

Anomalous interactions in network of Polish Football League

Andrzej Jarynowski^{a,b}

^aSmoluchowski Institute of Physics, Jagiellonian University,
30-059 Kraków, ul. Reymonta 4, Poland

^bUNESCO Chair of Interdisciplinary Studies, Wrocław University,
50-204 Wrocław, pl. M. Borna 9, Poland

Abstract

In recent years prosecution in Poland has been investigating several clubs, referees and players because of corruption procedures. We study the statistical properties of results in Polish League, looking for evidence of non-sport activity. We treat league as a complex system and we use tools from statistical physics to research some of its properties. Our analyse include: (1) comparing spectrum of correlation matrix with one of the 'cleaner' national leagues-Bundesliga and random matrixes (from different distributions related to this type of data); (2) investigating system of ranked elements in time serie and finding which clubs play for anothers profits; (3) analysing statistical situation before and after matches, which were stated by the court as those in which a crime has been committed, estimating parameters of Kopocinski model (stochastic) to predict propability of appearance of non-sport intervention in investigated match; (4) using other nonlinear measures of dependance to find anomalies. This research is dedicated to release Polish Football from problem of corruption.

słowa kluczowe: szeregi czasowe, statystyka sportowa, modelowanie korupcji, teoria gier, macierze przypadkowe, drzewa minimalnego zasięgu, entropia Shanona

1) Wstęp i motywacja

W naszej pracy spróbujemy zastanowić się nad pewnymi miarami zależności w układzie, które pozwalają lepiej go scharakteryzować i zidentyfikować. Wybraliśmy rozgrywki piłkarskie, ponieważ chcieliśmy sprawdzić: czy na podstawie pewnych właściwości statystycznych można odpowiedzieć na pytania, które stawia Prokuratura Rejonowa we Wrocławiu poszukując dowodów w sprawie procederu korupcyjnego. Na samym początku musimy zaznaczyć, że nasze badania nie mają na celu potwierdzenie bądź zaprzeczenie zależności między zjawiskiem korupcji a właściwościami

statystycznymi układu jakim jest liga. Postaramy się poznać pewne charakterystyczne cechy polskiej piłki nożnej, ocenę jednak pozostawiając czytelnikowi. Podobnie spekulacje teoretyczne, opisujące jakościowo zjawisko korupcji, nie doczekają się weryfikacji w części, gdzie znajduje się analiza ilościowa.

O specyfice polskiego futbolu obecnej dekady może świadczyć poniższy fragment „W śledztwie dotyczącym korupcji w PZPN, które prowadzi wrocławska prokuratura, zarzuty usłyszało około 180 osób. 17 już skazano, proces kolejnych 17 trwa od roku. W "ustawianie" meczów miało być zamieszanych około 50 klubów¹.” Świadomość społeczeństwa o procederze korupcyjnym jest tak duża, że do niedawna zakłady bukmacherskie nie przyjmowały zakładów na pojedyncze mecze, żeby nie ponosić strat na kuponach, które zostały kupione przez graczy, którzy skądś znali już wynik.

Statystyka sportowa jest dziedziną, która może posłużyć zarabianiu, więc np. zakłady bukmacherskie opracowały metody wyznaczania stawek do typowania wyników spotkań. Podstawową różnicą metodologiczną między badaniami niniejszymi a pracowników zakładów bukmacherskich jest cel analizy. Przewidzenie wyniku spotkania to dla nas cel pośredni, a dla nich ostateczny. Dlatego ograniczyłem się tylko i wyłącznie do analizy szeregów czasowych wyników spotkań, gdzie jednostką analizy jest rekord zawierający informacje: kto z kim gra, gdzie, kiedy oraz jaki był wynik. Bukmacherzy, w przeciwieństwie do mnie, posiłkują się wszelkimi dostępnymi informacjami, dzięki czemu mogą korygować prognozy o czynniki takie jak pogoda, kontuzje piłkarzy, problemy finansowe klubu itp. Przywołując analogię giełdową, to tak jakbyśmy pozostali na poziomie analizy instrumentalnej i nie wglębiali się w analizę fundamentalną². Nasze badania to tak naprawdę zespół często niezależnych metod badawczych, w kilku w ogóle nie muszę prognozować wyników poszczególnych spotkań, w związku z tym zdecydowałem się na pozostanie przy „zerowym przybliżeniu” i ograniczeniu się jedynie do „suchych” wyników spotkań. Momentami nasuwały się pokusy, żeby rozszerzyć dane do analizy, przykładowo jak dołączenie do analizy zachowania sędziów i szukanie anomalii w meczach przez nich prowadzonych. Zdać sobie należy sprawę, że zaniedbuję wiele czynników, takich jak finanse lub układy koleżeńskie, wykorzystywane jako dowody w prowadzonym przez Prokuraturę Wrocławską śledztwie. Zdecydowaną przewagą operacjonalizacji do analizy samych wyników meczy jest wykorzystanie wyłącznie obiektywnych źródeł.

Dotychczasowy dorobek nauki na temat piłki nożnej jest raczej niewielki. Ciekawym wyjątkiem jest książka Johna Wessona (fizyka z Uniwersytetu Zachodniej Anglii) pod tytułem „The Science of Soccer” [1]. Jest to obszerne zestawienie dotychczasowej wiedzy na temat motoryki kopania piłki, począwszy od prostej newtonowskiej mechaniki, przez zagadnienia turbulentne, do

1 Informacyjna Agencja Radiowa, 19 grudnia 2008

2 Różnica polega na wykorzystywaniu w analizie fundamentalnej wiedzy m.in. o kondycji przedsiębiorstwa i innych danych makro- i mikro-ekonomicznych w wycenie wartości akcji, gdzie analiza instrumentalna ogranicza się do historii notowań i mierzalnych czynników pochodnych wliczonych na podstawie historii notowań.

probabilistyki wraz ze statystyką z zakończeniem ekonomicznym. Jednak bezpośrednią inspiracją dla mnie były badania Bolesława Kopocińskiego [2,3] – profesora matematyki Uniwersytetu Wrocławskiego oraz Andreasa Heurera [4,5] – profesora chemii Uniwersytetu Centralnej Westfalii.

Kopociński w swoich pracach próbował odpowiedzieć na pytanie: jak mógłby zakończyć się sezon piłkarski 1938/1939, gdyby nie wybuch wojny. W celu ustalenia najbardziej prawdopodobnego rankingu drużyn postanowiono symulować ostatnie mecze.

Do tego został stworzony stochastyczny model składowych wyniku meczu (bramek strzelonych X i straconych Y):

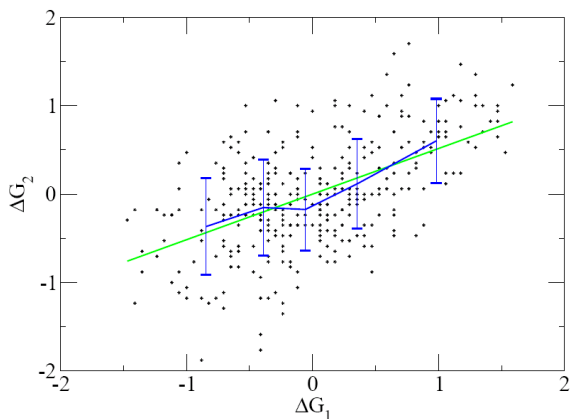
$$\begin{aligned} X_{ij} &= \Pi^{(3)}(a_1(r_{ij})) + \Pi^{(1)}(b(m_{ij})) + \Pi^{(4)}(c_1) \\ Y_{ij} &= \Pi^{(2)}(a_2(r_{ji})) + \Pi^{(5)}(c_2) \end{aligned} \quad (5.1)$$

gdzie:

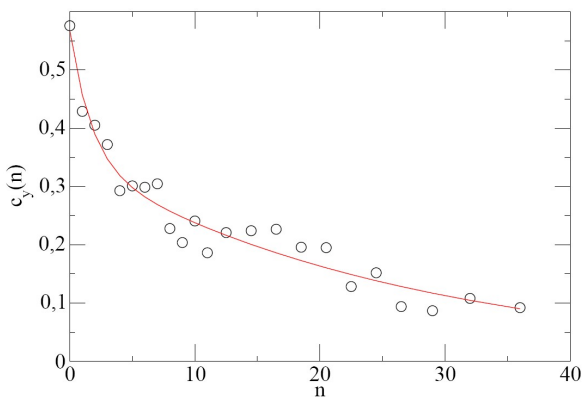
- w ogólności Π są liczbami losowymi z rozkładu Poissona
- $\Pi^{(1)}$ odpowiada za przewagę gry u siebie
- $\Pi^{(2)}$ i $\Pi^{(3)}$ różnicy w sile drużyny
- $\Pi^{(4)}$ i $\Pi^{(5)}$ czynniki losowe
- $a_1(), a_2(), b()$ – są znanymi funkcjami
- c_1, c_2 – parametry dla ligi
- r, m – parametry dla drużyny, z tym, że r_{ij} jest różnicą mocy drużyn i oraz j .

