



Dipartimento di Impresa e Management

Cattedra Statistica

Analisi Statistica sul fatturato dei migliori Club
calcistici d'Europa

RELATORE

PROF.SSA Livia De Giovanni

CANDIDATO
Benedetto Bacchetta
MATR. 189951

ANNO ACCADEMICO 2016-2017

*“Il calcio è l’arte di comprimere la storia
universale in 90 minuti”*

George Bernard Shaw

Sommario

Introduzione	4
1. Crescita economica del calcio europeo	6
1.1. Flusso dei proventi.....	6
1.2. L'importanza della UEFA Champions League e UEFA Europa League.....	7
1.3. Brand Value	8
1.4. La crescita del fatturato nelle squadre europee.....	11
1.4.1. Matchday, Broadcast, Commercial	11
1.4.2. The Big Four	12
1.4.3. Primi vs. ultimi	15
2. Teoria Statistica	17
2.1. La regressione lineare multipla.....	17
2.2. Il modello di regressione multipla	17
2.3. Stima dei parametri del modello con il metodo dei minimi quadrati.....	18
2.4. L'indice di determinazione R^2	19
2.5. Proprietà degli stimatori dei minimi quadrati ordinari.....	20
2.6. Stima dei parametri con il metodo della massima verosimiglianza	22
2.7. Inferenza sul modello di regressione multipla	22
2.8. Inferenza sui singoli coefficienti di regressione	23
2.9. Verifica della significatività dell'intero modello (test statistico F)	23
2.10. I Dati Panel.....	24
2.11. Modello ad effetti fissi	26
3. Outcome empirico.....	28
3.1. Raccolta dati.....	28
3.2. Le variabili	29
3.3. Statistica descrittiva.....	30
3.4. Serie storiche delle variabili.....	31
3.5. Diagrammi di dispersione	33
.....	34
.....	34
3.6. Risultati	35
Commenti e Conclusione	37
Bibliografia.....	39
Sitografia	41

Introduzione

Ogni giorno, 365 giorni l'anno, si parla di calcio: durante il lavoro, a scuola, tra amici, nei giornali e notiziari nazionali, il calcio è per molte persone una filosofia di vita; indipendentemente da età, sesso o classe sociale.

Il calcio è ormai diventato un fenomeno mondiale e un fondamento importante della nostra società. Milioni di persone ne sono interessate ed ogni anno il numero di affezionati cresce sempre di più.

Nel 2006 la FIFA (*Fédération Internationale de Football Association*) ha dichiarato che nel mondo si contavano 270 Milioni di persone attive nel calcio, pari al 4,13% della popolazione mondiale.¹

La diffusione e la popolarità del calcio sono attribuibili, fondamentalmente, a tre fattori:

1. è uno sport facile e divertente, accessibile a tutti e si può praticare ovunque;
2. può creare idoli ed eroi da ragazzi che hanno origini umili e popolari;
3. è un business.

Dal momento che, oggi, le Società di calcio sono delle vere e proprie multinazionali con un enorme giro di affari, diventa importante la valutazione della loro capacità di generare profitti. Obiettivo della tesi è quello di analizzare l'effetto che alcune variabili calcistiche e extracalcistiche hanno sul successo finanziario di alcune delle più importanti Società calcistiche europee. L'analisi è stata svolta attraverso il modello di regressione per dati panel, con lo scopo di individuare eventuali effetti fissi nel lasso di tempo considerato.

Il primo capitolo è strutturato in modo da avere una visione d'insieme sulla crescita economica delle società calcistiche europee negli ultimi anni. Si parte dall'analisi e comprensione del funzionamento dei proventi delle società calcistiche, per passare successivamente ad esaminare ed elaborare il significato del *brand value*, che sarà preso in considerazione come variabile indipendente ed extracalcistica nel modello di regressione, ed infine verrà illustrata la modalità in cui sono suddivise le entrate totali e l'incremento che hanno avuto nel corso degli anni.

Il secondo capitolo fornisce le basi teoriche riguardanti gli strumenti impiegati per sviluppare l'analisi dei dati, in particolare il modello di regressione lineare multipla con dati panel.

¹ http://www.fifa.com/mm/document/fifafacts/bcoffsurv/bigcount.statspackage_7024.pdf

Nel terzo capitolo viene sviluppato l'*outcome* empirico del modello di regressione per dati panel ad effetti fissi per valutare il peso che le variabili esplicative hanno sul successo finanziario delle società calcistiche esaminate.

1. Crescita economica del calcio europeo

In questo capitolo, verranno illustrati (in linea generale) i flussi dei proventi nelle società calcistiche e le modalità di crescita, negli ultimi anni, delle due maggiori competizioni internazionali, UEFA Champions League e UEFA Europa League, rilevando l'importanza dei crescenti incentivi finanziari che queste competizioni offrono.

Successivamente, verrà descritto il significato del *brand value* e la classifica delle squadre calcistiche con *brand value* più alto, esaminandole nel corso degli ultimi 5 anni

Infine verrà esaminato l'incremento del fatturato nelle squadre europee attraverso la variazione delle variabili prese in considerazione per calcolarlo, in particolare quelle dei 4 migliori campionati europei e il confronto tra il primo e il ventesimo Club per fatturato secondo il *Football Money League Report* di Deloitte.

1.1. Flusso dei proventi

Fino ad alcuni anni fa, le società calcistiche concentravano i loro sforzi e le loro risorse solo alla massimizzazione delle loro prestazioni in ambito agonistico sportivo. Attualmente, con il progressivo aumento di popolarità del gioco del calcio, una società sportiva è tenuta a prendere in dovuta considerazione la valorizzazione del proprio marchio. Una società di calcio è diventata una vera e propria impresa, con fatturati paragonabili a quelli delle multinazionali. La base da cui partire, per capire in linea generale il funzionamento dei proventi, è illustrata nel Grafico 1.



Grafico 1: flusso dei proventi.
Propria rappresentazione

Ogni squadra attraverso le proprie prestazioni che svolge in un determinato arco temporale, raggiunge un determinato risultato sportivo. Quest'ultimo influisce, a seconda dei casi, in maniera positiva o negativa nei confronti del numero di spettatori presenti allo stadio e tifosi/affezionati della squadra sparsi per il mondo. Gli spettatori ed i tifosi/affezionati agiscono sul *Footballmarket* apportando ricavi per le Società sportive in Milioni di euro, attraverso vari modi: merchandising, biglietti delle partite, tournée estive, ecc. Questi ricavi vengono, in un secondo momento, ripartiti tra: premi ai giocatori (per aver disputato un'ottima stagione) ed investimenti per rendere la squadra ancora più competitiva (gli investimenti non riguardano solo giocatori nuovi, essi possono riguardare un allenatore o un dirigente nuovo di fama mondiale). Tramite questo meccanismo, si può notare una crescita dei risultati sportivi; dando così origine ad un circolo vizioso/virtuoso.

È importante, tuttavia, puntualizzare non sempre questo schema viene rispettato perché uno degli elementi più affascinanti del mondo calcistico è rappresentato dalla variabile "imprevedibilità". A volte, infatti, favole come quella del Leicester FC (vincitrice della Premier League nella stagione 2015/2016) sono difficili da pronosticare, il risultato sportivo non è così immediato e scontato.

1.2. L'importanza della UEFA Champions League e UEFA Europa League

Alcuni dei motivi per cui il calcio è uno degli sport più seguiti al mondo sono la UEFA Champions League e la UEFA Europa League.

La Champions League è la competizione internazionale più importante, dove le migliori 32 squadre europee si affrontano partendo da 8 gironi di 4 squadre, che culminano nella partita finale. Per riuscire ad immaginare lo spessore che questa competizione ricopre, è utile presentare alcuni dati forniti dalla UEFA. L'importo netto distribuibile alle squadre partecipanti nella stagione 2016/2017 è di 1.268,9 miliardi di euro, di cui 761,9 milioni rappresentano quote fisse (premi per fase a gironi, prestazioni, bonus qualificazione) e 507 milioni quote variabile (market pool).²

Il Real Madrid che è stata la squadra vincitrice della stagione 2015/2016 ha incassato in totale 82,122 milioni di euro dalla UEFA.³

² <http://it.uefa.com/uefachampionsleague/news/newsid=2398829.html>

³ Real Madrid Management Report & Financial Statement 2015/2016 p.93

Per dimostrare ancora di più la grandezza di questa sfida è fondamentale riportare il numero di spettatori che hanno assistito dalle loro televisioni alla finale di Berlino il 6 giugno 2015. Ammontavano circa a 180 milioni di persone, connesse da oltre 200 nazioni.⁴

Al contrario, l'Europa League è stata per molti anni poco considerata per via dei minori introiti, della minore notorietà delle squadre e del livello di gioco inferiore rispetto a quello della Champions League. Lentamente però anche questa competizione sta emergendo. Le squadre partecipanti sono 48 e dai dati pubblicati nel "2015/2016 UEFA Financial Report" l'importo distribuito è stato di 411,2 milioni di euro, cresciuto del 72% (239,7 milioni) rispetto alla stagione 2014/2015.

Le vincitrici di queste due competizioni si affrontano nella Supercoppa UEFA. Da anni ormai la vincitrice di questa partita guadagna 4 milioni, mentre la perdente 3.

Tutti questi numeri hanno chiaramente delle conseguenze, soprattutto sulle scelte strategiche del Management delle società calcistiche. Come si può notare, partecipare a competizioni internazionali di questo tipo porta numerosi introiti nelle casse delle società. Attraverso questi incassi, una società potrebbe permettersi nuovi (e maggiori) investimenti, così da mantenere o aumentare la competitività a livello nazionale ed internazionale, con l'obiettivo di entrare a far parte del circolo vizioso del flusso dei proventi, spiegato nell'introduzione, in modo di poter sfruttare posizioni dominanti all'interno del panorama calcistico.

Questi guadagni, consentono inoltre, ai conti delle società un pò di respiro, rendendo il loro bilancio (considerato che non tutte vivono situazioni economiche ottimali) leggermente più sostenibile.

