



**HAL**  
open science

# Incertitude de classement final et affluence en Ligue 1 française de football : une nouvelle approche

Jean-Pascal Guironnet

► **To cite this version:**

Jean-Pascal Guironnet. Incertitude de classement final et affluence en Ligue 1 française de football : une nouvelle approche. *Revue d'économie politique*, 2018, 128 (4), pp.641-666. 10.3917/redp.284.0641 . halshs-02064147

**HAL Id: halshs-02064147**

**<https://shs.hal.science/halshs-02064147>**

Submitted on 11 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# “Incertitude de classement final et affluence en Ligue 1 française de football : une nouvelle approche”

Jean-Pascal Guirommet<sup>1</sup>

Cet article propose une nouvelle méthode pour évaluer les effets de l’incertitude d’une rencontre et de classement final sur l’affluence dans les stades de Ligue 1. L’issue incertaine d’un match est évaluée à partir des différences relatives de qualité des équipes. Contrairement aux précédentes études, ces forces intrinsèques sont estimées de manière aléatoire à partir des différences de budget salarial des clubs et elles sont, par la suite, réévaluées en fonction des résultats observés durant la saison. En outre, l’incertitude de classement final est ici modélisée dans un espace de probabilités permettant d’avoir une information statistique plus complète que des indicateurs agrégés. Suivant cette approche, trois résultats principaux émergent : (i) l’incertitude d’une rencontre présente un effet significatif non-linéaire et positif sur l’affluence dans les stades, (ii) les spectateurs ont des demandes antagonistes en espérant que leur club favori ne soit pas rejoint par des poursuivants durant la saison mais paradoxalement ils souhaitent une incertitude plus élevée concernant le classement final, (iii) l’intérêt du spectateur n’est pas linéairement décroissant en fonction de la hiérarchie du classement final mais de forme sinusoïdale avec un engouement plus important concernant l’attribution des places qualificatives en *Europa League* et pour les équipes luttant contre la relégation. L’intérêt pour la victoire finale peut s’avérer prépondérant pour l’attractivité globale du championnat, mais ce dernier est devenu plus volatile avec les nouvelles capacités financières du PSG.

*Classification-JEL: C11, D40, Z28*

**Mots clés : Incertitude, Calcul Bayésien, Football, Affluence**

---

<sup>1</sup>Université de Normandie (UCBN), CREM (CNRS), e-mail : jean-pascal.guirommet@unicaen.fr.

# “Championship Uncertainty and Soccer Attendance in French Ligue 1: A New Approach”

This study proposes a new method for estimating supporters' attendance in considering uncertainty about the games results and the final ranking of teams. Outcome uncertainty is expressed by some probabilities, which is determined according to the gap between team qualities. In this turn, the qualities of teams are randomly estimated by average wages of their players. However, unlike previous studies, these probabilities are further adjusted according to the past results of these specific teams during the championship. Moreover, instead of relying on an aggregate indicator to estimate the uncertainty about the final ranking, we estimate the probability of each possible final ranking. This procedure allows us to make use of a more complete information statistics in estimation than usually done in this kind of studies. We provide three main results. First, the effect that uncertainty about the match outcome has on supporters' attendance is positive, significant and non-linear. Second, concerning the final ranking, we identify two opposite effects on supporter's behaviour. While their attendance is inversely related to the gap between the position of their preferred team and its followers in the ongoing ranking, their attendance is also higher if the gap between each teams in the final ranking is sufficiently low. Third, supporters' attendance is not linearly decreasing with the position of the supported team in the championship, but this link presents a sinusoidal curve. The reason is that supporters' attendance also depends on the likelihood of their team to accede to Europea League during the next season, and also on the likelihood of their team being drawn out of "Ligue 1" at the end of the season. We also show that supporters' attendance is directly proportional to the uncertainty level about the championship final victory. In particular, the attractiveness of the championship is negatively affected by the recent financial power of PSG.

**Keywords : Uncertainty, Bayesian Approach, Football, Attendance**

# 1 Introduction

La Ligue de Football Professionnel (LFP) a enregistré sur ces cinq dernières années une affluence de spectateurs qui se stabilise désormais à un total d'environ 7,5 millions d'individus dans les stades par saison. Afin d'éclairer les responsables du football professionnel en matière d'attractivité du championnat français, l'objet de cet article est d'analyser l'intérêt que porte le spectateur à l'incertitude de classement final. L'étude de ce lien revient à appréhender la structure compétitive du championnat et ainsi mieux comprendre l'impact du format de cette compétition, et de ses éventuelles modifications, sur la demande. Suivant cet objectif, cette étude est réalisée sur deux récentes saisons, 2014-15 et 2015-16, de première division (Ligue 1).

Le championnat de France est composé de vingt équipes professionnelles qui s'affrontent en match aller-retour, soit un total de 380 rencontres de football réparties en 38 journées pour une saison complète. Le classement s'effectue en attribuant à chaque équipe trois points lors d'une victoire, un point pour un match nul et aucun point en cas de défaite. Le championnat est organisé comme une ligue ouverte, avec la relégation<sup>2</sup> en deuxième division (Ligue 2) des trois dernières équipes du classement. Les quatre premières places, ainsi que les vainqueurs de la Coupe de la Ligue (CdL) et de la Coupe de France (CdF), permettent de participer la saison suivante aux compétitions européennes. Les deux premières positions du classement donnent accès au premier tour de la Ligue des champions alors que l'équipe terminant troisième accède aux tours préliminaires de cette compétition. La quatrième place permet de participer à la *Ligue Europa*, les clubs terminant cinquième, ou sixième, peuvent être également qualifiés pour cette coupe d'Europe, si les vainqueurs de la CdL et de la CdF figurent parmi les quatre premiers du championnat.