Niemiecka praca natomiast jest statystyczną (bezmodelową) analizą Bundesligi, dla której punktem wyjścia są moce drużyn wyznaczone przez różnice goli w meczu. Autorzy postawili sobie za cele: obserwację zmiany mocy drużyny w czasie (a właściwie w przedziałach czasowych) oraz weryfikację panujących w środowisku mitów. Już na samym początku przeddefiniowana została miara mocy drużyn z punktowej, używanej w rankingu ligowym, (0-przegrana, 1-remis, 3-zwycięstwo), na mniej obciążoną i zarazem bardziej naturalną różnicę między bramkami strzelonymi a straconymi. Argumentem za wprowadzeniem miary różnic goli ΔG jest zachowanie informacji -w mierze punktowej nie ma różnicy czy drużyna wygrała 1:0, czy 6:0. Historyczne dane zostały podzielone na poszczególne okna czasowe po to, by móc zestawić różne charakterystyki w czasie. W celu porównywania skumulowana liczba goli w oknie czasowym musi być znormalizowana, poprzez podzielenie przez liczbę spotkań w danym oknie czasowym. ΔG będzie mówiło o różnicy goli przypadających na jeden mecz. Najbardziej naturalny podział to sezonowy, ale w sezonie również można znaleźć podział na połowy³. W ten sposób można przedstawić funkcję regresji między różnicą strzelonych bramek w pierwszym i drugim sezonie [Wyk. 5.1]. Autorzy podzielili sezony na kwarty, badali korelacje między nimi oraz wyznaczyli krzywą autokorelacji [Wyk. 5.2], jak również dopasowali rozkłady do samej różnicy bramek w oknie czasowym [Wyk. 5.3].

3 sezon piłkarski w krajach Europy Środkowej i Północnej ma przerwę zimową z powodów klimatycznych i zima dzieli sezon właśnie na dwie części, aczkolwiek w krajach Europy Południowej przerwa zimowa nie występuje, najwyższej przerwa świąteczna)



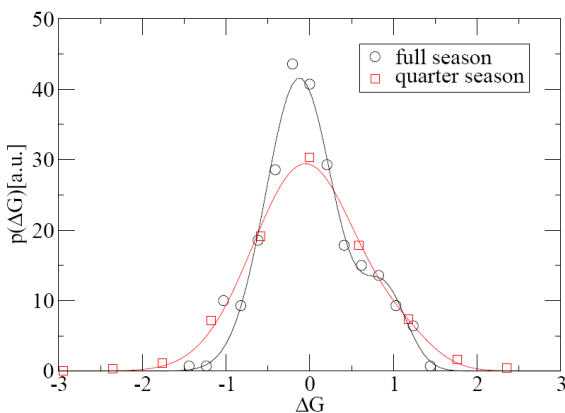
Wyk. 5.1 [Heuer⁴] Regresja między znormalizowaną liczbą goli w pierwszej oraz drugiej połowie sezonu. Współczynnik kierunkowy wynosi 0,53 a odchylenie standardowe 0,51.



Wyk. 5.2 [Heuer⁵] Przedstawiono zależność średnich autokorelacji ΔG dla połówek sezonów w kolejnych latach n . Jak widać tylko dla bezpośrednio następującej połówki korelacja jest większa od 0,5. Wynika to z właśnie podziału sezonu na część jesienną i wiosenną.

4 Heurer [5] Wybrane wykresy z przedstawionej publikacji

5 *ibidem*



Wyk. 5.3 [Heuer⁶] Gęstość prawdopodobieństwa znormalizowanej różnicy bramek dla całego sezonu i kwarty. Obie gęstości można przybliżyć rozkładem Gaussa z uwzględnieniem dodatkowej frakcji dla dużych ΔG . Autorzy przedstawili to jako sumę dwóch rozkładów Gaussa, przy czym ten odpowiadający za prawą frakcję jest $1/6$ intensywności głównego wyrazu. Proszę również zauważyć, że rozkład ΔG dla dłuższych szeregów jest węższy.

Niemcy podjęli również problem estymowania rzeczywistej mocy drużyny, mając jedynie skończone szeregi czasowe. Nie omawiając dokładnie argumentacji autorów, przytoczę poprawiony estymator i heurystycznie się do niego odnoję. Tak więc za punkt wyjścia należy przyjąć liniowe przekształcenie unormowanej różnicy bramek dla okna czasowego długości N , czyli $\Delta G(N)$, z tym że współczynnik kierunkowy jest zaczerpniętym z teorii pomiaru stosunkiem wariancji wartości teoretycznej (sygnału) do zaszumionej, a wyraz wolny odchyleniem standardowym pomiaru.

$$\Delta G = a_N (G(N) \pm \sigma_N) \quad (5.1)$$

Zamiast wyznaczać parametry a_N i σ_N możemy je w pewnych przypadkach, np. dla połowy sezonu, przybliżyć wielkościami wyliczonymi w regresji, odpowiednio: współczynnikowi kierunkowemu i odchyleniu standardowemu [Wyk. 2.1]. Moim zdaniem jednak powyższe poprawienie estymatora ΔG powoduje iż niepewność porównywania drużyn jest zbyt duża. Na przykład, wedle powyższego wzoru, drużyny, dla których na koniec sezonu ($N=34$) ΔG różni się o 12 bramek z prawdopodobieństwem $1/4$ mogą być równe sobie, bądź gorsza okazałyby się teoretycznie lepsza.

2) Zjawisko korupcji

Nie trzeba nikogo przekonywać, że korupcja jest poważnym problemem w Polsce, ale i nie tylko. Ośrodek Badania Opinii Publicznej opublikował, że o wszechobecności korupcji w Polsce było przekonanych 91% badanych. Trudno oszacować straty ekonomiczne, społeczne, czy moralne, jakie ona wyrządza. Afera korupcyjna wokół piłki nożnej jest tylko jedną z wielu i stosunkowo mało groźną w porównaniu z aferami gospodarczo-politycznymi. O tym, że wyniki spotkań ustalane są przy „zielonym stoliku” a nie na zielonej murawie mówiło się już od dłuższego czasu [Obr. 5.1]. Niestety, dosyć długo nic z tym stanem rzeczy nie robiono. Dopiero zmasowana akcja medialna i pierwsze zatrzymania odkryły skalę zjawiska [Obr. 5.2]. Właśnie dzięki działaniu prokuratury i niezależnych dziennikarzy mogłem przystąpić do analizy- informacje o skazanych zespołach pozwoliły mi zawęzić poszukiwania.



Obr. 5.1) Klatka z filmu „Piłkarski Poker” [reż. Janusz Zaorski]



Obr. 5.2) Zatrzymanie przez prokuraturę wrocławską sędziego Ryszarda F.

(pseudonim „Fryzjer”) oskarżonego o „drukowanie” (ustawianie) wyników spotkań [fotoreportaż PAP z 05.07.2006]

2a) Hipoteza korupcyjna

Nie dochodząc przyczyn społeczno-politycznych, przytoczę tylko powody bezpośrednie występowania korupcji, bo jest to istotne w zrozumieniu metod i technik, które zastosowałem. Poza niektórymi wpływami (jak przychód ze sprzedaży praw do transmisji telewizyjnej), dochód drużyny zależy głównie od zajęcia wysokich lub najniższych miejsc w tabeli. Domniemywając istnienia kontaktów pozasportowych między klubami można dojść do racjonalnych ekonomicznie transakcji sprzedaży- kupna meczu, wynikających z praw popytu. W obliczu korzyści etyka i sport może nie mieć znaczenia i w efekcie przejść na drugi plan. Dla drużyn ze środka stawki, dla których krańcowa użyteczność zwycięstwa jest niższa niż krańcowa użyteczność porażki.