1.3. Brand Value

Secondo l'ISO (*International Organization for Standardization*) il Brand è definito come "*a marketing-related intangible asset including, but not limited to, names, terms, signs, symbols, logos and designs, or a combination of these, intended to identify goods, services or entities, or a combination of these, creating distinctive images and associations in the minds of stakeholders, thereby generating economic benefits/value.*"

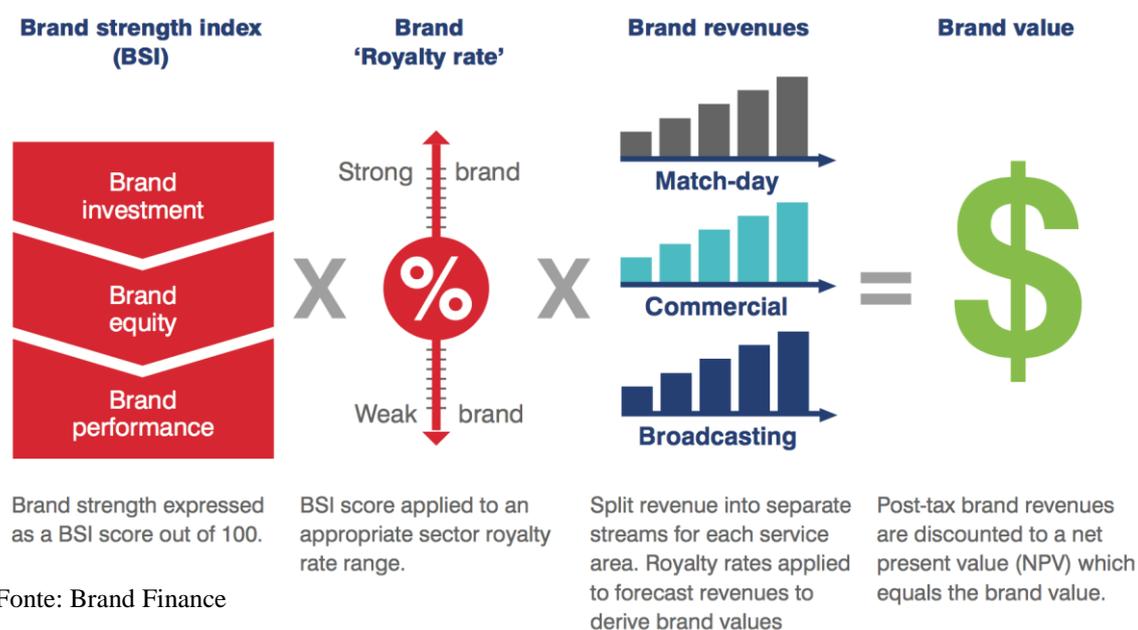
Annualmente, Brand Finance, pubblica report sulle 50 società calcistiche con Brand Value (valore del marchio) più alto. Vi sono numerose definizioni riguardo al valore del marchio di un'azienda; in estrema sintesi, il valore del marchio, non è altro che il *valore aggiunto* con il

⁴ <http://www.uefa.com/uefachampionsleague/news/newsid=2255318.html>

quale un marchio arricchisce il proprio prodotto. Questo valore aggiunto può essere visto dalla prospettiva dell'impresa, del settore o del consumatore.

Brand Finance calcola il Brand Value in questo modo:

Figura 1: Calcolo Brand Value



Chiaramente, più un marchio è globalmente riconosciuto, maggiori saranno i ricavi ottenuti dal suddetto marchio, che, quindi, in quanto “forte”, genererà per esempio, maggiori proventi dal settore commerciale e dal *ticketing*, rispetto ad un marchio “debole”.

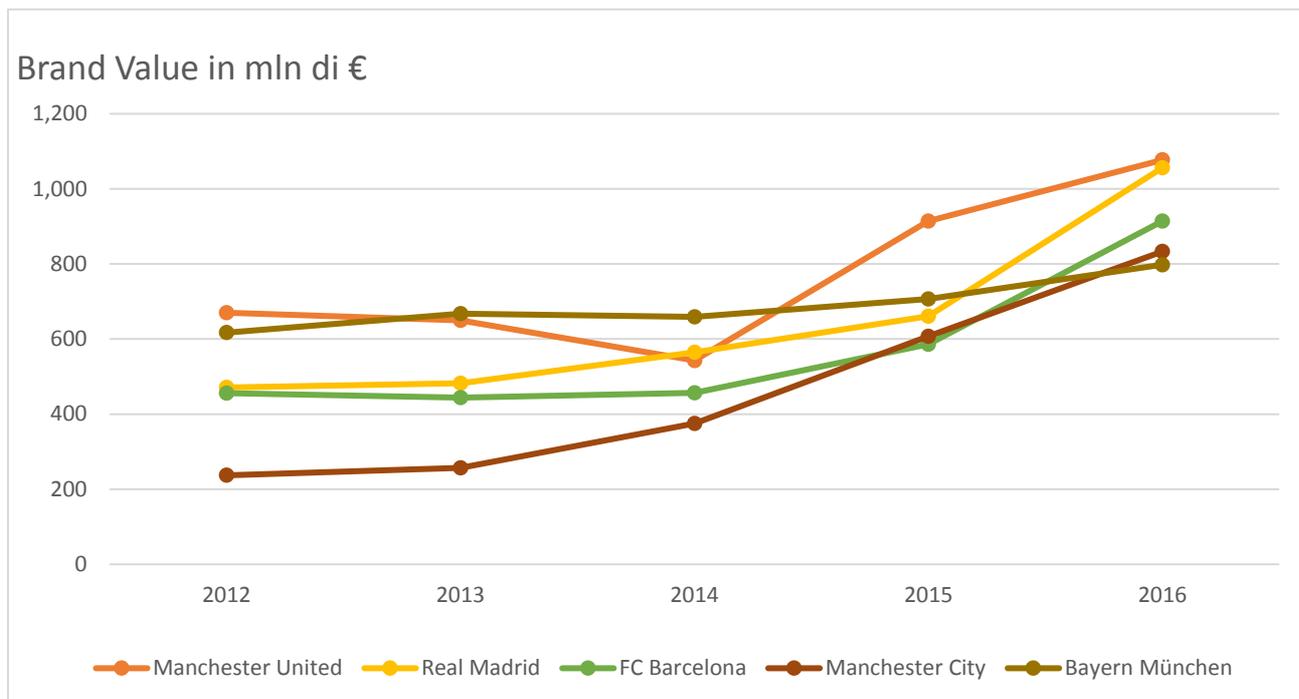
Tabella 1: Brand Value (2012-2016) Top 5 squadre classifica Brand Finance 2016 in mln di €

	2016	2015	2014	2013	2012	Crescita media annua (%)
1. Manchester United FC	1077	914	543	650	670	9.96%
2. Real Madrid CF	1056	661	565	482	471	17.52%
3. FC Barcelona	914	586	457	444	456	14.92%
4. Manchester City FC	833	607	375	257	237	28.58%
5. FC Bayern München	798	707	659	668	617	5.28%

Fonte: Brand Finance. Rappresentazione propria

Per la prima volta nella storia 2 squadre, il Manchester United e il Real Madrid, hanno superato una cifra record: 1 miliardo di euro. Come si può notare dagli andamenti questo numero è destinato a salire e probabilmente, già da quest'anno, il numero di squadre che supererà il miliardo di Brand Value sarà superiore alle 2.

Grafico 2: Andamento Brand Value Top 5



Fonte: Brand Finance. Rappresentazione propria

Andando ad analizzare i fatturati (nel prossimo paragrafo) e il Brand Value, si può notare come il mercato sia dominato da un numero ristretto di squadre che possiede un marchio molto forte. In questo caso, ipotizzando che il mercato sia composto dalle sole 50 squadre con fatturato più grande, secondo Brand Finance, si può osservare come le prime 10 con Brand Value più alto rappresentino il 58% del totale. Qualora il numero delle squadre prese in considerazione aumenti a 100, credo che la percentuale in possesso delle prime 10 rimarrebbe comunque elevata e attorno al 50%.

Questo perché il marchio in loro possesso è troppo più consolidato rispetto alle altre. Sicuramente i motivi per cui sono al vertice sono: la presenza costante di queste squadre nella Champions League, i ripetuti successi sportivi negli ultimi anni e l'acquisto mirato di giocatori che hanno una fama mondiale (che permettono quindi maggiori entrate nella vendita di magliette). Tutti questi fattori incidono positivamente sul Brand Value.

Quest'ultimo paragrafo può essere riassunto con: "chi è forte diventerà sempre più forte."

1.4. La crescita del fatturato nelle squadre europee

1.4.1. Matchday, Broadcast, Commercial

Per semplificazione si ipotizza che il flusso di entrate nelle squadre di calcio dipenda dalla gestione di 3 grandi macro aree: *matchday*, *broadcast* e *commercial*.

Nella macro area del giorno della partita (*matchday*) sono inclusi i ricavi dalla vendita dei biglietti e i servizi di ospitalità aziendale. Nei diritti televisivi (*broadcast*) sono compresi la ripartizione dei diritti tv per le competizioni nazionali (es. Serie A e Coppa Italia) e internazionali (es. Europa League e Champions League). Mentre nell'area commerciale (*commercial*) troviamo i flussi in entrata derivanti dai contratti di sponsorship, vendite del merchandising, tour dello stadio e altre operazioni commerciali.⁵

Deloitte, da ormai 20 anni, pubblica annualmente il Football Money League report, che mostra chi sono le 20 squadre con fatturati più alti del mondo, prendendo in considerazione le 3 aree sopra descritte.

Nella stagione 2015/2016 le entrate totali complessive delle migliori 20 ammontava a 7.417,6 miliardi di euro, di cui: il 18% proveniente dal *matchday*, il 39% dal *broadcast* e il 43% dal *commercial*. Le entrate sono aumentate dell'11,94% (791,2 milioni di euro) rispetto alla stagione 2014/2015.⁶

Il fatturato continuerà a crescere per via dell'aumento dell'acquisizione dei diritti tv dei 5 migliori campionati europei (Inghilterra, Germania, Spagna, Italia e Francia). Ad esempio, nella stagione appena conclusasi (2016/2017) la Premier League, nell'ambito della ripartizione dei diritti televisivi, ha distribuito alle squadre partecipanti complessivamente circa 2,4 miliardi di sterline (+46,4% rispetto alla stagione precedente), di cui, la prima classificata (Chelsea), ha incassato quasi 151 milioni di sterline, mentre l'ultima (Sunderland) 93,5 milioni di sterline.⁷

Il dato che più chiama l'attenzione riguarda la situazione del Sunderland, una squadra che, nonostante l'ultimo posto in classifica e la conseguente retrocessione, conta, dalla ripartizione, un incasso maggiore di quello ottenuto dal Real Madrid dopo la partecipazione e l'annessa vincita della Champions League nella stagione 2015/2016 (si veda paragrafo 1.1.) attribuitogli dalla UEFA. Gli introiti del broadcast potrebbero aumentare ancora di più per via di possibili nuove piattaforme su cui in futuro sarà possibile vedere le partite come Youtube o Facebook.

⁵ Deloitte Football Money League 2017 p.2

⁶ Deloitte Football Money League 2017 p.3

⁷ <https://www.premierleague.com/news/405400>

Sempre restando in Inghilterra, che attualmente è il campionato migliore dal punto di vista economico, l'Adidas ha firmato un contratto nel 2014 con il Manchester United per la fornitura tecnica che porterà nelle casse del club 750 milioni di sterline per i prossimi 10 anni.⁸

In Italia l'unica squadra che riesce a stare al passo delle inglesi è la Juventus che, grazie al nuovo stadio, ha realizzato dal 2011 un ricavo di 9,3 milioni di euro dal museo⁹. Per quanto riguarda la vendita dei biglietti per le partite casalinghe, il primo semestre della stagione 2016/2017 ha portato un ottimo risultato nelle casse del club: 27,7 milioni di euro.¹⁰ Inoltre ha *“perfezionato un accordo con Adidas per la sponsorizzazione tecnica a partire dalla stagione sportiva 2015/2016. Adidas diventerà sponsor tecnico di tutte le squadre Juventus a fronte di un corrispettivo fisso complessivo per i sei anni di durata del rapporto pari a 139,5 milioni di euro”*.¹¹

1.4.2. The Big Four

The big four è il termine che ho scelto di usare per descrivere i quattro migliori campionati europei.