Afin d'évaluer l'effet de l'incertitude de classement final (variant durant la saison), il est au préalable nécessaire de contrôler et d'identifier les autres facteurs pouvant potentiellement influencer les choix du spectateur. La relation entre aléa du résultat d'une rencontre et affluence de spectateurs a suscité un développement fécond depuis les articles

---

<sup>2</sup>Depuis 2016-17, un système de barrages a été introduit dans le format de cette compétition : les équipes terminant aux deux dernières positions sont automatiquement reléguées alors que le 18<sup>ième</sup> doit effectuer un barrage contre le troisième de Ligue 2.

fondateurs de Rottenberg [1956] et Neale [1964]. L'influence de cette incertitude sur la demande de spectateurs est ainsi un facteur probablement prépondérant qu'il est nécessaire d'approfondir. L'idée principale de ces travaux est qu'un championnat dominé par un certain nombre d'équipes véhicule un intérêt moindre qu'une compétition composée d'équipes de force similaire.

Les travaux sur ce type d'incertitude sont nombreux et ils raisonnent sur des différences relatives entre les deux équipes de la rencontre : par exemple, le différentiel de buts (Falter et Perignon [2000]), la différence de points au classement (Scelles *et al.* [2013] ; Scelles, *et al.* [2016]). Relativement à cette littérature, notre travail suggère une nouvelle modélisation de l'issue possible d'un match. L'idée est de proposer non seulement un pari sur une rencontre de football (Peel et Thomas [1992]) mais également d'envisager que celui-ci est incertain. Cette approche revient alors à considérer le calcul bayésien comme outil de modélisation de l'incertitude (Parent et Bernier [2007]). Sur la base de ce raisonnement, nous construisons un indicateur des attentes du spectateur, avant le début de la compétition, concernant l'issue finale du championnat. Ces croyances initiales sont par la suite réévaluées en fonction des résultats observés durant la saison et de la forme "du moment" des équipes. Cette révision est probablement aussi réalisée par les spectateurs et ce travail permet, dans une certaine mesure, d'appréhender l'évolution du raisonnement des individus durant la saison. Dans cette démarche, les performances à domicile et à l'extérieur des équipes sont également considérées dans l'estimation de l'incertitude d'une rencontre.

Après avoir déterminé l'ensemble des probabilités caractérisant une rencontre de Ligue 1, une espérance de points est calculée pour l'ensemble des équipes et cette dernière est réévaluée après chaque journée de championnat. A partir de l'évolution de cette espérance de points et du nombre de points effectivement attribués, nous déterminons l'évolution de l'incertitude de classement final au cours de la saison. A ce niveau-là, certaines études utilisent des mesures comme les indices de concurrence (Herfindahl, *cf.* Brandes et Franck [2007]), d'inégalités (Gini, indices d'entropie... *cf.* Gayant et Le Pape [2015] et [2016]) ou la mesure de Kuypers<sup>3</sup> [1996]. Toutefois, comme tout indicateur statistique, ces mesures

---

<sup>3</sup>Définie comme le produit entre le nombre de matchs restant à jouer et le nombre de points de retard

synthétisent l'information et elles n'exploitent pas pleinement toutes les caractéristiques de l'incertitude<sup>4</sup> : provient-elle d'un risque de se faire rattraper par un poursuivant et/ou de la probabilité de dépasser un concurrent ? Ces indicateurs ne peuvent, dès lors, pas être utilisés pour appréhender complètement l'incertitude de classement final. Comme le souligne Andreff [2009], il est alors nécessaire d'adopter en économie du sport une démarche probabiliste.

La détermination de ces deux types d'incertitude, portant sur une rencontre et sur le classement, va permettre d'identifier les potentiels liens de ces facteurs avec l'affluence observée dans les stades de Ligue 1. Cet article s'articule comme il suit : la section 2 présente, dans un environnement déterministe, notre raisonnement probabiliste et les calculs d'incertitude de classement final ; la section 3 introduit à partir d'un raisonnement Bayésien un aléa sur le calcul de ces probabilités subjectives ; la section 4 présente notre travail d'inférence Bayésienne et nos choix économétriques ; la dernière section discute, avant de conclure, les résultats obtenus suivant cette approche.

## 2 Probabilités de victoire & incertitude de classement final

Le raisonnement probabiliste, présenté dans cette section, s'inspire en grande partie des travaux de Parent et Bernier [2007]. Supposons que nous connaissons la qualité intrinsèque des équipes, notée  $q_i$ . Lorsque l'équipe  $i$  de force  $q_i$  affronte l'équipe  $j$  de force  $q_j$ , notre ensemble de probabilités, ou plus simplement le résultat d'une rencontre, est alors principalement gouverné par l'écart<sup>5</sup>  $\delta_{ij} = q_i - q_j$ . Notons les probabilités de victoire, nul et défaite de l'équipe  $i$  respectivement de la manière suivante  $(p_{ij}^+ ; p_{ij}^- ; p_{ij}^-)$ .

Plus l'écart  $\delta_{ij}$  est grand (et positif) plus l'équipe  $i$  est censée remporter le match et, dans ce cas, le rapport de la probabilité de victoire de  $i$  sur la probabilité de faire un

---

sur le *leader*.

<sup>4</sup>Le principe de ces indices repose, en outre, sur le calcul d'écart par rapport à un indicateur de tendance centrale. Lorsqu'une équipe est championne, ou qu'elle ne peut plus rattraper ou se faire rejoindre par d'autres équipes, la situation devient certaine et l'incertitude prend la valeur nulle. Or, dans le calcul de l'ensemble de ces indicateurs, la valeur nulle n'est pas considérée comme une situation certaine mais elle accroît la valeur de l'indice sélectionné pour mesurer l'incertitude.