Jeżeli przyjmujemy, że korupcja w polskiej piłce rzeczywiście występuje, to nasuwa się pytanie jak ją mierzyć? Samo zjawisko korupcji jest na tyle skomplikowane, że najpierw wypadałoby ją zdefiniować, przynajmniej operacyjnie (konceptualizacja, aczkolwiek ciekawa, nie mieści się w obszarze badań fizyka). Moim pierwszym założeniem, było wyciągnięcie informacji o systemie z obserwacji anomalii. Mogłoby się wydawać, że czynnik losowy w sporcie jest decydujący (świadczą o tym przytoczone w poprzednim rozdziale prace Heurera i Kopocińskiego). Dlatego właśnie występowanie anomalii⁷ może świadczyć o zaburzeniu naturalnej harmonii. W związku z tym postawiłem hipotezę zerową *„If it is true, that many of matches were „fixed”, so Polish League's statistics should be different than other „sport” leagues like Bundesliga. There should be also anomalies compared to stochastic pattern.”*

„Jeżeli wiele spotkań zostało „ustawionych”, to pewne statystyki polskiej powinny się różnić od innych „sportowych” lub „grających fair” lig jak Bundesliga. Powinny również występować anomalie w porównaniu do stochastycznego wzorca ligi.”

To właśnie poszukiwanie anomalii oraz anomalnych oddziaływań między zespołami jest sednem tej pracy. Jak okazało się w wyniku badań, można postawić hipotezę alternatywną: „(...)Anomalie powinny występować jednak rzadziej, niż by to wynikało z stochastycznego wzorca ligi”.

Do hipotezy zerowej pasowały by takie zachowania:

- zwycięstwa słabych drużyn, walczących o utrzymanie;
- podkładanie się przeciętnych drużyn, pewnych utrzymania w danej klasie rozgrywek, w meczach z drużynami walczącymi o utrzymanie;
- innego rodzaju przekupienie sędziów, piłkarzy czy działaczy, skutkujące niespodziewanym wynikiem spotkania.

Na poczet hipotezy alternatywnej wysuwają się:

- zwycięstwa dobrych drużyn, walczących o mistrzostwo bądź prawo gry w

7 Nie definiujemy tego pojęcia, bo będziemy wykorzystywać je w jego obszarze znaczeniowym

Europejskich Pucharach, jak również awans do wyższej klasy rozgrywek ;
 –podkładanie się przeciętnych drużyn, pewnych utrzymania w danej klasie rozgrywek, w meczach z drużynami walczącymi o mistrzostwo bądź Europejskie Puchary, jak również o awans do wyższej klasy rozgrywek;
 –innego rodzaju przekupienie sędziów, piłkarzy czy działaczy, skutkujące zapewnieniem spodziewanego wyniku spotkania.

2b) Próba modelowania korupcji – teoria gier

Spróbujmy do rozgrywki ligowej podejść w ten sposób, że kolejne mecze są sekwencjami gier (nie tylko sportowych, ale również ekonomiczno-politycznych). W pojedynczym spotkaniu drużyna może wybrać 2 strategie: zostać donorem (D) albo akceptorem (A) łapówki. Przypadek meczu uczciwego opisywałaby gra, kiedy zawsze obie drużyny chciałyby być akceptorem łapówki i nie dochodziłoby do transakcji, a o wyniku spotkania decydowałyby tylko czynniki sportowe. W przypadku wyboru strategii A przez jedną drużynę oraz D przez drugą dochodzi do transakcji. Kiedy natomiast obie drużyny chciałyby dać łapówkę może dojść do sytuacji, kiedy jedna drużyna kupuje mecz np. u sędziego i transakcja nie zachodzi między obiema drużynami, a tylko między jedną z nich a np. sędzią (dla ułatwienia przyjmijmy, że tylko gospodarz może przekupić sędziego). Przypadkowi D-D można by było przypisać również sportowe rozstrzygnięcie (obie drużyny chcą wygrać) i wtedy interpretowany byłby tak samo jak A-A. Dla konsystencji modelu uznałem jednak, że jeśli drużyny wybierają strategię D, to i tak wręczą łapówkę- jeśli nie innej drużynie, to sędziemu. Skonstruowana, wedle powyższych założeń, macierz wypłat bazuje na mierze użyteczności różnych efektów. Dlatego musimy sprowadzić do tych samych jednostek takie czynniki, jak: sportowa ilość strzelonych bramek z modelu stochastycznego Kopocińskiego, kwota pieniężna łapówki, zysk lub strata w punktach, czy koszt moralny uczestniczenia w procederze korupcyjnym.

$\frac{1}{2}$	D	A
D	$(W1-L, W2+G-C-M)^*$	$W1+G-C-M, W2-L+C-M$
A	$W1-L+C-M, W2+G-C-M$	$W1, W2$

Tab. 5.1 Macierz wypłat *(dla ułatwienia przyjmijmy, że tylko gospodarz może przekupić sędziego, choć w ogólności może to być drużyna, której bardziej może zależeć na zwycięstwie, bądź ma więcej pieniędzy)

Rozwijając symbole w [Tab. 5.1]:

$$\begin{aligned}
 w2 &= \Pi(ar_{ij} + bm_i + c) \\
 w1 &= \Pi(ar_{ji} + c)
 \end{aligned}
 \tag{5.2}$$

W1 i W2- Użyteczna wartość stochastycznych funkcji zdobycia bramek w meczu
 w1-gospodarz i w2-gość (parametry wewnętrzne: różnica w sile drużyn, przewaga gry u siebie oraz czynnik losowy zostały wyjaśnione w modelu Kopocińskiego)

G- Użyteczna wartość zysku w punktach

L- Użyteczna wartość straty w punktach

C- Użyteczna wartość kwoty pieniężnej łapówki
M- Użyteczna wartość kosztu moralnego

Analiza została przeprowadzona zgodnie z wytycznymi teorii gier, której autorstwo przypisuje się Johnowi von Neumanowi w roku 1928, aczkolwiek równoległe problem opisał polski matematyk związany z Wrocławiem: Hugo Steinhaus [6]. W naszym modelu graczami są kluby, a strategiami gry są D i A. Cały sezon odpowiadałby ewolucyjnym grom, a wtedy parametry modelu zależałyby od czasu. Niestety, powyższa analiza ma tylko charakter orientacyjny i jej dokładniejsze przeprowadzenie nie mieści się w zakresie tej pracy. Gdyby jednak udało się znaleźć sposób na sprowadzenie wszystkich czynników do jednostek użyteczności (o czym w ekonomii klasycznej i neoklasycznej dużo zostało powiedziane), można by było poszukać równowag (Nasha, Min-Max) i porównać otrzymane strategie. Zaznaczam, że teoria gier wielokrotnie została wykorzystana do opisu zjawiska korupcji [7]. Umiejętne jej zastosowanie dałoby więcej światła na same mechanizmy korupcyjne, a stąd już tylko krok do zwalczania tego procederu .

3) Na tropie korupcji

W tym rozdziale opiszemy podejścia, które wykorzystaliśmy w poszukiwaniu anomalnych interakcji. Sformułowanie „na tropie” jest po dwakroć słuszne: raz, że korupcja jest czynem karanym (KPA-Kodeks Postępowania Administracyjnego, KK-Kodeks Karny), ja staram się pomóc w zrozumieniu zjawiska, co później może zostać wykorzystane w wykryciu bądź prewencji; po drugie „tropem” są tylko poszlaki, „artefakty statystyczne”, pozostałe w gąszczu szumu stochastycznego.

3a) Macierze przypadkowe i analiza spektralna

Teoria macierzy przypadkowych została wymyślona na potrzeby fizyki teoretycznej kwantowych układów wielu ciał przez Eugena Wignera około roku 1950. Chodziło mu o przedstawienie hamiltonianów w reprezentacji macierzowej (Heisenberga), opisującej przypadkowość panującą w układzie. Bardzo szybko teoria znalazła swoje zastosowania [8] w wielu dziedzinach fizyki (np. w opisie lokalizacji), jak również w innych dziedzinach (np. przy minimalizacji ryzyka portfela finansowego). W przypadku ligi piłkarskiej zastosowałem tę metodę, traktując ligę jako układ o N stopniach swobody (gdzie N jest liczbą badanych drużyn w lidze), a zmiennymi będą szeregi czasowe wyników meczów. Aby móc w pełni skorzystać z teorii macierzy przypadkowych szeregi czasowe M powinny być nieskończenie długie, a przynajmniej jak najdłuższe. Tak samo w granicy N powinno dążyć do nieskończoności tak, żeby iloraz N/M był stały i skończony. Znajomość parametru N/M pozwala nam ustalić naturalny przedział występowania wartości własnych⁸.