Come detto in precedenza, il campionato migliore dal punto di vista economico è quello inglese, che possiede un elevato vantaggio competitivo rispetto ad altri campionati.

Innanzitutto, in Inghilterra, la maggior parte delle squadre possiede stadi di sua proprietà. Stadio di proprietà è sinonimo di maggiori ricavi.

Inoltre, la stragrande maggioranza dei calciatori, giustamente, ha l'ambizione di confrontarsi con i migliori al mondo e nel miglior campionato del mondo. Quindi, se un giovane fuoriclasse avesse, ad esempio, offerte da squadre inglesi e squadre tedesche, con ampia probabilità sceglierà le prime. La preferenza ricadrebbe, infatti, sulla Premier League, anche in vista delle capacità economiche maggiori di cui le squadre inglesi godono. Capacità che, si traducono, in offerte di contratti più corposi ai giocatori a differenza dei loro competitor.

Nel Football Money League report 2017 (che analizza la stagione 2015/2016) pubblicato da Deloitte, tra le 20 squadre con entrate più alte: 8 sono inglesi, 4 sono italiane, 3 sono tedesche, 3 sono spagnole, una è francese e una è russa. Come si può notare, su 20 squadre, solo due squadre sono al di fuori delle *“big four”*; le *big four*, sostanzialmente, posseggono uno strapotere finanziario.

Per il terzo anno consecutivo, la top 10 comprende le stesse squadre.

⁸ <http://www.bbc.com/news/business-28282444>

⁹ Financial report 30 June 2016 Juventus Football Club S.p.A. p.11

¹⁰ Relazione finanziaria semestrale al 31 dicembre 2016 Juventus Football Club S.p.A. p.8

¹¹ Comunicato ufficiale Juventus Football Club S.p.A.

Tabella 2: Top 20 squadre per fatturato

Classifica	Squadra	Nazione	Introiti in mln di €
1	Manchester United FC	Inghilterra	689
2	FC Barcelona	Spagna	620,2
3	Real Madrid CF	Spagna	620,1
4	FC Bayern Munich	Germania	592
5	Manchester City FC	Inghilterra	524,9
6	Paris Saint-Germain FC	Francia	520,9
7	Arsenal FC	Inghilterra	468,5
8	Chelsea FC	Inghilterra	447,4
9	Liverpool FC	Inghilterra	403,8
10	Juventus FC	Italia	341,1
11	Borussia Dortmund	Germania	283,9
12	Tottenham Hotspur FC	Inghilterra	279,7
13	Club Atlético de Madrid	Spagna	228,6
14	FC Schalke 04	Germania	224,5
15	AS Roma	Italia	218,2
16	AC Milan	Italia	214,7
17	FC Zenit St. Petersburg	Russia	196,5
18	West Ham United FC	Inghilterra	192,3
19	FC Internazionale	Italia	179,2
20	Leicester City FC	Inghilterra	172,1

Fonte: Deloitte Football Money League 2017

Gli elementi che saltano all'occhio sono:

- Nella top 10, 5 squadre sono inglesi.
- La prima squadra italiana, la Juventus, è solo al decimo posto.
- 2 spagnole su 3 si trovano nei primi 3 posti della classifica.
- La Francia ha solo una squadra (Paris Saint-Germain), che però si trova al sesto posto.

Analizzando la ripartizione dei ricavi totali della Premier League, Serie A, Bundesliga e Primera División, si può osservare come la somma dei ricavi totali di Italia, Germania e Spagna

sia quasi pari al ricavo totale della Premier League. Dimostrando ulteriormente la loro supremazia.

Questo perché, come è stato detto da Walter C. Neale nel 1964, le leghe sportive professionali sono caratterizzate da economie di scala.¹²

La ripartizione, oltretutto, ci dice quanto le diverse nazioni siano dipendenti dalle 3 macro aree. Chiaramente i ricavi da matchday, sono inferiori rispetto alle altre 2 aree poiché le partite casalinghe si svolgono in media una volta ogni due settimane e perché, nonostante gli stadi di proprietà portino a numerosi incassi, queste entrate non potranno mai raggiungere o ambire alle stesse cifre dei contratti di sponsorizzazione (tecnica e *main sponsor*) e dei diritti televisivi. L'Italia si dimostra essere troppo dipendente dai diritti tv, facendo fatica (tranne la Juventus) a generare introiti elevati nelle altre aree, così come la Germania è strettamente dipendente dal settore commerciale.

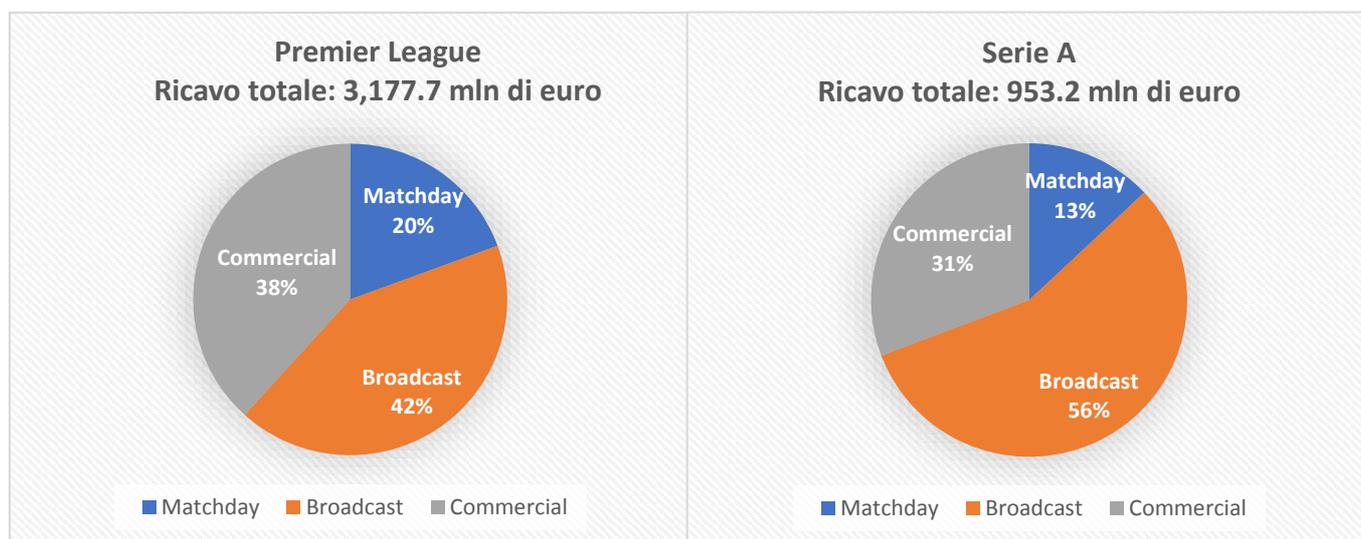


Grafico 4: Ripartizione ricavi totali Premier League. Fonte: Deloitte Football Money League. Rappresentazione propria

Grafico 5: Ripartizione ricavi totali Serie A. Fonte: Deloitte Football Money League. Rappresentazione propria

¹² Neale, W.C. The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. The Quarterly Journal of Economics 1964.

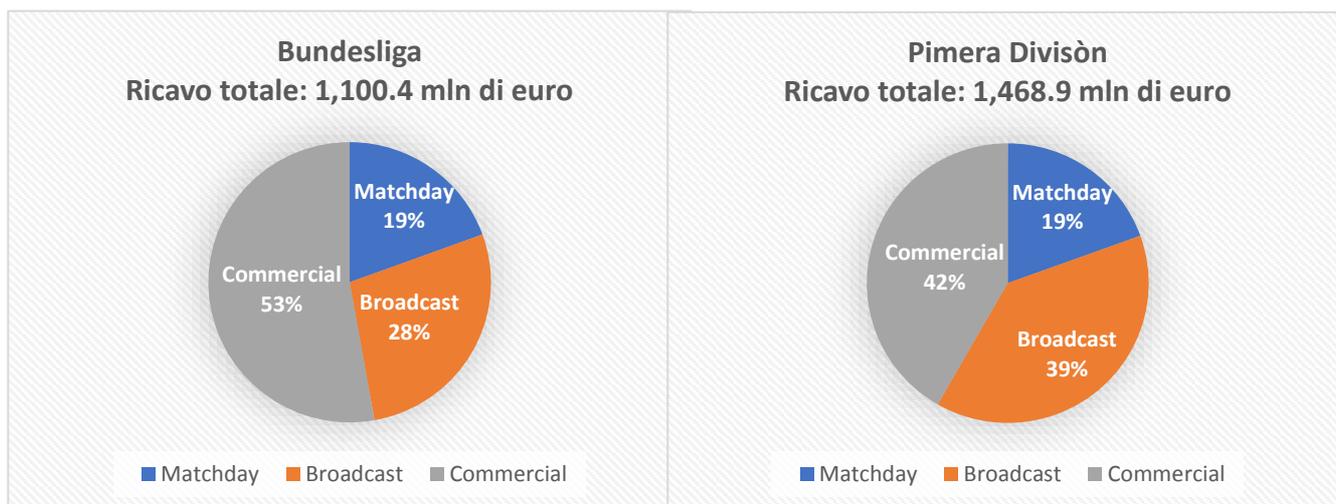


Grafico 6: Ripartizione ricavi totali Bundesliga. Fonte: Deloitte Football Money League. Rappresentazione propria

Grafico 7: Ripartizione ricavi totali Primera División. Fonte: Deloitte Football Money League. Rappresentazione propria

1.4.3. Primi vs. ultimi

Per concludere questo capitolo e dimostrare che il concetto “chi è forte diventerà sempre più forte” nel mondo del calcio è corretto, confronterò la prima e l’ultima per fatturato secondo il Football Money League report dalla stagione 2006/2007 alla stagione 2015/2016 (10 stagioni). Con questo concetto non voglio intendere che, per esempio, la AS Roma (quindicesima in classifica per fatturato nella stagione 2015/2016) non potrà mai aspirare ad arrivare tra le prime cinque posizioni della classifica. La Roma, infatti, per raggiungere quell’obiettivo, dovrebbe arrivare per anni tra le prime 8 squadre in Champions League e vincere spesso il campionato italiano. Per arrivare a ciò, però, servirebbero investimenti enormi in calciatori di fama mondiale ma, con il Financial Fair Play che impone un limite massimo di investimenti per squadra, questo non è pensabile nel breve termine. Inoltre il campionato italiano dovrebbe diventare più appetibile all’estero per consentire così alle squadre di ripartirsi più soldi. In sintesi: per una squadra come la Roma sarebbe più semplice uscire dalla classifica delle migliori venti squadre per fatturato, che arrivare nella top 5.