<sup>5</sup>Le potentiel avantage de jouer à domicile pour l'équipe  $i$  sera considéré ultérieurement.

match nul  $\frac{p_{ij}^+}{p_{ij}^-}$  doit être important, comme le rapport  $\frac{p_{ij}^-}{p_{ij}^+}$ . En conséquence, il doit exister deux fonctions croissantes de  $\delta_{ij}$  telles que :

$$\begin{cases} \frac{p_{ij}^+}{p_{ij}^-} = f_1(\delta_{ij}) \\ \frac{p_{ij}^-}{p_{ij}^+} = f_2(\delta_{ij}) \end{cases} \quad (1)$$

Supposons que la probabilité de gagner ou de faire match nul est proportionnelle à l'écart  $\delta_{ij}$ . Les fonctions dans (1) doivent donc être de la forme  $f_1(\delta_{ij}) = k_1 f(\delta_{ij})$  et  $f_2(\delta_{ij}) = k_2 f(\delta_{ij})$ , soit :

$$\begin{cases} \frac{p_{ij}^+}{p_{ij}^-} = k_1 f(\delta_{ij}) \\ \frac{p_{ij}^-}{p_{ij}^+} = k_2 f(\delta_{ij}) \\ p_{ij}^+ + p_{ij}^- + p_{ij}^0 = 1 \end{cases} \quad (2)$$

Le système (2) implique que  $\frac{p_{ij}^+}{p_{ij}^-} + 1 + \frac{p_{ij}^-}{p_{ij}^+} = \frac{1}{p_{ij}^0}$ , soit  $k_1 f(\delta_{ij}) + 1 + \frac{1}{k_2 f(\delta_{ij})} = \frac{1}{p_{ij}^0}$ . Autrement dit :

$$\begin{cases} p_{ij}^-(\delta_{ij}) = \left(\frac{1}{k_2 f(\delta_{ij})}\right) / \left[\frac{1}{k_2 f(\delta_{ij})} + 1 + k_1 f(\delta_{ij})\right] \\ p_{ij}^0(\delta_{ij}) = 1 / \left[\frac{1}{k_2 f(\delta_{ij})} + 1 + k_1 f(\delta_{ij})\right] \\ p_{ij}^+(\delta_{ij}) = k_1 f(\delta_{ij}) / \left[\frac{1}{k_2 f(\delta_{ij})} + 1 + k_1 f(\delta_{ij})\right] \end{cases} \quad (3)$$

Pour la rencontre entre  $i$  et  $j$ , la fonction  $f$  doit vérifier pour tout  $\delta_{ij}$ :

$$\begin{cases} p_{ij}^-(-\delta_{ij}) = p_{ij}^+(\delta_{ij}) \\ p_{ij}^0(\delta_{ij}) = p_{ij}^0(-\delta_{ij}) \end{cases} \quad (4)$$

D'après (4), il est ainsi nécessaire de vérifier l'égalité suivante :  $k_1 f(\delta_{ij}) + \frac{1}{k_2 f(\delta_{ij})} = k_1 f(-\delta_{ij}) + \frac{1}{k_2 f(-\delta_{ij})}$ .

Une solution simple vérifiant cette condition s'obtient en prenant la fonction  $f(\delta_{ij}) = \exp(\delta_{ij})$  et les constantes telles que  $k_1 * k_2 = 1$ , soit  $k_1 = \frac{1}{k_2} = k$ .

Au final, nous obtenons ainsi le système suivant :

$$\begin{cases} p_{ij}^+(\delta_{ij}) = k \exp(\delta_{ij}) / [k \exp(-\delta_{ij}) + 1 + k \exp(\delta_{ij})] \\ p_{ij}^{\bar{0}}(\delta_{ij}) = 1 / [k \exp(-\delta_{ij}) + 1 + k \exp(\delta_{ij})] \\ p_{ij}^-(\delta_{ij}) = k \exp(-\delta_{ij}) / [k \exp(-\delta_{ij}) + 1 + k \exp(\delta_{ij})] \end{cases} \quad (5)$$

$k$  est un paramètre lié à  $\delta_{ij}$ , son rôle est de diminuer la probabilité de faire match nul lorsque la différence de qualités des équipes est suffisamment importante pour permettre de les départager lors de la rencontre (Rao et Kupper [1967]).  $k$  est ainsi proportionnel à la volonté des équipes<sup>6</sup> à vouloir créer des différences pour remporter la confrontation.

Il existe des variantes de la modélisation des probabilités, décrites à partir des relations du système (5). À titre d'exemple, on peut citer le populaire modèle de Bradley et Terry [1952], qui a suscité de féconds développements (*cf.* Davidson [1970]). L'ensemble de ces travaux modélise, comme notre approche, une expression des probabilités déterminées à partir de la fonction  $\exp(\cdot)$ . Ce choix a l'avantage de faire le lien avec l'utilisation des distributions statistiques appartenant à la famille des exponentielles pour estimer les probabilités (*cf.* section 4). Une différence, par rapport au modèle de Bradley et Terry [1952], est que notre paramètre  $k$  intervient de manière similaire dans les expressions du système (5). Cette approche facilite ainsi l'estimation des probabilités de la section 4, puisque celle-ci revient dès lors à simplement estimer la quantité  $k \exp(\delta_{ij})$ . L'objet de notre étude n'étant pas de déterminer un ensemble de valeurs crédibles pour  $k$ , nous choisissons de ne pas distinguer dans la suite de notre travail  $k$  de l'expression  $\exp(\delta_{ij})$ .

A partir des probabilités définies par (5), nous pouvons déterminer l'espérance de points d'une équipe  $i$  confrontée à ses  $n$  adversaires<sup>7</sup>,  $E(\pi_i)$ , de la manière suivante :

---

<sup>6</sup>Les conditions d'une rencontre (comme par exemple l'état du terrain) peuvent également influencer la capacité à faire des différences.