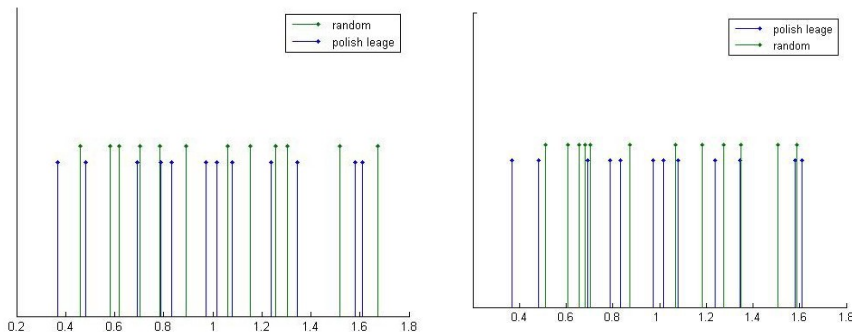
8 Wishard w teorii był w stanie nakreślić wykres gęstości spektrum mocy, ale my będziemy bardzo daleko od granicy

Metoda macierzy przypadkowych polega na porównaniu widma macierzy korelacji $N \times N$, gdzie każdy wiersz czy kolumna odpowiada jednemu zespołowi z widmem macierzy przypadkowej powstałej z procesu stochastycznego o średniej zerowej i podobnym zakresie co wyniki piłkarskie. Do opisu wyniku meczu można przyjąć kilka miar:

1. punktową (3-zwycięstwo, 1-remis, 0-porażkę)
2. różnicę liczby strzelonych bramek
3. "logiczną" (1, 0 -1)

Ostatnia miara logiczna jak i druga, bramkowa, spełniają zależność, że wartość średnia jest zerowa, jak i wartości są symetryczne, więc te miary są wiarygodne pod względem statystycznym (pierwszy i trzeci moment centrowe są zerowe). Dlatego też stosowałem najczęściej miarę drugą, ponieważ zawiera najwięcej informacji (choć traci na interpretacji wynikowej).

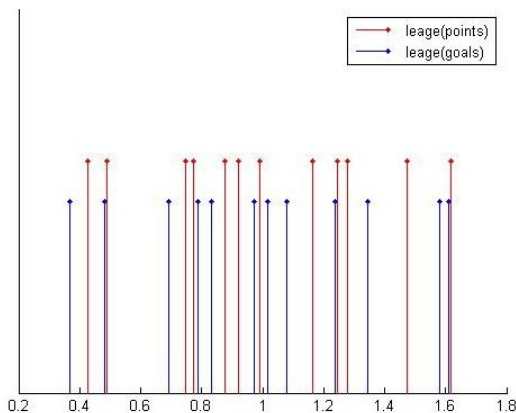
Wyjściowym zbiorem jest 12 drużyn, które w ciągu 3 sezonów (90 spotkań), nie spadły z Ekstraklasy. Są to wartości dalekie od nieskończoności. W związku z tym, że liczba drużyn – liczba stopni swobody, wynosi 12, to otrzymamy również tylko 12 wartości własnych. Z takiej ilości nie da się utworzyć histogramu ani wyestymować funkcji gęstości. Pociąga to za sobą konieczność porównywania widm, a nie gęstości widmowej. Do tego celu posłużyłem się pojedynczymi reprezentacjami losowych macierzy przypadkowych powstałych z rozkładu jednorodnego na przedziale $[-4;4]$, który mógłby opisywać różnicę bramek w meczu [Wyk. 5.4].



Wyk. 5.4) Spektra widmowe macierzy korelacyjnej ligi piłkarskiej z procesem losowym jednorodnym na $[-4;4]$. Lewy wykres przedstawia reprezentację ciągłego, a prawy dyskretnego procesu losowego, przy tych samych wartościach własnych ligi piłkarskiej. Proszę nie doszukiwać się wartości na osi y , ponieważ słupki odpowiadają pojedynczym wartościom własnym na osi x , a wysokość słupka nie ma żadnego znaczenia (pozwala jedynie rozróżnić wartości losowe-te wyższe od ligowych-niższe)

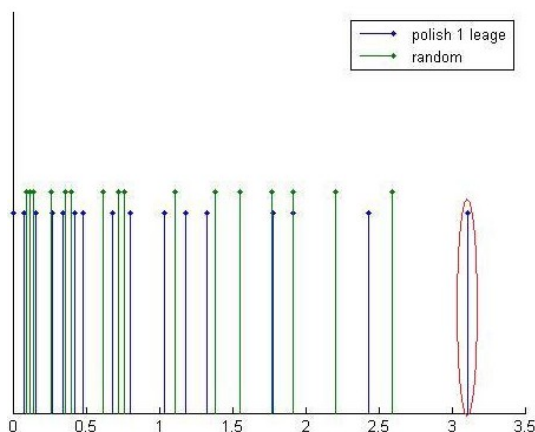
Warto by się było w tym miejscu przyrzeć jak prezentuje się widmo

sporządzone dla miary punktowej oraz jak różni się od miary bramkowej, którą wybrałem ze względu na własności statystyczne. Otóż nie różnią się widma w sposób znaczny co może znaczyć, że obie miary stanowią porównywalny wskaźnik dla ligi. Porównując [Wyk 5.4] i [Wyk 5.5] zauważyć można, że obszar wartości własnych ligowych niezależnie od miary leży w obszarze losowym. Przez to nie da się znaleźć pewnych wewnętrznych zależności.



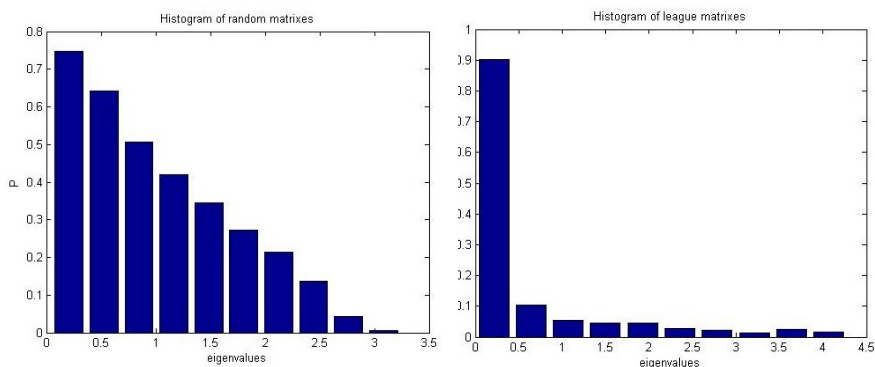
Wyk 5.5) Widma dla ligi piłkarskiej dla różnych miar: 1) punktowej oraz 2) bramkowej. Proszę nie doszukiwać się wartości na osi y, ponieważ słupki odpowiadają pojedynczym wartościom własnym na osi x, a wysokość słupka nie ma żadnego znaczenia (pozwala jedynie rozróżnić wartości miary punktowej-te wyższe od bramkowych-niższe)

Ponieważ na poziomie Ekstraklasy nie widać nic ciekawego, postanowiłem zbadać sezon pierwszej ligi, który obfitował w stwierdzone przez prokuraturę, sąd czy PZPN nieprawidłowości. Na [Wyk. 5.6] znajdziemy pierwszą anomalię. Na pierwszy rzut oka okazuje się, że może istnieć wewnętrzna relacja między zespołami ligi, która skutkowałaby pojawieniem się wartości własnej wystającej o 1/5 zakresu poza obszar przypadkowy.



Wyk. 5.6) Widmo pierwszej ligi wraz widmem macierzy losowej. Wartość własna 3,1 leży dosyć daleko od obszaru losowego.

Powracając do problemu krótkiej historii szeregu (tu w analizie pierwszej ligi mamy tylko 30 obserwacji, ale za to aż 16 zespołów) powraca pytanie o interpretację wyników. Aby powiększyć liczbę wartości własnych utworzyłem okno czasowe o szerokości 16 spotkań (najmniejsza liczba, żeby mieć tę samą lub większą liczbę stopni swobody, co danych) i po całym szeregu z tym oknem reprodukując nowe wartości własne. Dla macierzy przypadkowych również można zbierać większą ilość wartości własnych replikując macierze o ustalonym rozmiarze (30 spotkań). Dzięki temu możemy stworzyć histogramy widm i sprawdzić jakie jest prawdopodobieństwo anomalnej wartości z [Wyk. 3].