Un AC Milan invece, grazie a dei periodi entusiasmanti a livello nazionale e europeo dalla stagione 2002/2003 alla stagione 2007/2008 e grazie al livello alto del campionato italiano in quel periodo, raggiunge stabilmente le prime venti posizioni della classifica, nonostante non partecipi alla Champions League o Europa League da 3 anni; dimostrando come la consacrazione costante di qualche anno a livello europeo possa portare risultati nel lungo termine.

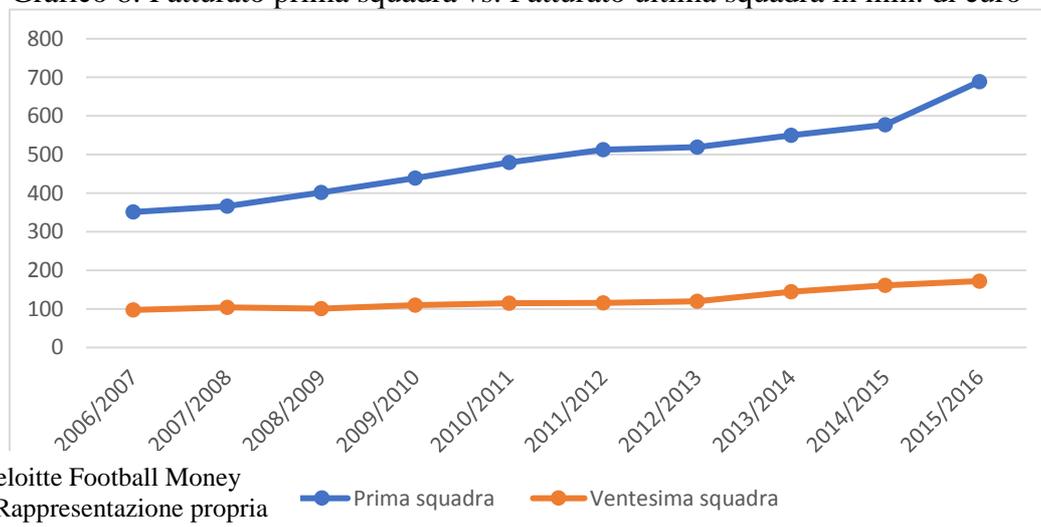
Tabella 3: Crescita fatturato prima e ventesima squadra dal 2006/2007 al 2015/2016

Stagione	Prima squadra	Fatturato in mln di €	Ventesima squadra	Fatturato in mln di €
2006/2007	Real Madrid CF	351	SV Werder Bremen	97.3
2007/2008	Real Madrid CF	365.8	Manchester City FC	104
2008/2009	Real Madrid CF	401.4	Newcastle United FC	101
2009/2010	Real Madrid CF	438.6	Aston Villa FC	109.4
2010/2011	Real Madrid CF	479.5	SSC Napoli	114.9
2011/2012	Real Madrid CF	512.6	Newcastle United FC	115.3
2012/2013	Real Madrid CF	518.9	Club Atletico de Madrid	120
2013/2014	Real Madrid CF	549.5	Everton FC	144.1
2014/2015	Real Madrid CF	577	West Ham United FC	160.9
2015/2016	Manchester United FC	689	Leicester City FC	172.1

Fonte: Deloitte Football Money League. Rappresentazione propria

Come volevo dimostrare, non è casuale il fatto che la prima squadra sia stata per 9 anni su 10 una sola o che, come detto in precedenza, per il terzo anno consecutivo la top 10 è rimasta invariata. La concorrenza per occupare il ventesimo posto è molto alta. Nonostante ciò però, i fatturati di chi è situato alla ventesima posizione non aumentano tanto quanto aumentano gli introiti di chi è al vertice della classifica. Dal grafico 8 si può evidenziare come la “forbice” tra la prima e la ventesima squadra sia destinata a salire. Il tasso di crescita medio annuo della ventesima squadra è del 5.87% mentre per la prima è del 6.98%.

Grafico 8: Fatturato prima squadra vs. Fatturato ultima squadra in mln. di euro



Fonte: Deloitte Football Money League. Rappresentazione propria

2. Teoria Statistica

L'importanza di questo capitolo risiede nelle basi teoriche che saranno rilevanti per l'elaborazione dei dati ottenuti che verranno illustrati nel capitolo 3. In particolare vengono illustrate le caratteristiche principali della regressione lineare multipla e dei dati panel con effetti fissi.

2.1. La regressione lineare multipla

La regressione lineare multipla è una estensione della regressione lineare semplice nel caso in cui si impiegano due o più variabili esplicative (o variabili indipendenti) per effettuare previsioni su una variabile risposta (o variabile dipendente).

2.2. Il modello di regressione multipla

Al fine di tener considerare più di una variabile indipendente, si estende il modello di regressione lineare semplice assumendo che tra la variabile dipendente (Y_i) e ciascuna delle variabili esplicative (X_i) vi sia una relazione lineare. Nel caso di p variabili esplicative, il modello di regressione multipla acquisisce la seguente espressione:

Il modello di regressione multipla con p variabili indipendenti

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad (2.1)$$

dove

β_0 = intercetta della retta di regressione

β_1 = inclinazione di Y rispetto alla variabile X_1 tenendo costanti le variabili X_2, X_3, \dots, X_p

β_2 = inclinazione di Y rispetto alla variabile X_2 tenendo costanti le variabili X_1, X_3, \dots, X_p

β_3 = inclinazione di Y rispetto alla variabile X_3 tenendo costanti le variabili $X_1, X_2, X_4, \dots, X_p$

β_p = inclinazione di Y rispetto alla variabile X_p tenendo costanti le variabili $X_1, X_2, X_3, \dots, X_{p-1}$

ε_i = errore in corrispondenza dell'osservazione i

Nel caso di *due variabili esplicative*, il modello di regressione multipla è espresso come segue:

Il modello di regressione multipla con due variabili indipendenti

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i \quad (2.2.)$$

dove

β_0 = intercetta della retta di regressione

β_1 = inclinazione di Y rispetto alla variabile X_1 tenendo costante la variabile X_2

β_2 = inclinazione di Y rispetto alla variabile X_2 tenendo costante la variabile X_1

ε_i = errore in corrispondenza dell'osservazione i

Nel modello di regressione multipla dell'equazione (2.2.) l'inclinazione β_1 indica come varia Y in corrispondenza di una variazione unitaria della variabile X_1 . quando si tiene conto anche degli effetti della variabile X_2 . β rappresenta, quindi, il *coefficiente angolare della retta di regressione*.

Come nella regressione semplice, i coefficienti di regressione campionaria (b_0 , b_1 , e b_2) vengono usati come stimatori dei corrispondenti parametri (β_0 , β_1 e β_2).

Pertanto, l'espressione campionaria dell'equazione di un modello di regressione multipla con due variabili esplicative ha la forma seguente:

L'equazione della regressione multipla con due variabili esplicative

$$\hat{Y} = b_0 + b_1 X_{1i} + b_2 X_{2i} \quad (2.3)$$

I valori dei coefficienti di regressione campionari si possono calcolare con il metodo dei minimi quadrati.

2.3. Stima dei parametri del modello con il metodo dei minimi quadrati

Come nella regressione lineare semplice, anche per la stima dei coefficienti di regressione si utilizza il metodo dei minimi quadrati. Con questo metodo si assegnano a β_0 , β_1 , ..., β_p quei valori, b_0 , b_1 , ..., b_p che rendono minima la quantità (ai fini della stima si usano i simboli y_1 , ..., y_n per indicare i valori osservati della variabile Y).

Si considerino il vettore dei valori osservati della variabile dipendente

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

e la matrice

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

dei valori osservati delle variabili esplicative. Allora, la stima dei minimi quadrati, del vettore $\beta = [\beta_0, \beta_1, \dots, \beta_n]$ dei coefficienti di regressione del modello è data da:

$$b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y \quad (2.4)$$

Una volta determinati i coefficienti di regressione, il modello di regressione stimato è espresso da:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (2.5)$$

Geometricamente, è l'equazione di un iperpiano nello spazio a $p+1$ dimensioni.

Sostituendo a x_1, x_2, \dots, x_p i valori osservati delle variabili esplicative, si ottengono i cosiddetti valori teorici. Il vettore dei valori teorici $\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]$ è dato da

$$\hat{y} = \mathbf{X}b \quad (2.6)$$

2.4. L'indice di determinazione R^2

La devianza totale della variabile risposta (Y) può essere scomposta nel modo seguente

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.7)$$

dove \bar{y} è la media aritmetica dei valori osservati di Y. Nell'equazione (2.7.), il primo addendo a destra dell'uguale è la cosiddetta *devianza spiegata*, mentre il secondo è la cosiddetta *devianza residua*.

Sulla base dell'equazione (2.7.), è possibile definire una misura della bontà dell'adattamento del modello di regressione lineare ai punti osservati. Si ha:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.8)$$

La formula (2.8) è detta *indice o coefficiente di determinazione* ed è un indicatore dell'idoneità del modello di regressione lineare a rappresentare la relazione statistica tra la variabile risposta e le variabili esplicative. Essa assume valori nell'intervallo [0, 1]: prende il valore 0 quando la conoscenza dei valori delle variabili esplicative non dà alcun contributo alla previsione del valore della variabile risposta ($\hat{y}_i = \bar{y}$, per qualsiasi i); risulta uguale a 1 quando la variabile risposta presenta una relazione lineare perfetta con le variabili esplicative ($\hat{y}_i = y_i$, per qualsiasi i). Di conseguenza, un valore elevato di R^2 indica un ottimo adattamento del modello ai dati analizzati; viceversa, un valore dell'indice vicino allo 0 indica che il modello è inadeguato a rappresentare i dati osservati.

Alcuni ricercatori ritengono che quando si ricorre a un modello di regressione multipla, sia opportuno utilizzare un indice che tenga conto anche del numero di variabili esplicative incluse nel modello e dell'ampiezza del campione, l' \tilde{R}^2 *corretto*. Il ricorso a questo tipo di indice diventa necessario soprattutto qualora si vogliano confrontare modelli di regressione che intendono spiegare la medesima variabile dipendente, impiegando un numero diverso di variabili esplicative. L' \tilde{R}^2 *corretto* è dato dalla seguente formula:

$$\tilde{R}^2 = R^2 - \frac{p(1-R^2)}{n-p-1} \quad (2.9)$$

Un'altra ragione per cui si ricorre all'indice \tilde{R}^2 risiede nel fatto che, quando n è piccolo e il numero delle variabili esplicative è relativamente elevato rispetto a n , R^2 tende a sopravvalutare l'adattamento del modello ai dati.