<sup>7</sup>Les équipes s'affrontant en match aller-retour, nous supposons similairement aux précédents travaux, que les potentiels avantages à domicile se compensent sur l'ensemble de la saison et qu'ils n'influencent donc pas le calcul de l'espérance de points pour le classement final.



$$\begin{aligned}
E(\pi_i) &= \pi^+ \sum_{j=1}^n k \exp(\delta_{ij}) / [k \exp(-\delta_{ij}) + 1 + k \exp(\delta_{ij})] \\
&\quad + \pi^- \sum_{j=1}^n 1 / [k \exp(-\delta_{ij}) + 1 + k \exp(\delta_{ij})] \\
&= \pi^+ \sum_{j=1}^n p_{ij}^+ + \pi^- \sum_{j=1}^n p_{ij}^-
\end{aligned} \tag{6}$$

avec  $\pi^+$  et  $\pi^-$ , le nombre de points attribués en cas de victoire et de match nul respectivement.

- *Hypothèse 1 : en considérant que le marché des joueurs et des entraîneurs<sup>8</sup> dans le milieu du football professionnel est concurrentiel, ces acteurs sont alors rémunérés<sup>9</sup> à leur productivité/performance. Les spectateurs fondent dès lors leurs croyances initiales sur le classement final à partir du budget salarial moyen des joueurs et de l'entraîneur de l'équipe première des clubs<sup>10</sup>.*

Comme le soulignent Kuper et Szymanski [2016], pour une saison donnée, la masse salariale explique environ 70% de la variation du classement final, la partie non expliquée pouvant provenir de perturbations aléatoires souvent non observables (blessures, adaptation au système de jeu, performances surprises, etc.), dont nous tiendrons compte dans la section suivante afin d'assouplir l'hypothèse 1. La masse salariale d'une équipe est donc un bon candidat pour apprécier la qualité initiale d'une équipe avant le début de la saison. Afin de permettre des comparaisons inter-temporelles et d'éviter d'être affecté par des effets d'inflation ou de la spécificité du système de rémunération dans le football professionnel, nous considérons les différences relatives entre équipes, soit :

$$q_i = \frac{\bar{y}_i - \bar{Y}}{\sigma_{\bar{Y}}} \tag{7}$$

---

<sup>8</sup>Les salaires des autres membres du *staff* ayant une variabilité interclub faible, ils ne sont ici pas considérés pour évaluer les différences de qualités des équipes.

<sup>9</sup>Contrairement à la valeur d'un joueur qui représente une estimation de sa productivité potentielle pour les saisons futures et pas simplement de sa performance actuelle.

<sup>10</sup>Ce lien entre performance et rémunération est également significatif pour le marché des joueurs "vedettes" (Bourg [2008])

avec  $\bar{y}_i$  le salaire moyen de l'équipe première de  $i$ ,  $\bar{Y}$  le salaire moyen sur l'ensemble du championnat et  $\sigma_{\bar{Y}}$  l'écart-type de ces salaires moyens.

A ce niveau, nous pouvons distinguer deux types d'incertitude sur la structure du championnat. Premièrement, un aléa de classement dont le calcul repose sur les différences de points observées,  $\pi^{obs}$ , entre une équipe  $i$  et  $j$  durant la saison :

$$I_{obs} = \max \left( 1 - \frac{|\pi_i^{obs} - \pi_j^{obs}|}{114 - 3 * J}, 0 \right) \quad (8)$$

Le dénominateur indique que l'incertitude  $I_{obs}$  décroît avec le nombre de points restant à gagner, soit un maximum de 114 points sur une saison de Ligue 1 moins 3 points par journées,  $J$ , jouées. Cette modélisation permet notamment, par rapport à la mesure de Kuypers [1996], d'avoir une incertitude qui vaut<sup>11</sup> 0 lorsqu'une équipe  $i$  ne peut plus dépasser ou se faire rejoindre par un concurrent  $j$ , si l'écart de points est supérieur à celui restant à attribuer pour terminer le championnat. Nous calculons  $I_{obs}$  pour toutes les positions du championnat, à l'exception des places relégables, où l'écart est calculé par rapport à la 17<sup>ième</sup> place.

A partir de (6) et des masses salariales des clubs publiées par la Commission de Contrôle des Clubs Professionnels (nommée DNCG), nous déterminons une espérance de points pour chaque club avant le début de la saison (*cf.* l'exemple donné en annexe 7.1). Le deuxième type d'incertitude de classement, que nous considérons, est ainsi déterminé selon la relation (8) appliquée aux espérances de points de chaque équipe à l'issue de la saison. L'un des avantages de considérer cette forme d'incertitude est que l'on peut, notamment, expliquer l'affluence de spectateurs lors de la première journée de championnat pour laquelle il n'existe pas de distribution préalable de points ; ce classement est par la suite réévalué après le déroulement de chaque journée. Le nombre estimé de points des équipes est ainsi recalculé en fonction des points effectivement gagnés plus l'espérance de points sur les matchs restant à jouer. Le classement final attendu est ainsi corrigé suivant la performance observée des équipes au fur et à mesure du déroulement de la saison. Cette méthodologie permet d'aboutir à la dernière journée à un nombre de points espérés qui correspond au

---

<sup>11</sup>L'incertitude ne pouvant être inférieure à 0, les valeurs négatives sont remplacées par la valeur nulle.

nombre de points effectivement gagnés. Contrairement à la précédente mesure l'ensemble de ces écarts de chaque journée est rapporté au nombre de points maximal du championnat, soit 114, assurant une incertitude de même amplitude au cours de la saison.  $I_{obs}$  détient par conséquent un poids moins important au début de la compétition, relativement à l'incertitude calculée sur l'espérance de points, et elle s'accroît lorsqu'on se rapproche de la fin du championnat. Autrement dit, l'influence des résultats observés l'emporte au fur et à mesure du déroulement du championnat sur les croyances initiales du début de saison.