Wyk. 5.7) Unormowane histogramy wartości własnych. Gęstość macierzy przypadkowej (wykres lewy) maleje prawie że liniowo w przeciwieństwie do eksponentjalnego zaniku dla ligi piłkarskiej (wykres prawy). W związku z tym, że pole pod wykresami jest równe 1, prawdopodobieństwo wystąpienia wartości własnej rzędu 3-3,5 jest dla macierzy przypadkowej mniejsze od jednego promila, za to bardzo prawdopodobne dla ligi piłkarskiej.

Po wyluskania z anomalnej wartości własnej [Wyk. 5.6] wektora własnego jej odpowiadającego dowiadujemy się, że podstawowy wkład ma drużyna, która wysoko przegrywała mecze w takich stosunkach jak nawet 0:7. Dlatego anomalność sprowadza się w tym przypadku do bardzo słabych wyników tej drużyny a nie, jak zakładałem, do charakterystycznego dla procederu korupcyjnego wewnętrznego uporządkowania.

Podsumowując metodę macierzy przypadkowych muszę przyznać, że się na niej trochę zawiodłem. Brak żadnych jakościowych różnic (nie powodowanych wytłumaczalnym czynnikiem) między widmami ligowymi a stochastycznymi można zrzucić na karb krótkich szeregów czasowych. Jakkolwiek, wydaje mi się ważne dla całości pracy, aby powyższe rozumowanie przedstawić jako unikalne przedstawienie statystyki ligi piłkarskiej.

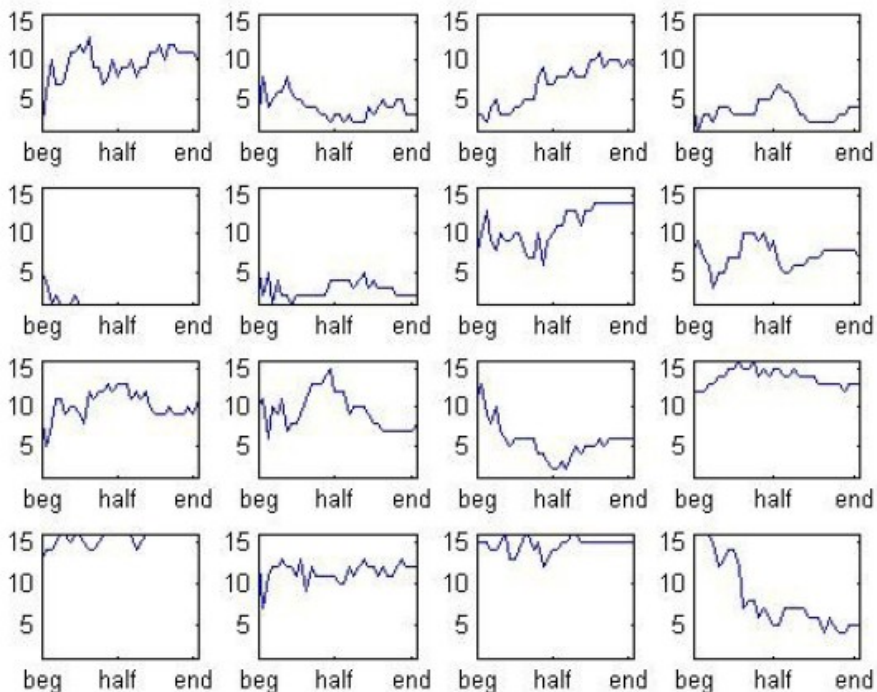
2b) Zmiana miejsca w rankingu

Pozycja w rankingu jest najważniejszym efektem rozgrywek ligowych (zwłaszcza na koniec sezonu), dlatego warto przyrzeć się dynamice zmian. Nasza intuicja skłania nas do zastanowienia się nad właściwościami układu w czasie i współzależnościami między drużynami. Do tego celu użyte zostały wyniki z sezonu 2003/2004 Polskiej Pierwszej Ligi, punktem startowym był [Tab. 4.2] stan po pierwszej kolejce (stan „zerowy”, czyli jeszcze przed rozpoczęciem rozgrywek jest listą w kolejności alfabetycznej).

1	Piast Gliwice
2	Cracovia
3	Jagiellonia Białystok
4	GKS Bełchatów
5	Pogoń Szczecin
6	Zagłębie Lubin
7	Tłoki Gorzyce
8	Podbeskidzie Bielsko-Biała
9	Ruch Chorzów
10	Łódzki KS
11	RKS Radomsko
12	Arka Gdynia
13	Aluminium Konin
14	Stasiak Opoczno
15	Polar Wrocław
16	Szczakowianka Jaworzno

Tab. 5.2) Ranking I Ligi po pierwszym spotkaniu sezonu 2003/2004.

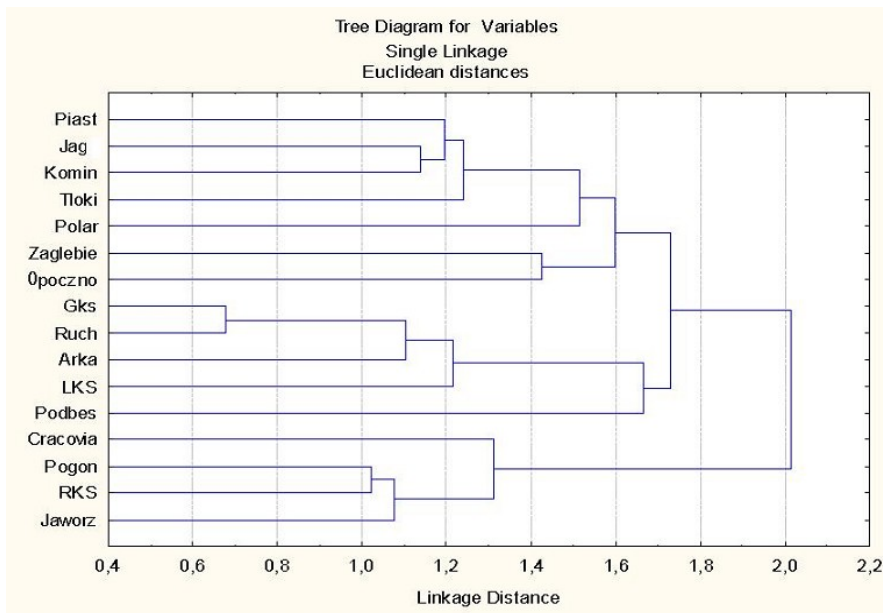
Dla tak ustalonego warunku początkowego [Tab. 5.3] możemy zbudować zależność czasową rankingu poszczególnych drużyn [Wyk. 5.8]. Oczywiście początek rozgrywek przynosi dużą rotację drużyn, gdyż ranking jest ustalany przede wszystkim na podstawie sumy zdobytych punktów. Z punktu widzenia analizy korupcyjnej najbardziej będzie nas interesować końcówka sezonu. Sprawia to wrażenie, że drużyny dążą do equilibrium odpowiadającemu ich mocy, z czego interesujące są fluktuacje, nad którymi zastanowimy się później.



Wyk. 5.8) Miejsce w tabeli od czasu (od początku do końca sezonu). Proszę zwrócić uwagę, że najlepsze miejsce w tabeli to 1 (najniżej na pojedynczym wykresie) a najgorsze 16 (najwyżej). Kolejność wykresów (od lewego górnego rogu idąc w prawo, przechodząc do następnej linii aż w końcu do prawego dolnego rogu) odpowiada kolejności drużyn z [Tab. 5.2], czyli od Piasta Gliwice potem Cracovii aż do Szczekowianki Jaworzno.