2.5. Proprietà degli stimatori dei minimi quadrati ordinari

Nell'esposizione del metodo dei minimi quadrati, le quantità y_1, y_2, \dots, y_n sono state considerate come numeri dati. In una visione predittiva, si possono prendere in considerazione, in luogo dei numeri y_1, y_2, \dots, y_n , altrettante variabili casuali, Y_1, Y_2, \dots, Y_n . Con questo cambiamento di prospettiva, è utile riscrivere la (2.4.) utilizzando la lettera maiuscola anche per il vettore dei coefficienti di regressione stimati:

$$B = (X'X)^{-1}X'Y \quad (2.10)$$

Nella formula (2.10), Y e B sono vettori di variabili casuali. In particolare, le variabili casuali componenti del primo vettore sono indipendenti e con la stessa varianza (omoschedastiche), mentre quelle del secondo, come combinazioni lineari di Y_1, Y_2, \dots, Y_n , sono tra loro correlate.¹³ Il vettore B è lo stimatore dei minimi quadrati del vettore dei parametri β . Le proprietà dello stimatore B possono essere riassunte come: *Lo stimatore B è non distorto. Nell'ipotesi che le v.c. Y_1, Y_2, \dots, Y_n siano indipendenti e omoschedastiche, vale il Teorema di Gauss-Markov secondo cui B è lo stimatore più efficiente nell'insieme degli stimatori non distorti espressi da combinazioni lineari delle v.c. Y_1, Y_2, \dots, Y_n .*

Secondo il Teorema Gauss Markov B è BLUE (Best Linear Unbiased Estimator):

- Best = il più efficiente nella classe degli stimatori lineari corretti.
- Linear = lineare rispetto a Y
- Unbiased = corretto
- Estimator = stimatore

La varianza di B è data da

$$\text{Var}(B) = \sigma^2 (X'X)^{-1} \quad (2.11)$$

Da questa stima della varianza di B si deduce che la varianza della generica componente B_j del vettore B è data da

$$\text{Var}(B_j) = \sigma^2 c_{j+1, j+1} \quad (2.12)$$

dove $c_{j+1, j+1}$ è l'elemento della diagonale di $(X'X)^{-1}$ individuato dalla riga $j + 1$ e dalla colonna $j + 1$.

Poiché σ^2 rappresenta la varianza del termine d'errore ε di cui conosciamo n determinazioni campionarie $[e_i, i = 1, \dots, n]$ è possibile basare la stima di σ^2 su una misura della variabilità dei residui campionari. Lo stimatore viene definito in modo da risultare *corretto* per σ^2 .

Uno stimatore *non distorto* della varianza σ^2 della componente di errore ε del modello (2.1.) è dato da

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - p - 1} \quad (2.13)$$

¹³ Il vettore B è il prodotto della matrice $(X'X)^{-1}X'$ per il vettore Y , pertanto la singola componente di B (come prodotto dell'appropriato vettore riga di $(X'X)^{-1}X'$ per il vettore colonna Y) è una combinazione lineare delle v.c. Y_1, Y_2, \dots, Y_n .

2.6. Stima dei parametri con il metodo della massima verosimiglianza

Se nel modello (2.1.) si assume che la componente di errore ε sia distribuita normalmente con media 0 e varianza σ^2 , le stime dei parametri $\beta_0, \beta_1, \dots, \beta_p$ e σ^2 possono essere effettuate con il metodo della massima verosimiglianza e sono date da

$$b = (X'X)^{-1}X'y, \quad (2.14)$$

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}$$

dove y è il vettore dei valori osservati della variabile risposta e X è la matrice dei valori osservati delle variabili esplicative. Lo stimatore di massima verosimiglianza (2.14) può essere riscritto come

$$\tilde{\sigma}^2 = \frac{(y - Xb)'(y - Xb)}{n} = \frac{D_R}{n} \quad (2.15)$$

Lo stimatore di massima verosimiglianza di $\tilde{\sigma}^2$ è distorto e, asintoticamente, la distorsione tende a 0.

2.7. Inferenza sul modello di regressione multipla

Le tecniche di inferenza che saranno presentate da qui in avanti richiedono l'assunzione che la componente di errore ε , che appare nel modello (2.1.), abbia distribuzione normale $\varepsilon \sim N(0, \sigma^2)$. Ciò implica che le v.c. Y_1, Y_2, \dots, Y_n (indipendenti e omoschedastiche) siano normali. Da questa assunzione vale quanto di seguito indicato:

- Il vettore B , stimatore dei minimi quadrati (e di massima verosimiglianza) di β , ha distribuzione normale multipla

$$B \sim N_{p+1}(\beta, \sigma^2 (X'X)^{-1});$$

- La singola componente B_i del vettore B di cui al punto precedente ha distribuzione normale

$$B_j \sim N(\beta_j, \sigma^2 c_{j+1, j+1}), j = 0, 1, \dots, p;$$

- La v.c. $\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{\sigma^2}$ ha distribuzione chi-quadrato con $n - p - 1$ gradi di libertà;
- Se nel modello (2.1.) $\beta_1 = \beta_2 = \dots = \beta_n = 0$, la v.c. $\sum_{i=1}^n \frac{(\hat{Y}_i - \bar{Y}_i)^2}{\sigma^2}$ ha distribuzione chi-quadrato con p gradi di libertà;
- Le v.c. B_i e $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ sono indipendenti;
- Le v.c. $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ e $\sum_{i=1}^n (\hat{Y}_i - \bar{Y}_i)^2$ sono indipendenti.

2.8. Inferenza sui singoli coefficienti di regressione

Come conseguenza della normalità della distribuzione B_j e del fatto che la variabile casuale

$$\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{\sigma^2}$$

ha distribuzione chi quadrato, il rapporto

$$t = \frac{B_j - \beta_j}{\hat{\sigma}_{B_j}}$$

essendo $\hat{\sigma}_{B_j} = \hat{\sigma} \sqrt{c_{j+1,j+1}}$, ha distribuzione t di Student con $n - p - 1$ gradi di libertà. È così possibile costruire intervalli fiduciali o verificare ipotesi sui singoli β_j .

Fissato un livello di significatività α , i limiti fiduciali al livello $1 - \alpha$ per i singoli coefficienti del modello di regressione sono dati da

$$L_1 = B_j - t_{\alpha/2} \hat{\sigma}_{B_j}, \quad L_2 = B_j + t_{\alpha/2} \hat{\sigma}_{B_j} \quad (2.16)$$

Se, invece, si vuole sottoporre a verifica l'ipotesi $H_0 : \beta_j = \bar{\beta}_j$, dove $\bar{\beta}_j$ è un valore prefissato di β_j , contro l'alternativa $H_1 : \beta_j \neq \bar{\beta}_j$, la statistica test da impiegare è data da

$$t = \frac{B_j - \bar{\beta}_j}{\hat{\sigma}_{B_j}} \quad (2.17)$$

E' semplice stabilire che la zona di rifiuto del test è $R = [t : |t| \geq t_{\alpha/2}]$. Un altro modo di procedere è quello di utilizzare il livello di significatività osservato. Se l'ipotesi alternativa è unidirezionale, la zona di rifiuto deve essere opportunamente modificata.

In corrispondenza del valore osservato della statistica test possiamo calcolare il *p-value* che rappresenta la misura del "grado di disaccordo" rispetto all'ipotesi nulla: quanto più è piccolo il *p-value*, tanto maggiore è l'evidenza contro l'ipotesi nulla.

2.9. Verifica della significatività dell'intero modello (test statistico F)

Si consideri di sottoporre a verifica l'ipotesi $H_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$, contro l'alternativa che almeno uno dei coefficienti sia non nullo. Si tratta in realtà di confrontare il modello nel suo complesso con il modello nullo, cioè quello che contiene solo l'intercetta. Un test appropriato per verificare la significatività dell'intero modello è dato dal rapporto

$$F = \frac{D_s/p}{D_r/(n - p - 1)} \quad (2.18)$$

rapporto che sotto H_0 ha distribuzione F con p e $n - p - 1$ gradi di libertà. La zona di rifiuto del test è pertanto $R = [f : f \geq F_\alpha]$. Questo problema viene in genere illustrato tramite lo schema della tabella dell'analisi della varianza (Tabella 4).

Tabella 4: Tabella dell'analisi della varianza

Fonte di variabilità	Devianza	Gradi di libertà	Stima della varianza	F
Regressione	D_S	p	D_S/p	$F = \frac{D_S/p}{D_R/(n-p-1)}$
Errore	D_R	$n - p - 1$	$D_R/(n - p - 1)$	
Totale	D_Y	$n - 1$		

La statistica test (2.18) può essere espressa in funzione di R^2 . Considerando che

$$D_S = R^2 D_Y \text{ e } D_R = (1 - R^2) D_Y ,$$

si può scrivere

$$F = \frac{R^2/p}{(1 - R^2)/(n - p - 1)} \quad (2.19)$$

2.10. I Dati Panel

I dati in formato panel combinano le informazioni relative alle caratteristiche di N individui nello stesso istante temporale con quelle rilevate per gli stessi individui in T diversi periodi di tempo. Nei modelli di tipo panel i dati disponibili hanno perciò entrambe le caratteristiche di:

- Dati *Cross Section* (cross-sezionali): per un dato istante sono osservate le caratteristiche di più individui (N unità cross section),
- Dati *Time Series* (serie storiche): per un dato collettivo di individui sono rilevate le diverse caratteristiche in diversi istanti di tempo (T unità temporali).

Il totale delle osservazioni sarà quindi $N \times T$.

Attraverso i dati panel si possono controllare fattori che: a) variano tra le unità ma non nel tempo; b) potrebbero causare distorsione da variabili omesse se fossero omessi; c) sono inosservati o non misurati, e quindi non possono essere inclusi in una regressione multipla.

La seguente matrice mostra la disposizione dei dati in formato panel relativi ad una variabile Y ; ogni colonna si riferisce ad un diverso individuo per cui la variabile è stata rilevata, mentre per riga sono disposte le diverse osservazioni nel tempo. Ovviamente la variabile Y è composta di NT osservazioni.

$$Y_{(N \times T)} = \begin{bmatrix} y_{11} & y_{21} & \dots & y_{i1} & \dots & y_{N1} \\ y_{12} & y_{22} & \dots & y_{i2} & \dots & y_{N2} \\ \vdots & \vdots & & \vdots & & \vdots \\ y_{1t} & y_{2t} & \dots & y_{it} & \dots & y_{Nt} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ y_{1T} & y_{2T} & \dots & y_{iT} & \dots & y_{NT} \end{bmatrix} \quad (2.20)$$

Poiché i dati *cross section* e quelli *time series* hanno ciascuno le proprie peculiarità, essi portano con sé tutte le complicazioni soprattutto per quanto riguarda il venir meno di alcune ipotesi classiche del modello di regressione lineare

$$Y = X\beta + \varepsilon \quad (2.21)$$

Attraverso la 2.21 è possibile introdurre la notazione. Il vettore Y di dimensione $(NT \times 1)$ è ottenuto applicando l'operatore *vec* alla matrice (2.20) e rappresenta la variabile dipendente, la matrice dei regressori X ha dimensione $(NT \times k)$, mentre il vettore k -dimensionale β contiene i parametri incogniti da stimare. Il termine di disturbo ε ha le stesse dimensioni della variabile dipendente.

Molto spesso i modelli per i dati panel vengono presentati nella formulazione che tiene conto della singola osservazione, quindi l'equazione del modello lineare di regressione diventa

$$y_{it} = x'_{it}\beta + \varepsilon_{it} , \quad (2.22)$$

dove tutte le variabili si riferiscono all'osservazione relativa all' i -esimo individuo nell'istante t -esimo di tempo; in questo contesto y_{it} e ε_{it} sono scalari, mentre la matrice dei regressori è data da un vettore riga con k componenti.

