of:
 - The Major League salary provided under my Player Contract, up to one million dollars (\$1,000,000), *and* (if applicable)
 - Seventy-five percent (75%) of the amount by which the Major League salary provided under my Player Contract exceeds one million dollars (\$1,000,000.00).
- B. The amount for which my life is insured under the Policy will never exceed a maximum limit of thirty million dollars (\$30,000,000.00). The amount of coverage may be reduced under certain circumstances in order to reflect (if applicable) other insurance coverage on my life. The amount of insurance coverage will decrease over time as the amount owed under my Player Contract is paid to me.
- C. My Club (or a Trust established by my Club and other Major League Baseball clubs) will be the owner of the Policy. My Club will be the beneficiary of the Policy.
- D. Neither my heirs nor I will receive any rights or benefits, including the payment of a death benefit, under the Policy. The death benefit under the Policy will be payable to my Club. This consent has no effect on any other life insurance policies I hold or that any other person holds on my life.

Proposed Insured (please complete)

Name (First, Middle Initial, Last): _____

Date of Birth: _____

(Home address: street/city/state/zip) _____

Print Name of Insured

Date

BIBLIOGRAFIA

- Bollinger C.R., Hotchkiss J.L. (2003) *The Upside Potential of Hiring Risky Workers: Evidence from the Baseball Industry*. Journal of Labor Economics 21, no. 4: 923-944.
- Brealy R., Myers S., Allen F., Sandri S. (2014) *Principi di finanza aziendale*. McGraw Hill Libri Italia
- Collective Bargaining Agreement (2012)
- Cox M. (2000) *The development of computer-assisted reporting*. Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium, 17-18 Marzo 2000, University of North Carolina, Chapel Hill
- Djaouti D., Alvarez J., Jessel J.P. (2011) *Classifying Serious Games: the G/P/S model*
- Dorsey, N. (2015) *An Exploration into the Relationship of MLB Player Salary and Performance*.
- Friendly M. (2008) *Milestones in the history of thematic cartography, statistical graphics, and data visualization*.
- Heckman, J. J. (1977) *Sample selection bias as a specification error (with an application to the estimation of labor supply functions)*.
- Holovaty A. (2006) *A fundamental way newspaper sites need to change*
- James B (2009) *The 96 Families of Hitters*. The Bill James Gold Mine, p.24 ss..
- Lazear, Edward P. (1998) *Hiring Risky Workers*. In *Internal Labour Market, Incentives, and Employment*, edited by Isao Ohashi and Toshiaki Tachibanaki. New York: St. Martin's
- Lewis, M. (2003) *Moneyball: The Art of Winning an Unfair Game*. W.W. Norton & Company Inc
- Link, C. R., & Yosifov, M. (2012) *Contract Length and Salaries Compensating Wage Differentials in Major League Baseball*. Journal of Sports Economics
- Lorenz M. (2010) *Data driven journalism: What is there to learn?* Edited conference documentation, Amsterdam, Olanda
- Palmer, M. C., & King, R. H. (2006). *Has Salary Discrimination Really Disappeared From Major League Baseball?* Eastern Economic Journal
- Roberts S. (2007) *Just How Long Does the Average Baseball Career Last?* New York Times
- Sandler S. (2012) *NFL, MLB, NHL, MLS & NBA: Which Leagues and Players Make the Most Money?*
- Schwarz A. (2004) *The Numbers Game*. pp. 165, 233.
- Smiciklas M. (2012) *The Power of Infographics: Using Pictures to Communicate and Connect with Your Audience*.
- Thibodeaux T. (2011) *5 tips for getting started in data journalism*. Poynter.org
- Thorn J., Palmer P. (1984) *The Hidden Game of Baseball*. pp. 69-70
- Wasserman, T. (2013) *Determinants of Major League Baseball Player Salaries*.
- Werner, S., & Mero, N. P. (1999). *Fair or foul?: The effects of external, internal, and employee equity on changes in performance of major league baseball players*. Human Relations
- Wooldridge, J. M. (2015) *Introductory econometrics: A modern approach*. Nelson Education

SITOGRAFIA

- About Sean (2017). Disponibile su: <http://www.seanlahman.com/about/>
- Baseball Reference (2017a) *Single-Season Leaders & Records for Batting Average*. Baseball-reference.com.
Disponibile su: https://www.baseball-reference.com/leaders/batting_avg_season.shtml
- Baseball Reference (2017b) *Single-Season Leaders & Records for On-Base%*. Baseball-reference.com.
Disponibile su: https://www.baseball-reference.com/leaders/onbase_perc_season.shtml
- Baseball Reference (2017c) *Single-Season Leaders & Records for Slugging %*. Baseball-reference.com.
Disponibile su: https://www.baseball-reference.com/leaders/slugging_perc_season.shtml
- Bleacher Report. *NFL, MLB, NHL, MLS & NBA: Which Leagues and Players Make the Most Money?* Disponibile su: <http://bleacherreport.com/articles/1109952>
- Gaines C (2011) *Take A Close Look At An Actual Major League Baseball Contract*. BusinessInsider
Disponibile su: <http://www.businessinsider.com/uniform-major-league-baseball-contract-2011-4?IR=T>
- Glossary MLB(2017a) *25-man Roster*. Disponibile su: <http://m.mlb.com/glossary/transactions/25-man-roster>
- Glossary MLB(2017b) *40-man Roster*. Disponibile su: <http://m.mlb.com/glossary/transactions/40-man-roster>
- Heckmanequation.org (2017) *About Professor Heckman*. Disponibile su: heckmanequation.org/about-professor-heckman
- McMahon T. (2017) *Historical Consumer Price Index (CPI-U) Data*. Disponibile su:
https://inflationdata.com/Inflation/Consumer_Price_Index/HistoricalCPI.aspx?reloaded=true
- Read Me (2016). Disponibile su: <http://seanlahman.com/files/database/readme2016.txt>
- Seanlahman.com (2017) *Download Lahman's Baseball Database*. Disponibile su:
<http://www.seanlahman.com/baseball-archive/statistics/>
- Uchicago.edu (2017a) *Profilo del Professor Heckman all'interno del Department of Economics della University of Chicago*. Disponibile su: economics.uchicago.edu/directory/james-j-heckman
- Uchicago.edu (2017b) *Profilo del Professor Heckman all'interno della Law School della University of Chicago*. Disponibile su: law.uchicago.edu/faculty/heckman
- Uchicago.edu (2017c) *Professor James Heckman | Center for the Economics of Human Development | The University of Chicago*. Disponibile su: cehd.uchicago.edu/?page_id=71