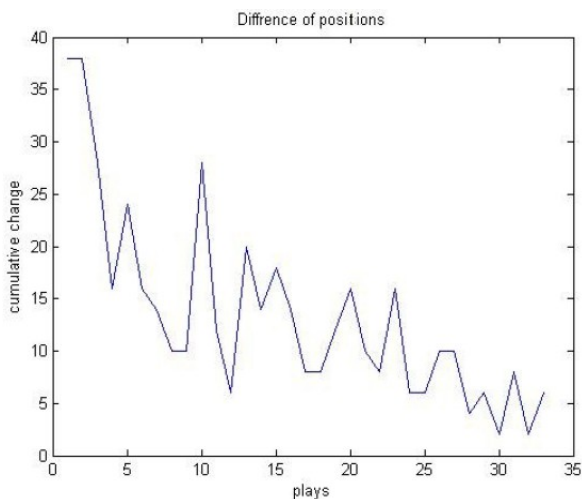
Indywidualne trajektorie [Wyk. 5.8] warto uzupełnić o miary współzależności. Do tego posłużą nam metoda tworzenia drzewa minimalnego zasięgu – MST (Minimal Spanning Tree), choć w celu prezentacji użyłem tożsamej reprezentacji graficznej – diagramu hierarchicznego. Dokładniejsze informacje można znaleźć w książce Rosario Mantegna [9] jak i we wstępie do niniejszej książki. Do zbudowania drzew potrzebne są szeregi czasowe pozycji w rankingu drużyn (a dokładnie zmiany miejsca w rankingu). Na ich podstawie budujemy macierz korelacji (podobną do tej z podrozdziału 4a) i przekształcamy ją w macierz odległości euklidesowych. Teraz nie zostaje nam nic innego jak przedstawić graficznie wyniki oraz ich interpretacja. Powyższa analiza pozwala na odnalezienie klastrów podmiotów zmieniających się podobnie w czasie, co wskazywałoby na ich współzależność. W przypadku drzewa hierarchicznego dla rankingu [Wyk. 5.9] możemy wskazać jedynie bliskość klubów sponsorowanych przez koncerty węglowe: Ruch Chorzów i GKS

Bełchatów.



Wyk. 5.9) Diagram hierarchiczny I Ligi sezonu 2003/2004 na podstawie rankingu drużyn. Etykietami są skrócone nazwy drużyn, a połączeniami euklidesowskie odległości między drużynami.

Jak już można było zauważyć na [Wyk. 5.8] w miarę upływu czasu pozycja drużyn względnie się stabilizowała. Tak naprawdę to stan na koniec sezonu jest obrazem zapamiętywanym przez kibiców i to co dzieje się w końcówce sezonu budzi najczęściej emocji. W naszym kraju zdarzały się tak zwane „kolejki cudów” kiedy, to drużyny walczące o utrzymania wygrywały mecze zapewniające im byt w dosyć kontrowersyjny sposób. Dlatego też na pierwszy rzut oka pod koniec sezonu, mogłaby wystąpić roszczenia miejsc, co skutkowałoby dużą fluktuacją w analizie rankingowej całego układu [Wyk. 5.9]. Nic takiego jednak nie obserwujemy. Można spekulować, że w układzie występuje dodatkowa siła skłaniająca drużyny do utrzymania swojej pozycji, co kompensowałoby nagły wzrost fluktuacji. Podstawą do takiej argumentacji jest brak obserwacji nagłego skoku fluktuacji, jak i zmiana sposobu myślenia o korupcji, o czym w następnym podrozdziale.



Wyk. 5.10) Zmierzenie układu do equilibrium wraz z fluktuacjami. Zależność skumulowanej zmiany pozycji drużyn w kolejce w funkcji czasu (następujące po sobie kolejki).

2c) Miara przewidywalności/anomalności wyniku.

Poprzednie próby analizy jakkolwiek dotykające nieliniowych zależności, bazują na współczynnikach korelacji liniowej (odpowiednio przetwarzanych). Idea nam przyświecającą jest możliwość oceny, czy analizowany wynik spotkania jest prawdopodobny (w jakim sposób). Do tego wykorzystamy różnicę mocy drużyn. Aby uniknąć żmudnego estymowania parametrów modelu Kopocińskiego (przypomnę, że dla najprostszego układu parametrów jest więcej niż meczy-punktów pomiarowych), za różnice mocy weźmiemy różnicę pozycji w rankingu. To jest też zasadnicza różnica wobec modelu Heuera, który za moc drużyny brał różnice bramek strzelonych i straconych przez drużynę. Dalej uwzględniona została przewaga gry na własnym boisku. Porównując wyniki Heurera (drużyna u siebie strzela o 0,7 bramek więcej) czy Kopocińskiego (drużyna grająca u siebie w ponad 68% przypadków nie przegra na własnym boisku) dokonaliśmy przeskalowania różnicy mocy drużyn tak, że grając u siebie ma się 22% więcej szans na nieprzegranie meczu (wynik obserwacji empirycznych). Efektem jest miara przewidywalności spotkania. Od razu należy wyjaśnić, że celem tej pracy w żadnym wypadku nie miało zbliżyć się do prognozowania wyniku meczu. Pokazany algorytm może wydawać się takim narzędziem, ale był on stworzony z myślą o wyznajdowaniu anomalii i tak należy go rozumieć.

Pomiar anomalności: wyniki, które miały niskie prawdopodobieństwo wystąpienia, zostaną uchwycone przez przeskalowaną różnicę miejsc w rankingu. Skalowanie opiera się na apriorycznym prawdopodobieństwie przewagi gry u siebie. Ujemny współczynnik oznacza, że wynik można było przewidzieć, a dodatni że wynik był niespodziewany. Algorytm wyliczania współczynnika anomalności przedstawia

[Schem. 1] razem z opisem:

$A < 1 < H$ - ten warunek musi być spełniony, aby skalowanie miało sens

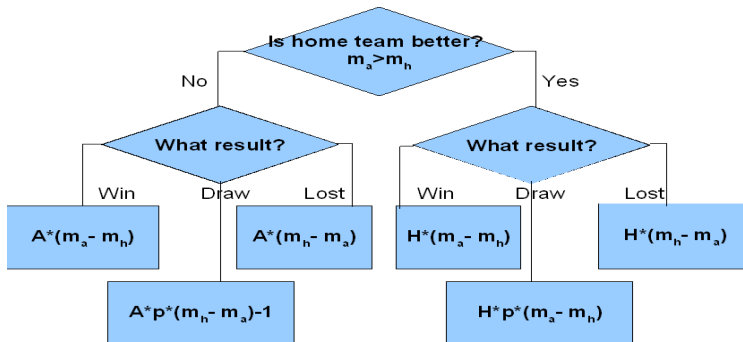
H- aprioryczna przewaga gry u siebie (u nas 1,22)

A- aprioryczna trudność gry na wyjeździe (u nas 0,78)

p- aprioryczne prawdopodobieństwo remisu (u nas 0,326, z obserwacji empirycznych)

m_a - miejsce w tabeli drużyny grającej u siebie

m_b - miejsce w tabeli drużyny grającej na wyjeździe



Sch. 5.1) Sposób wyznaczania współczynnika przewidywalności.

Obliczając współczynnik przewidywalności/anomalności możemy brać aktualny (to znaczy przed badanym meczem) ranking, bądź też korzystać przez cały sezon z rankingu końcowego (mamy taką szansę, ponieważ badamy historyczne rozgrywki). Dlatego też możemy mówić o dynamicznym, bądź stacjonarnym współczynniku, z czego do dalszej analizy będziemy używać tego drugiego. Dodatkowo aby otrzymać wartość z przedziału $[-1;1]$ normalizujemy współczynnik, dzieląc go przez liczbę drużyn w lidze. Z tak wyznaczonymi wskaźnikami możemy rozpocząć poszukiwania nietypowych wyników. Najprostszym sposobem jest analiza wszystkich wyników w lidze i odnalezienie tych z wysokim współczynnikiem anomalności. Można też podejść do problemu od drugiej strony i mając podejrzenie o niesportowe zachowanie w jakichś spotkaniach - wyznaczyć ów współczynnik.