A volte è possibile incontrare una notazione che accorpa tutte le osservazioni relative all' i -esimo individuo per il quale vengono rilevate T osservazioni. L'equazione che ne deriva è la seguente:

$$y_{it} = x_i \beta + \varepsilon_i \quad (2.23)$$

(T×1) (T×k) (k×1) (T×1)

La matrice delle varianze e delle covarianze del termine di errore del modello panel è quadrata, simmetrica, ha dimensione $(NT \times NT)$ ed è definita come

$$\Omega = \text{Var}(\varepsilon) = E(\varepsilon\varepsilon')$$

La convenienza di utilizzare i modelli di tipo panel risiede soprattutto nel guadagnare efficienza della stima perché il maggior numero di osservazioni che si ha rispetto alla sola dimostrazione *cross section* o *time series* genera uno stimatore con varianza più piccola.

2.11. Modello ad effetti fissi

Prendendo in considerazione l'*i*-esimo individuo, il modello ad effetti fissi si configura come segue

$$y_i = \alpha_i + x_i\beta + \varepsilon_i \quad (2.24)$$

dove y_i e ε_i hanno dimensione $(T \times 1)$, x_i ha dimensione $(T \times k)$ e β è il vettore contenente k parametri da stimare. La peculiarità della (2.24) riguarda la costante che si configura come un vettore di T elementi costanti pari ad α_i . Questa caratteristica indica, essenzialmente, che per ciascun individuo occorre stimare un solo valore della costante e che, se $\alpha_i \neq \alpha_j$ per ogni $i \neq j$, tale costante misura l'*effetto individuale*, cioè quell'insieme di caratteristiche specifiche proprie di ciascun individuo che però restano immutate nel tempo. In pratica, nel modello ci sono in tutto $k + N$ parametri da stimare, k contenuti nel vettore β ed N costanti per i diversi individui. Queste costanti rappresentano l'eterogeneità presente tra gli individui nel sistema, caratteristica peculiare dei panel data.

Generalizzando la (2.24) riscrivendola in forma matriciale si ottiene:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \\ y_N \end{bmatrix} = \begin{bmatrix} \iota_T & 0 & \dots & 0 & X_1 \\ 0 & \iota_T & \dots & 0 & X_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & X_{N-1} \\ 0 & 0 & \dots & \iota_T & X_N \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{N-1} \\ \varepsilon_N \end{bmatrix}$$

dove ι_T è un vettore contenente T elementi pari a 1. In forma compatta si ha quindi

$$Y = \left[(I_N \otimes \iota_T) X \right] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon \quad (2.25)$$

oppure

$$Y = \underbrace{(I_N \otimes \iota_T)}_{(NT \times N)} \underbrace{\alpha}_{(N \times 1)} + X \underbrace{\beta}_{(k \times 1)} + \underbrace{\varepsilon}_{(NT \times 1)} \quad (2.26)$$

Poiché i valori del vettore α non sono osservabili essi rientrerebbero all'interno dell'errore del modello, ma, se così fosse, essi potrebbero essere correlati con le variabili esplicative X_i e la stima risulterebbe distorta.

La formulazione (2.26) permette di stimare il modello attraverso la stima dei minimi quadrati, l'OLS, in quanto tutte le ipotesi classiche sono rispettate. Il modello prende il nome di *modello*

a variabile dummy poiché occorre costruire N (numerosità degli effetti individuali) variabili dummy da inserire all'interno della matrice dei regressori. Lo stimatore che si ottiene è, quindi, non distorto, consistente e BLUE.

3. Outcome empirico

Nel terzo capitolo viene sviluppato l'*outcome* empirico del modello di regressione per dati panel ad effetti fissi per valutare il peso che le variabili esplicative hanno sul successo finanziario delle squadre di calcio.

Per la presentazione dell'elaborato, ho scelto di dividere l'ultimo capitolo in 6 parti, partendo dalla modalità di raccolta dei dati per arrivare all'esposizione dei risultati dell'analisi statistica effettuata.

3.1. Raccolta dati

Il lavoro si basa su una raccolta di dati panel delle migliori 20 società di calcio migliori per fatturato secondo il Football Money League 2017 pubblicato da Deloitte. Sono state valutate, ai fini della mia analisi, 14 Club per un arco di tempo di 10 anni, con un totale di 140 osservazioni. Il periodo di tempo considerato va dal 2006 (riferito alla stagione 2004/2005) al 2016 (stagione 2014/2015), escludendo però il 2008 per mancanza di dati disponibili riferiti al *brand value* delle squadre.

Le 14 società sono elencate in ordine al miglior fatturato:

1. Manchester United FC
2. FC Barcelona
3. Real Madrid CF
4. FC Bayern München
5. Manchester City FC
6. Arsenal FC
7. Chelsea FC
8. Liverpool FC
9. Juventus FC
10. Tottenham Hotspur
11. FC Schalke 04
12. AS Roma
13. AC Milan
14. FC Internazionale

Gli altri 6 club (Paris Saint-Germain FC, Borussia Dortmund, Club Atletico de Madrid, FC Zenit Saint-Petersburg, West Ham United FC, Leicester City FC) non sono state oggetto di

analisi a causa della loro presenza non costante all'interno dei report pubblicati da Deloitte, con conseguente mancata disponibilità di dati completi all'interno dell'arco temporale considerato. Dai Football Money League reports di Deloitte provengono sia i dati riguardanti il fatturato sia quelli relativi alle 3 macro aree dei proventi delle 14 squadre. Le 3 macro aree dei proventi comprendono: *matchday*, *broadcast* e *commercial*.

I dati relativi alla vincita delle coppe nazionali e delle competizioni internazionali sono stati collezionati, rispettivamente, dagli albi d'oro dei diversi campionati europei e dalla UEFA. Nelle vincite delle coppe nazionali sono state incluse sia la Coppa Nazionale sia il Campionato; solo in Inghilterra, oltre la Premier League e FA Cup, è stata considerata anche la English Football League Cup (EFL).

Nelle competizioni internazionali sono state considerate la Champions League e l'Europa League.

La media punti a partita della singola squadra è stata ottenuta, grazie agli albi d'oro, dividendo i punti totali conseguiti da una squadra con le partite giocate all'interno del campionato.

I dati relativi al *brand value* sono stati considerati quelli che il *Brand Finance* pubblica annualmente. Il *brand value* dei migliori Club europei è stato considerato come *club value* al netto delle immobilizzazioni materiali fisse (stadio, campo di allenamento), immobilizzazioni immateriali comunicate (giocatori acquistati) ed altre immobilizzazioni immateriali non divulgate (giocatori del vivaio). Il *brand value* include, anche, le informazioni pubbliche disponibili (fornite ad esempio da Bloomberg o dalle relazioni finanziarie annuali dei club) come quelle riguardanti le quote di mercato, la crescita del mercato e la struttura finanziaria del club. Per gli anni 2006, 2007, 2009, 2010, 2011 ho convertito i dati del brand value da sterline a euro secondo il tasso di cambio al 31/12 di ciascun anno.

3.2. Le variabili

Le variabili impiegate sono rappresentate nella Tabella 5.

Dal momento che le società di calcio possono essere considerate vere e proprie aziende, l'obiettivo da loro perseguito è quindi la massimizzazione dei ricavi e il successo sportivo.

La massimizzazione dei ricavi è rappresentata dai *total revenue* delle squadre mentre il successo sportivo è illustrato dai trionfi nazionali, internazionali e la media punti a partita delle squadre. I *matchday revenue*, *broadcast revenue* e *commercial revenue* vengono riportati per mostrare la divisione del ricavo totale all'interno di una società calcistica.

Il *brand value* indicherà la forza del marchio dei club.

Verrà utilizzato il logaritmo naturale dei *total revenue* (l_Total Revenue) e *brand value* (l_Brand Value) per ottenere una più corretta distribuzione delle variabili.

La vincita delle Coppe Nazionali e delle competizioni internazionali sono variabili *dummy* e assumono valore 0 in caso di mancato successo e valore 1 in caso di successo.

Tabella 5: Variabili

Variabile	Scala
Total Revenue	<i>Metric</i>
Matchday Revenue	<i>Metric</i>
Broadcast Revenue	<i>Metric</i>
Commercial Revenue	<i>Metric</i>
Brand Value	<i>Metric</i>
Vincita coppe nazionali	<i>Dummy</i>
Vincita competizioni internazionali	<i>Dummy</i>
Media punti a partita	<i>Metric</i>

3.3. Statistica descrittiva

La statistica descrittiva ottenuta dalle variabili impiegate è illustrata nella Tabella 6.

Dal 2006 (stagione 2004/2005) al 2016 (stagione 2014/2015), escludendo il 2008 (stagione 2006/2007), le 14 squadre sopra elencate hanno prodotto un fatturato totale medio pari a 285,86 milioni di euro.

Nei rispettivi campionati nazionali hanno totalizzato una media di 1,97 punti a partita, che equivale all'incirca al 66% dei punti totali possibili.

Il 39% dei Club ha vinto almeno una coppa nazionale nell'arco di una stagione e l'8% una competizione internazionale.

Il *brand value* medio ammonta a 301,51 milioni di euro (superiore rispetto ai ricavi totali) presentando un'enorme differenza tra il *brand value* minimo e il *brand value* massimo.

Nei ricavi del *matchday*, *broadcast* e *commercial* non ne sono stati calcolati due a causa della mancanza di dati disponibili del Manchester City nella stagione 2005/2006 e della AS Roma nella stagione 2012/2013.