### 3 Incertitude d'une rencontre & calcul Bayésien

Notre calcul de probabilités pour caractériser une rencontre de Ligue 1, dans la section précédente, ne considère pas que ces dernières sont subjectives et qu'elles représentent un pari incertain. L'approche Bayésienne de cette section va permettre d'introduire une incertitude dans le calcul de ces probabilités.

Posons  $p_{ij}$  la probabilité de ne pas perdre, soit  $p_{ij}^+ + p_{ij}^-$ . Les chances de match nul et de victoire peuvent être alors déterminées relativement à  $p_{ij}$ , soit  $\frac{p_{ij}^-}{p_{ij}}$  et  $\frac{p_{ij}^+}{p_{ij}}$ . Pour considérer que nos probabilités sont incertaines, nous supposons que  $p_{ij}$  est une variable aléatoire qui peut être décrite suivant une loi *Bêta* de paramètre  $(a_{ij}; b_{ij}) = [\frac{p_{ij}^-}{p_{ij}} * p_{ij} ; \frac{p_{ij}^+}{p_{ij}} * p_{ij}]$ , soit  $(\pi_{ij}^- ; \pi_{ij}^+)$ . Cette hypothèse permet ainsi de déterminer la probabilité de faire match nul et de gagner, le risque de défaite étant le complémentaire de la somme de ces deux précédentes probabilités. Cette modélisation bivariée n'apporte toutefois pas beaucoup d'information supplémentaire par rapport à notre cadre déterministe de la section précédente.

En particulier, nous connaissons le résultat des rencontres et on peut ainsi, selon le théorème de Bayes, se demander ce que deviendrait notre calcul de probabilités sachant l'issue  $x$  du match. Les précédentes performances avant la rencontre  $x$  donnent également des informations exploitables sur la qualité de l'équipe et les spectateurs tiennent compte de la "forme du moment" d'une équipe pour réviser leurs croyances initiales du début de saison. Il est donc nécessaire d'exploiter ces informations pour réviser le calcul des probabilités suivant les résultats observés.

Une solution est alors de considérer que notre espace probabiliste suit *a posteriori*

(après l'observation du résultat  $x$ ) un processus de *Dirichlet*, noté  $D$ . Ce dernier appartient à la famille des lois exponentielles, similairement au calcul des probabilités présenté dans le système (5), et il revient à considérer que  $p_{ij}$  suit une loi *Bêta* (Parent et Bernier [2007]). En outre, cette loi multivariée permet d'estimer directement les trois probabilités caractérisant une rencontre de Ligue 1. Afin de spécifier correctement les paramètres de cette distribution *a posteriori*, un certain nombre d'hypothèses supplémentaires, portant sur le résultat du match concerné  $x$  et sur les rencontres antérieures à  $x$ , doit être au préalable discuté.

- *Hypothèse 2* : les chances de l'équipe recevant de remporter la rencontre sont corrélées positivement à la probabilité de victoire,  $p_{ij}^+$ , déterminée selon (5). On suppose également que  $p_{ij}^+$  est une fonction croissante du nombre de matchs gagnés,  $n_i^+$ , de  $i$  ( $x$  inclus si  $i$  a remporté l'actuelle rencontre). Un indice  $j'$  est alors introduit pour différencier les adversaires que  $i$  a vaincu,  $j'$  variant de 0 (en cas d'aucune victoire de  $i$ ) à  $n_i^+$ . En outre, la qualité de l'équipe  $i$  est une fonction croissante de la force de l'adversaire  $j'$ ,  $(1 - p_{ij'}^+)$ , qu'elle a vaincu. Si  $i$  a gagné contre une équipe supposée moins forte la mise à jour des chances de victoire lors de la rencontre entre  $i$  et  $j$  dans  $D$  sera alors  $1 - (p_{ij'}^+ - p_{ij'}^-)$ . Si elle a par contre gagné une équipe supposée plus forte, la mise à jour sera  $1 + (p_{ij'}^- - p_{ij'}^+)$ . Soit au final, pour l'ensemble des victoires de  $i$  :

$$D(p_{ij}^+ + \sum_{j'=0}^{n_i^+} [1 + p_{ij'}^- - p_{ij'}^+]) = D(p_{ij}^+ + n_i^+ + \sum_{j'=0}^{n_i^+} [p_{ij'}^- - p_{ij'}^+]) \quad (9)$$

- *Hypothèse 3* : nous supposons également que la probabilité de victoire de l'équipe recevant est une fonction croissante du nombre de défaites,  $n_j^-$ , de l'adversaire  $j$ . Dans une démarche similaire à la précédente hypothèse, cette probabilité de victoire est une fonction croissante de la force des équipes  $i'$  contre lesquelles  $j$  a perdu ( $x$  inclus si  $j$  a perdu l'actuelle rencontre),  $i'$  représentant désormais la  $i^{\text{ème}}$  équipe contre laquelle  $j$  a perdu ( $i'$  varie ainsi de 0 à  $n_j^-$ ). Nous réalisons ensuite une

moyenne arithmétique des processus de ces deux premières hypothèses.

$$D(p_{ij}^+ + \frac{n_i^+ + n_j^- + \sum_{j'=0}^{n_i^+} [p_{ij'}^- - p_{ij'}^+] + \sum_{i'=0}^{n_j^-} [p_{i'j}^+ - p_{i'j}^-]}{2}) \quad (10)$$