Piast Gliwice	Arka Gdynia	Podbeskidzie Bielsko-Biala	Jagiellonia Bialystok
-0,08	-0,44	0,07	0,12
0,04	0,02	-0,08	0,02
-0,16	-0,12	-0,32	0,28
-0,12	0,12	-0,2	-0,03
-0,03	0,05	-0,16	0,24
-0,32	0,1	-0,04	0,08
-0,28	-0,48	0,05	-0,2
-0,12	0	0,08	-0,32
0,05	0,03	-0,12	0,07
0,1	-0,03	-0,16	0,04
-0,24	0,51	0,2	-0,2
-0,08	0,13	0,12	-0,16
-0,16	-0,03	0,05	0,24
-0,2	-0,02	0,57	0
-0,2	0,07	-0,03	-0,02
-0,08	-0,44	-0,28	0,19
0,06	0,02	-0,08	0,02
0,25	0	-0,32	0,07
0	0,12	0,03	-0,04
-0,04	-0,08	0,02	0,05
0,08	-0,24	-0,04	0,13
0,44	-0,36	-0,24	-0,2
0,19	0,15	0,13	-0,32
-0,24	0	0,19	0,44
0,57	0,03	-0,16	-0,03
0,05	-0,03	0,32	0,03
-0,08	-0,32	0,19	0,25
-0,16	-0,03	-0,24	-0,24
0,03	-0,08	-0,36	0

Tab. 5.3) Współczynniki anomalności w szeregu czasowym dla kilku drużyn ukaranych za korupcję. Poczernione wartości oznaczają mało prawdopodobne wyniki.

Miarę anomalności/przewidywalności zastosowałem również dla kilku sezonów Bundesligi (przyjmując niemiecką ligę jako wzór czystej, sportowej rywalizacji). Dla przykładowego sezonu 2007/2008 zobrazowaliśmy macierz współczynników [Tab. 5.4]. Taka prezentacja w postaci stacjonarnej macierzy niesie za sobą utratę informacji o zależnościach czasowych. Nie będzie to jednak przeszkadzało w statystycznych porównaniach ligi polskiej z niemiecką.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1: VfB Stuttgart	0	0,06	-0,11	-0,17	-0,22	0,09	-0,33	-0,39	0,44	0,16	-0,56	-0,61	-0,67	-0,72	0,25	-0,83	-0,89	0,31
2: FC Bayern München	-0,04	0	0,02	0,11	0,17	0,07	0,09	-0,33	0,13	0,14	0,16	-0,56	-0,61	0,22	0,24	-0,78	0,27	0,89
3: Bayer 04 Leverkusen	-0,07	-0,04	0	-0,06	-0,11	-0,17	0,07	0,09	0,33	-0,39	-0,44	-0,5	0,18	-0,61	0,67	-0,72	0,78	0,83
4: 1. FC Nürnberg	0,11	0,07	0,04	0	-0,06	-0,11	-0,17	0,07	-0,28	0,33	-0,39	-0,44	0,16	0,18	-0,61	0,67	-0,72	-0,78
5: VfL Bochum	0,01	-0,14	-0,07	0,04	0	-0,06	-0,11	-0,17	-0,22	-0,28	-0,33	-0,39	0,14	-0,5	-0,56	-0,61	-0,67	-0,72
6: Borussia Dortmund	0,01	-0,14	-0,11	-0,07	-0,04	0	0,02	-0,11	0,17	0,07	-0,28	0,11	0,13	-0,44	0,16	-0,56	-0,61	0,67
7: Hannover 96	-0,21	0,01	-0,14	-0,01	-0,07	0,04	0	0,06	0,11	-0,17	-0,22	0,28	0,11	0,13	0,44	-0,5	-0,56	-0,61
8: Eintracht Frankfurt	0,25	0,21	0,18	0	-0,01	0,07	0,04	0	0,06	0,04	0,05	0,22	0,09	0,11	0,39	0,44	0,5	-0,56
9: VfL Wolfsburg	-0,28	0,04	0,02	-0,18	0	0,11	0,07	-0,03	0	0,06	-0,11	0,17	0,07	-0,28	0,33	-0,39	0,14	0,5
10: FC Schalke 04	0,32	-0,28	0,25	0,21	0,01	0,14	-0,01	-0,02	0,04	0	-0,06	-0,11	0,05	-0,22	0,28	0,11	0,13	0,14
11: Werder Bremen	-0,36	0,06	0,28	0,04	-0,21	0,18	0,14	0,11	-0,07	-0,03	0	0,06	0,11	-0,17	0,07	0,28	-0,33	0,13
12: Hamburger SV	0,39	0,07	-0,32	-0,28	-0,25	0,21	0,18	0	-0,11	-0,02	-0,04	0	0,06	0,11	0,05	0,22	-0,28	-0,33
13: Hertha BSC Berlin	0,43	0,08	0,36	-0,32	-0,28	0,25	0,02	0,01	0,14	-0,01	-0,07	0,04	0	0,06	0,11	0,05	0,22	0,28
14: Arminia Bielefeld	-0,46	0,43	-0,39	0,07	0,32	0,28	-0,25	0,21	0,18	0,14	-0,11	0,07	0,04	0	0,06	0,11	0,05	0,07
15: FC Energie Cottbus	0,5	-0,46	0,43	-0,39	-0,36	0,32	0,28	0,25	0,21	0,18	0	0,11	-0,07	0,04	0	0,06	-0,11	-0,17
16: Karlsruher SC	0,53	0,5	-0,46	-0,43	0,39	0,36	0,06	0,28	0,04	0,02	-0,18	0	0,11	-0,07	-0,04	0	0,02	0,04
17: FC Hansa Rostock	0,57	0,13	-0,12	-0,46	-0,43	-0,39	0,36	0,06	0,28	0,25	0,21	0,18	-0,14	-0,01	0,07	0,04	0	0,06
18: MSV Duisburg	-0,6	0,57	-0,53	-0,5	0,46	0,43	-0,39	-0,36	0,32	-0,28	-0,25	0,02	0,18	-0,14	-0,11	0,07	0,04	0

Tab. 5.4) Stacjonarna macierz współczynników przewidywalności/anomalności dla Bundesligi sezonu 2007/2008

Porównując trzeci kwantyl rozkładu empirycznego wartości współczynnika (interesują nas największe jego wartości) otrzymamy:

- 0,14 - niemiecka liga

•0,12 - polska liga

Balansując na granicy statystycznej istotności możemy wnioskować, że niemiecka liga jest bardziej nieprzewidywalna (jest więcej wysokich wartości współczynnika anomalności). Niestety, gdybyśmy postawili hipotezę o równości trzecich kwantyli, to dla małego przedziału ufności, nie mielibyśmy podstaw do odrzucenia tej hipotezy.

W tym miejscu warto wrócić do hipotezy zerowej z początku pracy, a zwłaszcza drugiej jej części, dotyczącej anomalii. Może warto było by pomyśleć o postanowieniu alternatywnej, która brzmiałaby :

(...)Wyniki spotkań powinny być mniej przypadkowe, niż by to wynikało ze stochastycznego wzorca.

Czy jak to zostało zaprezentowane jako hipoteza alternatywna:

(...)Anomalie powinny występować jednak rzadziej, niż by to wynikało z stochastycznego wzorca ligi.

Powyższe stwierdzenie świadczyło by o przewadze zjawisk korupcyjnych, wynikających z dążenia do zachowania równowagi – zespoły zadowolone ze swojego miejsca w tabeli mogłyby używać wszystkich dostępnych środków, aby zachować swoją pozycję.

2d) Quasi-entropowa analiza przewidywalności/anomalności

W poprzednim podrozdziale wprowadzona została miara przewidywalności/anomalności dla wyników poszczególnych spotkań. Proponuję teraz podejście ujmujące ligę jako całość. Do tego celu wykorzystam quasi-entropię Shanona:

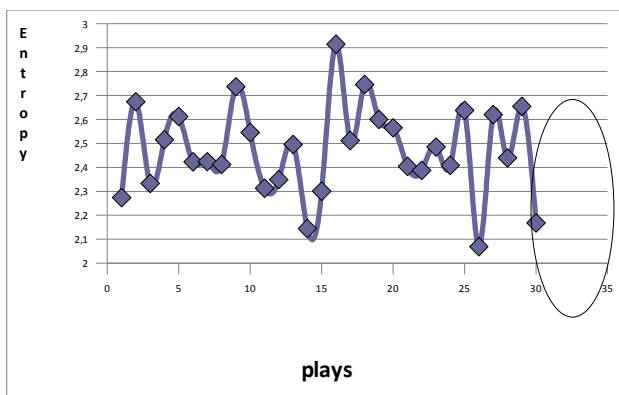
$$S_q = - \sum p_i \ln(p_i) \quad (5.4)$$

gdzie:

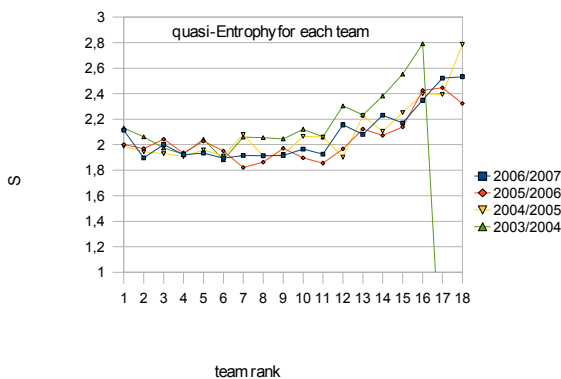
p_i – jest wartością współczynnika przewidywalności/anomalności dla pojedynczego spotkania przeskalowaną do miary prawdopodobieństwa $[0;1]$, tak że przewidywalny wynik byłby bliski 0, a anomalia blisko 1 ;

S_q – jest quasi-entropią Shanona określonej na pewnym zbiorze współczynników p_i , czy czym w ogólności p_i nie sumują się do jedynki (dlatego przedrostek quasi)

Rozpatrując zależność S_q od czasu, a dokładniej od kolejek (tak, że S_q jest każdorazowo określone na wszystkich p_i z jednej kolejki) w sezonie otrzymamy [Wyk. 5.11]. Kontynuując rozumowanie z części teoretycznej pracy przyjrzyjmy się końcówce sezonu (zaznaczonej na wykresie). Obserwujemy w badanym obszarze delikatny wzrost wahań. Nie widać też, żeby układ dążył do uporządkowania (funkcja S_q nie maleje). Ciekawsze obserwacje możemy uzyskać budując S_q dla poszczególnych drużyn [Wyk. 5.12] w taki sposób, że miejsce rankingu szereguje S_q . Dla różnych sezonów polskiej I Ligi widać, że dla drużyn z końcówki tabeli quasi-entropia jest dużo większa. Wskazuje to na łamanie symetrii między drużynami z początku tabeli a tymi z końca i mniej równomierny rozkład współczynników przewidywalności/anomalności drużyn z ostatnich miejsc w rankingu- w przeciwieństwie do reszty drużyn.



Wyk. 5.11) Zależność quasi-entropowa do kolejek dla Polskiej I Ligi sezonu 2003/2004 z zaznaczonym obszarem pod koniec sezonu.



Wyk. 5.12) Quasi-entropia dla różnych sezonów Polskiej I Ligi wyznaczana dla każdej z drużyn (na osi rzędnych miejsce w rankingu badanej drużyny).

5) Podsumowanie

Piłka nożna od dawna ma spore znaczenie w życiu wielu mieszkańców naszej planety- również w Polsce. W związku z dużym zainteresowaniem, pieniądze odgrywają bardzo dużą rolę w kształtowaniu rozgrywek na wszystkich poziomach od lig regionalnych aż po zawody międzynarodowe. Nawiązując do problematyki reglamentacji pozycji w lidze i różnych możliwości finansowych korupcja może być pewniejszą inwestycją niż np. zakup nowych piłkarzy. Istnieją modele wyjaśniające pojawianie się korupcji, w związku z nadwyżką pieniędzy w

stosunku do ograniczonych zasobów strukturalnych [10]. Tłumacząc zjawisko makroekonomiczne na sportowe wystarczy zauważyć, że dystrybucja miejsc w tabeli jest ograniczona, a zasoby finansowe drużyn mogą być teoretycznie o rzędy wielkości większe. Istnieje oczywiście kwestia uczciwości czy wykrywalności przestępstw. Moje podejście ilościowe wykorzystuje informacje jedynie o wynikach spotkań, w związku z tym może stanowić tylko bazę do późniejszych, bardziej dogłębnych analiz. Podjęte próby zasygnalizowania anomalnych zachowań mogą być wyjaśnione na wiele sposobów. Dla przykładu odstająca wartość własna [Wyk. 5.6] związana jest z „hokejowymi” wynikami jednej z drużyn (np. 7:0). Bliskość Ruchu Chorzów z GKS Bełchatów na drzewie minimalnego zasięgu [Wyk. 5.9] wynikać może z powiązania finansowego (sponsorami drużyn były spółki górnicze). Z drugiej strony pojawiły się wyniki nieintuicyjne, jak wysoka wartość quasi-entropii [Wyk. 5.12] dla drużyn z dołu tabeli. Pochodną przeprowadzonych analiz jest w związku z tym szczegółowa analiza ligi piłkarskiej jako takiej. Nowatorskie metody w tym obszarze jak: macierze przypadkowe, analiza rankingowa, współczynniki przewidywalności/anomalności, czy quasi-entropia rzucają nowe światło na dosyć dobrze opisany już przez statystyków obiekt.

Metodologia badania ilościowego zawęziła obszar analizy do samych wyników spotkań. Pierwszym naturalnym rozwinięciem byłoby dodanie sędziów i szukaniu związku z pojawieniem się anomalii dla poszczególnych nazwisk. Należy zauważyć, że to działanie nie naruszyłoby warunku obiektywności badania. W związku z tym nasuwa się wykorzystanie współczynnika przewidywalności/anomalności oraz quasi-entropii nie tylko dla drużyn podejrzanych o korupcję, ale również dla podejrzanych sędziów, co mogłoby być częścią materiału dowodowego w postaci ekspertyzy w prowadzonym śledztwie. Ponadto wynik porównania kwantyli rozkładu współczynnika przewidywalności/anomalności dla ligi polskiej i niemieckiej daje również dużo do myślenia nad samym zjawiskiem korupcyjnym. Co prawda, nie jesteśmy w stanie podać wniosków istotnych statystycznie, ale nasza intuicja wskazuje na przewagę zachowań korupcyjnych, związanych z zachowaniem ustalonego porządku. Początkowo sądziłem, że jest odwrotnie i całe badanie zostało skonstruowane pod kątem poszukiwania anomalii. Powyższy wniosek można bardzo łatwo przenieść na inne sfery życia. Gdy przyjrzymy się, dla przykładu, prywatyzacji przedsiębiorstw w Polsce w okresie transformacji, zauważymy, że najczęściej nowymi właścicielami zostali byli dyrektorzy i inni działacze PZPR (nie mylić z Polskim Związkiem Piłki Ręcznej), co świadczyć może o korupcyjnych tendencjach do zachowania równowagi (odpowiednik fizycznej reguły Lenza). Przykłady potwierdzające alternatywną hipotezę korupcyjną można by mnożyć w nieskończoność, ale kolejnym punktem, aż proszącym się o rozwinięcie, jest teoria gier. Niestety wiąże się to z utratą obiektywnych miar, ale pozwala na wykorzystanie czynników społecznoekonomicznych, przeze mnie pominiętych. Pierwszym krokiem mogłoby być przesymulowanie rozgrywek ligowych, działających zgodnie z tabelą wypłat i porównanie wyników z rzeczywistymi ligami. Następnie, można byłoby się

pokusić o próbę oszacowania różnych czynnik społecznoekonomicznych na przebieg rozgrywek. Dlatego też powyższa praca jest cegiełką w budowie aparatu wyjaśniającego mechanizm korupcji i ogólnych właściwości lig piłkarskich.

Bibliografia

- [1] Wesson, J. *The Science of Soccer*, IoP, Bristol 2002
- [2] Kopociński, B. Components of the game result in a football league, *Appl. Math.* **28** (2001), 1-18.
- [3] Kopociński, B. Mathematical modelling of results of a football league, *Human Movement* **1**(3) (2001), 129-134.
- [4] Heuer, A. et al. *Soccer: is scoring goals a predictable poissonian process?* *Europhys. Lett* **89**, 38007 (2010)
- [5] Heuer, A. et al. *Fitness, chance and myths: an objective view on soccer results* *Eur. Phys. J. B* **67**, 445-458 (2009)
- [6] Steinhaus, H. *Myśl akademicka*, Lwów 1925
- [7] Strzelecki T. *Przyczyny korupcji. Studium krajów postkomunistycznych w okresie transformacji.* „Korupcja. Oblicza, uwarunkowania, przeciwdziałanie.”, Ossolineum, Wrocław 2006
- [8] Burda, Z. et al. Signal and noise in correlation matrix, *Physica A* **343** (1-4), pp. 295-310 (2004)
- [9] Mantenga, R., Stanley, E. *Introduction to Econophysics*, Cambridge University Press, Cambridge 2007
- [10] Burda, Z. et al. Wealth condensation in pareto macroeconomics *Physical Review E* **65** (2), art. no. 026102, pp. 026102/1-026102/4 (2002)