Tabella 6: Statistica descrittiva

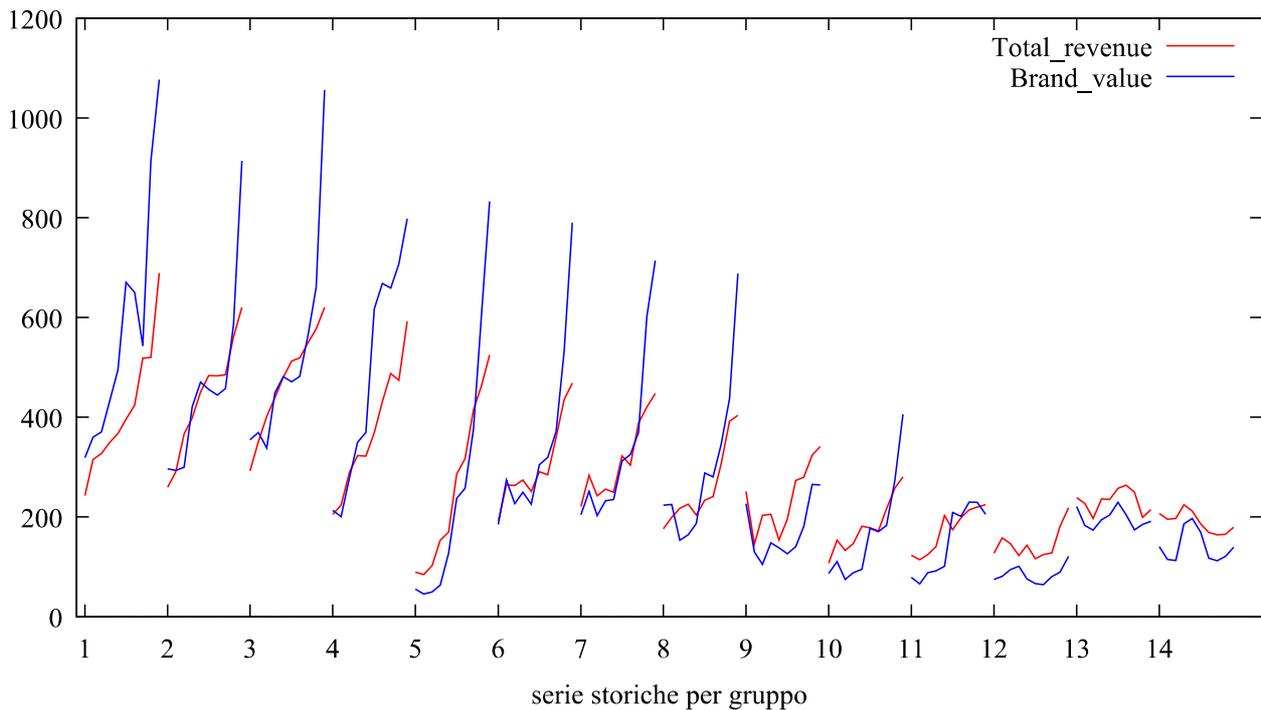
Variabile	N	Media	Mediana	Minimo	Massimo	Dev. Std.
Tot. Rev.	140	285,86	250,45	84,50	689	132,98
Match. Rev.	138	65,45	53,35	7,70	137,50	38,46
Broad. Rev.	138	115,49	107,35	25	227,70	46,95
Comm. Rev.	138	107,45	86,05	21,10	363,80	70,81
Brand Value	140	301,51	227,68	44,99	1077	220,28
Vinc. Cop. Naz.	140	0,39	-	0	1	-
Vinc. Comp. Int.	140	0,08	-	0	1	-
Media punti/partita	140	1,97	1,99	1,13	2,68	0,36

3.4. Serie storiche delle variabili

Tramite l'utilizzo del software multi-piattaforma per l'analisi statistica, Gretl, è stato possibile ricavare grafici delle singole squadre rispetto alle variabili selezionate. Il grafico 9 mostra il modo in cui il *brand value* è cresciuto in maniera esponenziale rispetto al *total revenue*. Questo perché il *brand* è un bene intangibile che include nome, logo e numerosi altri fattori che contribuiscono a determinare immagini distintive e associazioni nella mente del consumatore/stakeholder/mercato generando quindi un beneficio economico all'azienda (si veda paragrafo 1.2.). Essendo il *brand* un valore aggiunto con il quale un marchio arricchisce il proprio prodotto, il dato non corrisponde a delle vere e proprie entrate per le società. È chiaro, tuttavia, che la popolarità di un *brand* influirà sui ricavi attraverso un aumento delle vendite dei prodotti commerciali, che accrescono le entrate totali (effetto traino).

La potenza del brand delle squadre inglesi (1, 5, 6, 7, 8, 10) è evidente: il Tottenham Hotspur (10), nonostante non abbia vinto nulla negli ultimi 20 anni se non la *Football League Cup*, ha paradossalmente un brand value più alto rispetto alla Juventus (9), che negli ultimi 20 anni ha vinto 11 scudetti, 3 Coppe Italia, 6 Supercoppe italiane, 1 Supercoppa UEFA e 1 Coppa Intercontinentale. Probabilmente, per la Juventus ha influito il fatto di essere stata retrocessa in Serie B nella stagione 2006/2007 a causa dello scandalo "Calciopoli". Successivamente, il *brand* Juventus avrebbe dovuto avere un andamento migliore rispetto a quello mostrato. Purtroppo, così come le altre squadre, anche la Juventus ha risentito del livello mediocre e non attraente del campionato italiano, la cui popolarità mondiale è diminuita drasticamente nell'ultimo decennio.

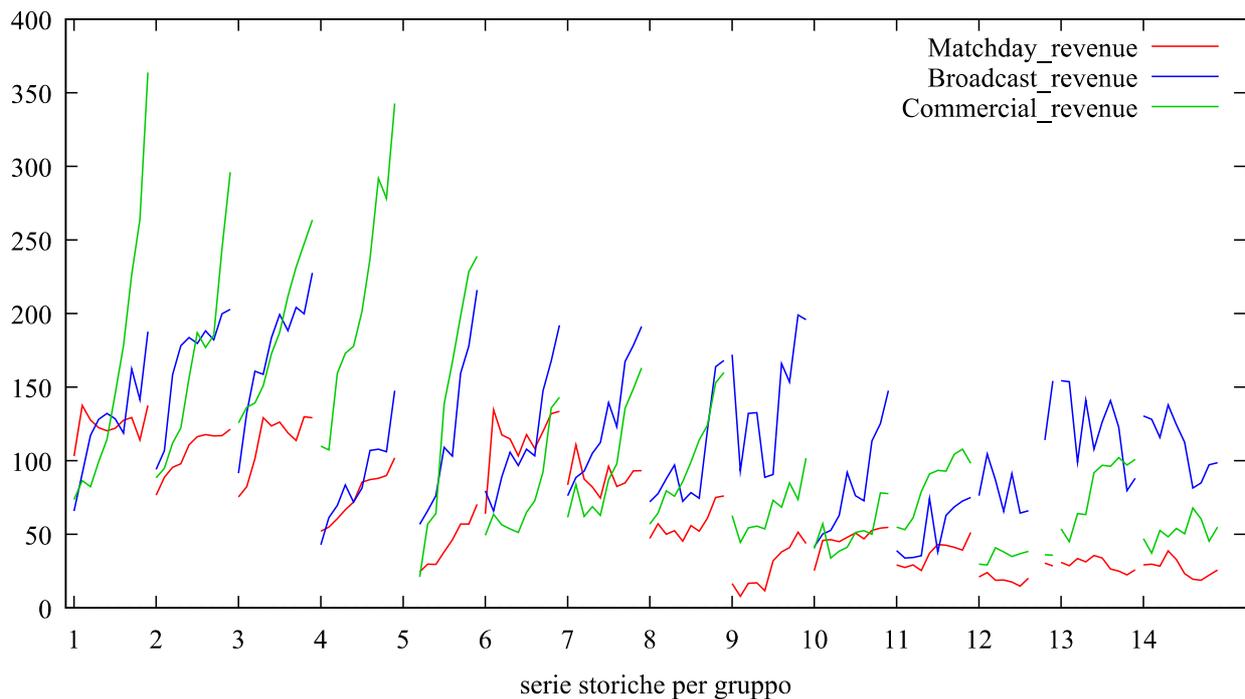
Grafico 9: Andamento Total Revenue e Brand Value delle 14 squadre



Il grafico 10 mostra la distribuzione della macro area con la maggiore influenza sulle entrate totali. Come si può osservare, per 7 squadre (1, 2, 3, 4, 5, 11, 13) gli introiti dal settore commerciale rappresentano il maggiore flusso in entrata, mentre per le altre 7 (6, 7, 8, 9, 10, 12, 14) i flussi maggiori derivano dalla distribuzione dei diritti televisivi. Questi dati suggerirebbero che per le società sportive sarebbe più vantaggioso fare maggiore affidamento sull'area commerciale visto il suo andamento più regolare rispetto al broadcast, che presenta un andamento più instabile.

La macro area *matchday* ha un andamento proprio e non segue logiche ben precise. Osservando il grafico, è evidente come le squadre italiane (9, 12, 13, 14), esclusa la Juventus (9) che dalla stagione 2011/2012 ne usufruisce, abbiano bisogno di uno stadio di loro proprietà per fare il salto di qualità e quindi aumentare le proprie entrate.

Grafico 10: Andamento matchday revenue, broadcast revenue e commercial revenue



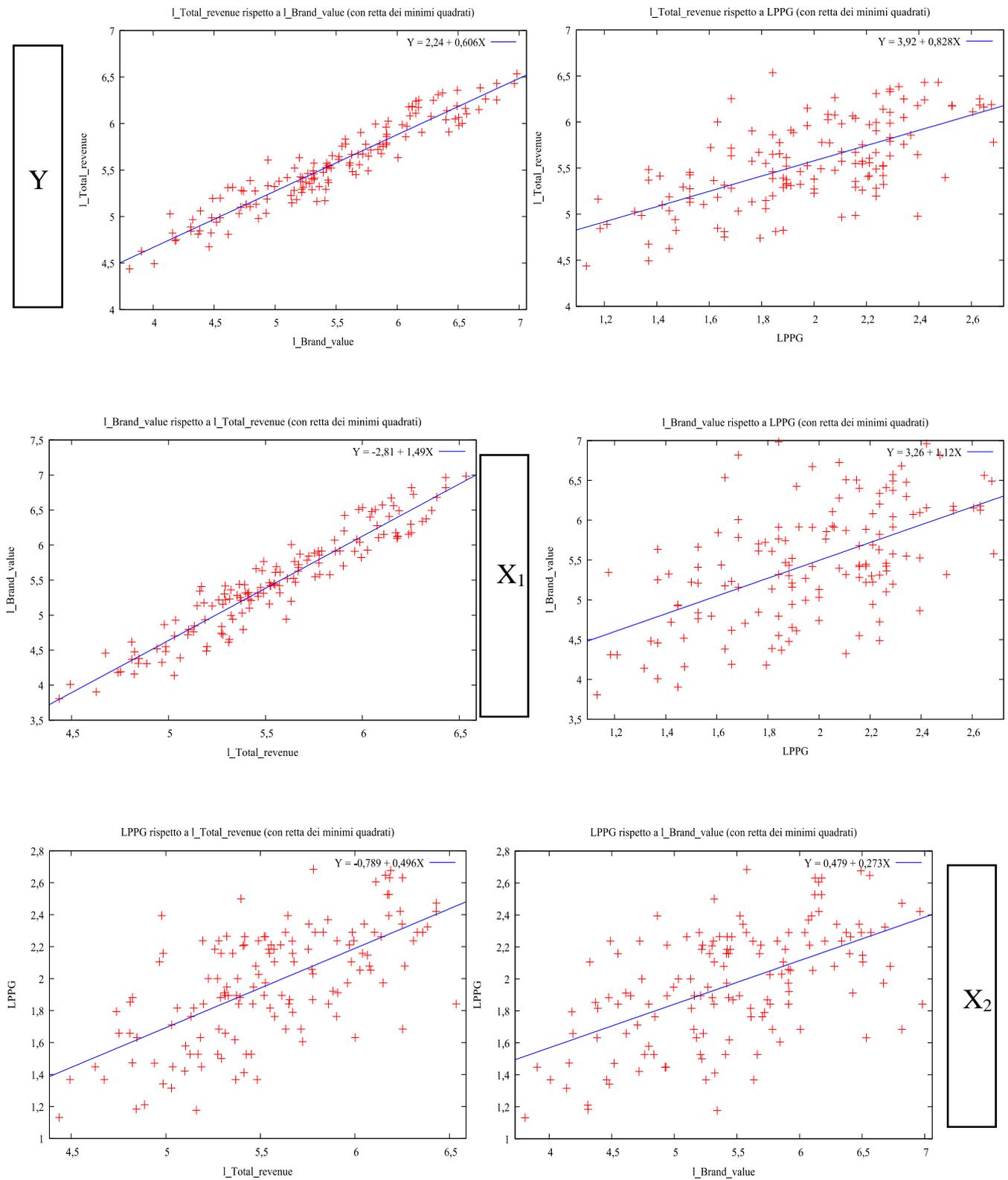
3.5. Diagrammi di dispersione

Prima di mostrare i risultati è opportuno illustrare la matrice dei diagrammi di dispersione. La rappresentazione è ottenuta attraverso la presentazione congiunta dei diagrammi di dispersione conseguiti associando la variabile dipendente con ciascuna delle variabili indipendenti e la singola variabile indipendente con ciascuna delle altre.