- *Hypothèse 4* : nous considérons que la probabilité de défaite de l'équipe  $i$  peut être construite de manière symétrique à celle de victoire. Un raisonnement similaire est également appliqué à la probabilité d'un score égal en considérant que celle-ci est corrélée positivement aux chances de ne pas faire match nul,  $(1 - p^=)$ , lors des précédents résultats nuls des équipes  $i$  et  $j$  ( $x$  inclus si la rencontre présente se termine par une égalité). Sachant que  $p^- + p^+ + p^= = 1$ , la mise à jour de la probabilité de match nul suivant les résultats de  $i$  face aux adversaires  $j'$  peut se réécrire :  $\sum_{j'}^{n_i^-} (1 - p_{ij'}^=) = \sum_{j'}^{n_i^-} (1 + p_{ij'}^+ + p_{ij'}^- - 1) = n_i^- + \sum_{j'}^{n_i^-} (p_{ij'}^+ + p_{ij'}^- - 1)$ . Puisque  $(p_{ij'}^+ + p_{ij'}^- - 1) < 0$ , nous réajustons ce dernier terme pour avoir une amplitude similaire aux mises à jour des probabilités de victoire et défaite (cf. hypothèses 3 et 2) de la manière suivante :  $\frac{3}{2}(p_{ij'}^+ + p_{ij'}^-) - 1$ . Nous notons dans la suite du texte,  $m_{ij}^+$ ,  $m_{ij}^=$  et  $m_{ij}^-$ , les trois paramètres de la distribution  $D$  :

$$\begin{aligned} D(m_{ij}^+ = p_{ij}^+ + \frac{n_i^+ + n_j^- + \sum_{j'=0}^{n_i^+} [p_{ij'}^- - p_{ij'}^+] + \sum_{i'=0}^{n_j^-} [p_{i'j}^+ - p_{i'j}^-]}{2}; \\ m_{ij}^= = p_{ij}^= + \frac{n_i^- + n_j^- + \sum_{j'=0}^{n_i^-} [\frac{3}{2}(p_{ij'}^+ + p_{ij'}^-) - 1] + \sum_{i'=0}^{n_j^-} [\frac{3}{2}(p_{i'j}^+ + p_{i'j}^-) - 1]}{2} \\ m_{ij}^- = p_{ij}^- + \frac{n_i^- + n_j^+ + \sum_{j'=0}^{n_i^-} [p_{ij'}^- - p_{ij'}^+] + \sum_{i'=0}^{n_j^+} [p_{i'j}^+ - p_{i'j}^-]}{2}) \end{aligned} \quad (11)$$

- *Hypothèse 5* : afin de prendre en compte le potentiel avantage de recevoir, les performances à domicile et à l'extérieur des équipes  $i$  et  $j$  sont respectivement distinguées. Avec des probabilités déterminées selon des fonctions exponentielles, l'avantage du terrain revient à considérer qu'un facteur multiplicatif  $h$  affecte les différences de qualité  $\delta_{ij}$  (Rao et Kupper [1967]). Pour estimer un  $h\delta_{ij}$  spécifique à chaque rencontre, une mise à jour des paramètres de  $D$  est réalisée suivant les résultats à domicile et à l'extérieur des équipes  $i$  et  $j$  respectivement. Autrement dit, nous ne considérons

que les victoires à domicile de  $i$  dans l'hypothèse 2, les défaites à l'extérieur de  $j$  dans l'hypothèse 3 et les matchs nuls à domicile et à l'extérieur de  $i$  et  $j$ , respectivement (cf. hypothèse 4).

- *Hypothèse 6 : les résultats des dernières rencontres doivent détenir une importance prépondérante pour capter la "forme du moment" d'une équipe relativement à des performances plus anciennes. Dans cette perspective, nous effectuons une moyenne pondérée<sup>12</sup>, en accordant un poids  $J$  au résultat de la  $J^{\text{ième}}$  journée divisée par la somme totale des poids attribués à chaque journée. Autrement dit, l'influence des premiers matchs se réduit au fur et à mesure du déroulement de la saison au profit de résultats plus récents.*
- *Hypothèse 7 : les mises à jour des paramètres de  $D$  présentent une variance plus importante en début de championnat. Cette volatilité diminue pour atteindre une tendance plus stable à partir de la huitième journée. Nous attribuons par conséquent une pondération décroissante entre les probabilités, déterminées selon (5), et les résultats des matchs observés<sup>13</sup>. La distribution a posteriori, incluant l'historique des rencontres, est ainsi dominée par les croyances en début de championnat, alors qu'à partir de la neuvième journée les résultats observés dominent les croyances initiales.*

Cette variance plus importante des paramètres de notre distribution *a posteriori* est synonyme d'une incertitude élevée lors des premières rencontres. Cette dernière se réduit lorsqu'il y a suffisamment de résultats pour avoir une meilleure appréhension de la qualité de l'équipe pour le reste de la saison. Ce processus statistique est similaire au comportement du spectateur qui va commencer à réviser ses croyances initiales uniquement lorsqu'il

---

<sup>12</sup>Permettant, également, d'éviter des effets d'échelle dans la mise à jour des différences relatives entre équipes.

<sup>13</sup>En considérant l'hypothèse 5, il est donc nécessaire d'avoir deux distributions identiques de poids : l'une pour les résultats à domicile, l'autre pour ceux à l'extérieur. Puisque le calendrier d'une saison repose sur un principe d'alternance entre matchs à domicile et à l'extérieur, les pondérations doivent se modifier toutes les deux journées. Par conséquent, nous attribuons pour la première et la deuxième journée des poids de 5/6 et 1/6 entre  $p_{ij}^+$  et sa mise à jour *a posteriori* respectivement. Cette pondération décroît par la suite pour se stabiliser à 1/6 et 5/6 à partir de la neuvième journée.

estimera avoir suffisamment d'information (nombre de rencontres jouées) pour se forger une opinion crédible.

## 4 Echantillonnage & spécification économétrique

L'estimation de l'écart de qualité,  $\delta$ , entre les équipes  $i$  et  $j$  s'effectue par simulation. Cette dernière est implémentée par la méthode d'échantillonnage pondéré/préférentiel (*Importance Sampling*, notée IS par la suite).