L'interpretazione del grafico permette di constatare visivamente la fondatezza di probabili associazioni di tipo non lineare tra la variabile dipendente e le variabili indipendenti, nonché di determinare la correlazione tra le coppie di variabili indipendenti.

La matrice dei diagrammi di dispersione è utile per il riconoscimento di potenziali dati anomali. La variabile dipendente scelta sarà il logaritmo dei *Total Revenue* ($l_Total_Revenue$), mentre le variabili indipendenti saranno il logaritmo del *Brand Value* (l_Brand_Value , X_1) e la media punti a partita delle squadre (LPPG, League Points Per Game, X_2).

Grafico 11: Matrice dei diagrammi di dispersione



3.6. Risultati

Attraverso il campione dei dati è stato costruito un modello di regressione lineare per dati panel ad effetti fissi. L'obiettivo del modello è di individuare l'influenza che le variabili esplicative hanno sul successo finanziario delle squadre di calcio. Il modello è rappresentato dalla seguente equazione:

$$Y = \alpha + B + P + \varepsilon$$

Dove

Y = le entrate totali in forma logaritmica (l_Total_Revenue)

α = è la costante

B = il Brand Value in forma logaritmica (l_Brand_Value)

P = la media punti a partita

ε = la componente di errore

Attraverso il campione di dati empirici è possibile misurare i valori delle variabili con l'utilizzo di del software Gretl. Dopo aver inserito il campione di dati empirici in questione su Gretl e aver scelto il modello panel ad effetti fissi si ottengono le risposte necessarie per analizzare le relazioni intercorrenti tra le variabili selezionate e la significatività delle variabili indipendenti rispetto alla variabile dipendente.

Il risultato ottenuto è il seguente:

Tabella 7: Risultati della regressione

	coefficiente	Errore std.	Rapporto t.	p-value
const	2,19	0,13	16,98	3,80e-34 ***
l_Brand_Value	0,56	0,02	23,55	1,35e-47 ***
Media punti a partita	0,16	0,04	4,07	8,33e-05 ***

Media var. dipendente 5,55 SQM var. dipendente 0,46

Somma quadr. residui 1,83 E.S. della regressione 0,12

R-quadro LSDV 0,94 R-quadro intra-gruppi 0,85

Note: SQM = Scarto Quadratico Medio (o devianza standard); E.S. = Errore Standard

Test congiunto sui regressori

Statistica test: $F(2, 124) = 349,63$

Con p-value = $P(F(2, 124) > 349,63) = 1,07e-51$

Test per la differenza delle intercette di gruppo –

Ipotesi nulla: i gruppi hanno un'intercetta comune

Statistica test: $F(13, 124) = 2,65$

Con p-value = $P(F(13, 124) > 2,646) = 0,0027$

Come detto nel paragrafo 2.4., l' R^2 è un indicatore dell'idoneità del modello di regressione lineare a rappresentare la relazione statistica tra la variabile risposta e le variabili esplicative. Prende il valore uguale a 1 quando la variabile dipendente mostra una relazione lineare perfetta con le variabili indipendenti.

Nel modello da me elaborato R^2 è pari a 0,94 dimostrando, quindi, un ingente potere esplicativo. Tramite il test congiunto sui regressori, il test statistico F (basato su 2 e 124 gradi di libertà) permette di rifiutare H_0 e quindi concludere che vi è una relazione di tipo lineare tra le variabili considerate.

Il test per la differenza delle intercette di gruppo rifiuta l'ipotesi nulla con un p-value pari a 0,0027, dimostrando effetti fissi delle unità.

I dati più importanti che ho ritenuto opportuno analizzare sono quelli che riguardano i coefficienti delle 4 variabili indipendenti. Il coefficiente indica quanto incide la variabile indipendente sulla variabile dipendente a livello percentuale. I valori indicano quanto segue:

- Il *Brand Value* incide sul *Total Revenue* del 56%
- La media punti a partita incide sul *Total Revenue* del 16%

Questi valori indicano che, se la media punti a partita aumentasse di 1 punto, il *Total Revenue* aumenterebbe del 16%.

Il *Brand Value*, invece, essendo calcolato come logaritmo (al pari del *Total Revenue*), influenza direttamente la variabile dipendente (*total revenue*), indicando come un aumento dell'1% del *brand value*, porterebbe il *total revenue* ad aumentare del 56%.

In conclusione, della valutazione di questi valori emerge l'importanza per le squadre di accrescere il proprio *brand value* per far aumentare in maniera consistente le proprie entrate accrescendo il proprio successo finanziario.

Commenti e Conclusione

Attraverso l'analisi svolta, alla domanda iniziale “*quali sono le variabili calcistiche ed extracalcistiche che influenzano di più le entrate totali?*”, può essere data una risposta descrivendo il coefficiente di ogni variabile indipendente all'interno del modello di regressione. L'elevata influenza del *brand value* (56%) può essere ricollegata ai successi sportivi presenti e passati delle relative squadre. Sir Alex Ferguson, è stato alla guida del Manchester United FC per 27 anni, conquistando numerosi successi sportivi e ottenendo il rispetto di tutti gli appassionati di calcio. Dopo il suo ritiro come allenatore, la squadra del Manchester United non ha più riscosso i successi degli anni precedenti. Nonostante ciò, il Manchester United FC continua a trovarsi al vertice alla classifica del *Brand Value* e a quella riguardante le entrate totali. Questo accade perché il fascino del marchio, conseguito e consolidato da decenni di successi a livello Europeo, risulta importante per il consumatore (tifosi e appassionati di calcio) e lo stakeholder (sponsor). Il marchio Manchester United FC è un'immagine distintiva e unica per l'immaginario degli appassionati di calcio.

Il successo nazionale e internazionale di una squadra non comporta, quindi, solo un impatto positivo immediato in termini di entrate totali, ma anche una ricaduta positiva a lungo termine sul valore del *brand* di un Club. Come visto, un aumento del valore del club comporta una crescita esponenziale alle entrate totali.

Un'altra fonte di aumento di valore del brand sono rappresentati dagli investimenti oculati in nuovi calciatori e allenatori volti ad accrescere i successi sportivi della squadra e quindi aumentare indirettamente il valore del Club.

Le società tendono a massimizzare il proprio successo finanziario, onde raggiungere questo obiettivo, bilanciando innanzitutto il loro successo sportivo nelle competizioni nazionali ed internazionali, massimizzando, conseguentemente, il rispettivo *Brand Value*.

Per esempio, poniamo che una società calcistica decida di acquistare un calciatore cinese emergente pagandolo una cifra superiore al suo reale valore di mercato. Sembrerebbe, ad un giudizio sommario, una scelta poco oculata. In realtà, l'investimento consiste nelle diverse opportunità che la società vede nell'acquisto di quel calciatore: oltre alla prestazione calcistica, sta acquistando anche la possibilità di vendere il suo merchandising nel mondo orientale, aprendosi così a nuovi mercati, che genereranno ulteriori introiti. Così, il Club espande l'immagine positiva del suo marchio al mondo Cinese, accrescendo così il proprio *Brand Value*.

Questo è solo un esempio che spiega le cifre record che le società calcistiche stanno spendendo per acquisire le performance sportive dei migliori prospetti mondiali.

Tali strategie sono finalizzate ad accrescere il proprio *Brand Value* e, perciò, i rispettivi introiti totali.

Bibliografia

Rohde M. e Breuer C., (2016). *Europe's Elite Football: Financial Growth, Sporting Success, Transfer Investment, and Private Majority Investors*. International Journal of Financial Studies.

FIFA Communications Division, Information Services, (2007). *FIFA Big Count 2006: 270 million people active in football*.

Neale, W.C. *The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition*. The Quarterly Journal of Economics 1964.

Real Madrid Management Report & Financial Statement. 2015/2016. (p.93)

Deloitte. Football Money League. Sport Business Group. January 2017

Deloitte. Football Money League. Sport Business Group. January 2016

Deloitte. Football Money League. Sport Business Group. January 2015

Deloitte. Football Money League. Sport Business Group. January 2014

Deloitte. Football Money League. Sport Business Group. January 2013

Deloitte. Football Money League. Sport Business Group. February 2012

Deloitte. Football Money League. Sport Business Group. February 2011

Deloitte. Football Money League. Sport Business Group. March 2010

Deloitte. Football Money League. Sport Business Group. February 2009

Deloitte. Football Money League. Sport Business Group. February 2008

Deloitte. Football Money League. Sport Business Group. February 2007

Brand Finance. Football 50 2016. *The annual report on the world's most valuable football brands*. June 2016

Brand Finance. Football 50 2015. *The annual report on the world's most valuable football brands*. June 2015

Brand Finance. Football 50 2014. *The annual report on the world's most valuable football brands*. 2014

Brand Finance. Football 50 2013. *The annual report on the world's most valuable football brands*. May 2013

Brand Finance. European Football Brands 2011. *The annual report on the world's most valuable football brands*. September 2011

Brand Finance. Top 25 Football Club Brands. *The annual report on the world's most valuable football brands*. September 2010

Brand Finance. European Football Clubs. *A review of the top European Football Club Brands*. February 2008

Juventus Football Club S.p.A. Financial report 30 June 2016. (p.11)

Juventus Football Club S.p.A. Relazione finanziaria semestrale al 31 dicembre 2016. (p.8)

Stock J.H. e Watson M.W., (2010). *Introduzione all'econometria*. Pearson

Levine D.M., Kreibel T.C., Berenson M.L., *Statistica*, Pearson, 2010. Capitolo 13

Università di Perugia, Dipartimento di Economia, Finanza e Statistica. *Regressione Lineare Multipla*. Capitolo 22

Palomba G., *Panel Dati*, 2008.

Sitografia

Sito istituzionale della UEFA. 25.08.2016. *“Distribuzione ricavi Champions League 2016/2017”*

Sito istituzionale della UEFA. 08.06.2015. *“Berlin final captures the world’s imagination”*

Sito istituzionale della Premier League. 01.06.2017. *“Premier League value of central payments to Clubs”*

Sito della BBC. 14.07.2014. *“Manchester United and Adidas in £750m deal over 10 years”*