L'échantillonnage préférentiel consiste à générer un échantillon selon une distribution instrumentale  $g(\delta)$ , de manière à obtenir  $n$  observations, notées  $\delta_1, \dots, \delta_i, \dots, \delta_n$ . On souhaite alors transformer cet échantillon afin de se rapprocher des caractéristiques, ou importance, de la distribution *a posteriori*  $f(\delta)$ . Les observations simulées sont alors pondérées selon des poids notés  $w$  afin de compenser les disparités entre les échantillons selon la fonction  $f(\delta)$  et  $g(\delta)$ . L'estimateur de la moyenne pondérée  $U$  est déterminé tel que :

$$\tilde{U} = \frac{1}{n} \sum_{i=1}^{i=n} u(\delta_i) w_i \text{ avec } w(\delta_i) = \frac{f(\delta_i)}{g(\delta_i)} \quad (12)$$

Dès lors, si la fonction instrumentale  $g$  surreprésente un point  $\delta_i$  relativement à la densité de la loi cible  $f$  alors le poids associé sera  $w(\delta_i) = \frac{f(\delta_i)}{g(\delta_i)} < 1$ . Réciproquement cette pondération sera supérieure à 1 pour les points sous-représentés.

$f$  étant, dans notre cas, connue à une constante près (*cf.* la vraisemblance *a posteriori* définie par 17), nous utilisons la méthode IS approchée. En posant  $u = 1$  dans (12), nous obtenons :

$$\tilde{U} = \frac{1}{n} \sum_{i=1}^{i=n} w_i \sim E(w_i) = 1 \quad (13)$$

$$E(w_i) = \int \frac{f(\delta_i)}{g(\delta_i)} g(\delta_i) d\delta_i = \int f(\delta_i) d\delta_i = 1 \quad (14)$$

Ainsi :

$$\frac{1}{n} \sum_{i=1}^{i=n} w_i \rightarrow 1 \text{ lorsque } n \rightarrow \infty \quad (15)$$

L'estimateur approché de  $U$  devient alors :

$$\tilde{U} = \frac{\sum_{i=1}^{i=n} u(\delta_i) w_i}{\sum_{i=1}^{i=n} w_i} \text{ quand } n \rightarrow \infty \quad (16)$$

La relation (16) est un estimateur biaisé mais convergent quand  $n$  tend vers l'infini. Ainsi, la fonction  $f(\delta)$  n'a besoin d'être connue qu'à une constante près pour le calcul de la pondération relative. Suivant notre approche bayésienne<sup>14</sup>,  $f(\delta)$  est proportionnelle à la vraisemblance de la distribution de *Dirichlet*, à une constante multiplicative près :

$$[p_{ij}^+(\delta_{ij})]^{m_{ij}^+} [p_{ij}^-(\delta_{ij})]^{m_{ij}^-} [p_{ij}^0(\delta_{ij})]^{m_{ij}^0} \quad (17)$$

Nous désirons obtenir un *prior* informatif tel que  $N(0, \delta) \sim N(0, g)$ ,  $g$  étant spécifié comme une distribution de probabilité tel que  $g \sim \text{Gamma}(\alpha; \beta)$ . Autrement dit, nous considérons initialement que les deux équipes ont une force similaire (espérance nulle), puis nous échangeons des joueurs entre les deux équipes afin de créer l'écart le plus adapté à la vraisemblance de la distribution *a posteriori*. Les hyper-paramètres  $\alpha$  et  $\beta$  de la loi *Gamma* sont élicités tels que<sup>15</sup>  $\alpha = c * \text{variance}(\text{prior non informatif})$  et  $\beta = c$ , avec  $c$  une constante positive qui optimise la précision de l'estimation.

Afin d'initialiser l'algorithme, le *prior non informatif*, qui traduit notre méconnaissance initiale sur les différences de qualités entre les équipes, est généré selon une loi normale  $N(0;1)$ , constituant un échantillon de 10 000 variables aléatoires. Le  $c$  optimal est alors celui qui minimise le rapport de la vraisemblance d'obtenir  $g$  suivant la distribution *a posteriori* ( $f$ ) sur celle de parvenir à  $g$  suivant la loi *Gamma*, menant à l'estimation de l'écart de qualités entre les équipes. Cette démarche permet ainsi de construire un

<sup>14</sup>Des détails supplémentaires sont donnés en annexe sur la méthode IS appliquée dans un cadre de raisonnement bayésien.

<sup>15</sup>Puisque  $E(g) = \frac{\alpha}{\beta}$  et  $Var(g) = \frac{\alpha}{\beta^2}$  alors  $E(g) = \frac{\alpha}{c} * c = \alpha$  et  $Var(g) = \frac{c\alpha}{c^2} = \frac{\alpha}{c}$ .



*prior* informatif, incluant de manière aléatoire l'information contenu dans la distribution *a posteriori* et permettant, par conséquent, d'assouplir<sup>16</sup> l'ensemble de nos hypothèses précédemment discuté.

Nous savons que l'échantillon de la distribution *a posteriori* est similaire à celui généré pour le *prior* si les poids de la méthode IS ne sont pas trop hétérogènes. Un critère d'information couramment utilisé, dans ce cadre, est la taille d'échantillon effective (*Effective Sample Size*, abrégé *EES* par la suite). Selon Robert et Casella [2009], le critère *ESS* peut se définir comme un ratio de variance :

$$ESS = \frac{\left(\sum_{i=1}^{i=n} w_i\right)^2}{\sum_{i=1}^{i=n} w_i^2} \quad (18)$$

Afin de considérer une estimation acceptable de l'écart, il est nécessaire de minimiser ce critère. Un algorithme a été programmé à cette fin sur le logiciel *R*, dans lequel l'estimation s'arrête lorsque *ESS* est inférieur à 5 (Goldstick [2009]). Un exemple de l'ensemble des probabilités estimées suivant le *prior* informatif est donné en annexe (7.3). A ce stade, il est utile de donner quelques informations concernant la qualité d'inférence de nos probabilités. Le tableau 1 résume l'ensemble des prédictions correctes, relativement aux résultats observés lors des deux saisons considérées.

**Tableau 1 : pourcentage de prédictions correctes (saisons 2014-16)**

Résultats	Probabilités déterministes	Distribution de Dirichlet	<i>Prior</i> Informatif
<i>Victoire domicile</i>	63.93%	80.35%	69.21%
<i>Match nul</i>	74.08%	87.52%	74.08%
<i>Défaite domicile</i>	61.05%	85.40%	75.53%
Ensemble	66.35%	84.42%	72.94%

Source : LFP & calculs de l'auteur

<sup>16</sup>Permettant de considérer, en particulier, que le salaire moyen de l'équipe première d'un club est un indicateur imparfait de la qualité d'une l'équipe.

Les équations déterministes suivant (5) prédisent correctement environ 66% des résultats observés à l'issue des rencontres. En tenant compte de l'historique des matchs, des performances à domicile et à l'extérieur des équipes ainsi que du résultat  $x$  (distribution *a posteriori*), les prédictions correctes augmentent à plus de 84%. L'information incluse dans la distribution *a posteriori* améliore donc significativement l'appréhension du déroulement du championnat, confirmant dans une certaine mesure l'intérêt des précédentes hypothèses discutées. La combinaison de la distribution *a posteriori* et du *prior non informatif* permet d'aboutir à un *prior informatif*, qui inclut une partie des croyances du spectateur. Cette approche évalue comme attendu une proportion de prédictions correctes qui se situe entre celle donnée par les fonctions déterministes et celle de la distribution *a posteriori*. Ce *prior informatif* est alors utilisé pour déterminer l'ensemble des probabilités caractérisant une rencontre de Ligue 1.

Etudier les déterminants de l'affluence de spectateurs au cours d'une saison peut s'avérer une tâche difficile, étant donné les spécificités de chaque club participant à cette compétition. Notre objectif n'est pas d'essayer de maximiser le public total dans les stades de Ligue 1, mais d'identifier les facteurs de l'affluence quels que soient les clubs concernés. Dans cette perspective, cette étude analyse comme variable expliquée les variations de remplissage au-dessus du minimum de spectateurs observé durant la saison pour un club donné  $i$ , soit formellement :

$$t_i = \frac{AFF_{ij} - \text{Min}(AFF_i)}{C_i} \quad (19)$$

avec  $AFF_{ij}$  l'affluence observée lors d'une rencontre de ligue 1 entre l'équipe recevant  $i$  et l'équipe visiteuse  $j$ ,  $\text{Min}(AFF_i)$  est l'affluence minimum observée au cours de la saison de Ligue 1 pour le club  $i$  et  $C_i$  la capacité maximale du stade de  $i$ . Ce choix de variable expliquée permet d'éliminer l'influence des "fans" inconditionnels venant au stade, quel que soit la performance de leur équipe. En divisant ce différentiel d'affluences par la capacité des stades, certaines caractéristiques structurelles des clubs (comme par exemple la densité locale de population) deviennent moins prépondérantes.

Afin d'identifier correctement l'effet de l'incertitude de classement à l'issue du championnat, un ensemble de variables de contrôle est utilisé pour capter les différents effets

potentiels pouvant influencer le choix du spectateur d'assister ou non à une rencontre de Ligue 1. Dans ce cadre, un indicateur de température est construit puisque lors des périodes hivernales les individus sont probablement moins incités à venir supporter leur équipe. Cette variable a été renseignée à partir des températures observées dans les stations météo lors du lieu de la rencontre, le jour et l'horaire<sup>17</sup>.

Nous incluons également une variable "rivalité" qui représente les oppositions historiques entre équipes (définition plus large que les *derby*). La définition de cet indicateur comporte toujours une part relativement arbitraire, mais les spectateurs raisonnent également de manière subjective, en adoptant chacun leur propre définition qui peut varier selon les localisations et les périodes. Pour être certain que notre variable d'incertitude ne soit pas parasitée par un effet d'oppositions historiques, nous avons retenu<sup>18</sup> un nombre relativement important de rivalités, que l'ensemble des spectateurs peuvent potentiellement considérer.

Enfin, comme le suggèrent Forrest et Simmons [2002], l'incertitude d'une rencontre est mesurée par la différence de probabilités de victoire des deux équipes concernées. Au niveau du classement, le calcul de probabilités permet de considérer le risque de se faire rattraper et les chances de rejoindre un concurrent. Ces deux probabilités peuvent être calculées après chaque journée de championnat sur la répartition du nombre de points observée et/ou sur l'espérance de points à l'issue de la saison (*cf.* section 2).

## 5 Résultats

Etant donné la potentielle colinéarité entre les probabilités de dépasser ou de se faire rattraper durant la saison, nous retenons les plus significatives : (*i*) la probabilité observée

---

<sup>17</sup>Ces informations sont consultables sur le site Meteociel.fr. Des données sur la pluviométrie ont été également testées, mais celles-ci ne se sont pas révélées significatives.

<sup>18</sup>Soit le *derby* de la Côte d'Azur (OGC Nice *vs.* AS Monaco), le *derby* Rhônalpin (Olympique Lyonnais *vs.* AS Saint-Etienne), le *derby* Breton (Stade Rennais *vs.* FC Nantes), le *derby* du Sud-Est (OGC Nice *vs.* Olympique de Marseille,) le *Classico* (Paris Saint-Germain *vs.* OM), l'*Olympico* (Olympique Lyonnais *vs.* Olympique de Marseille), le *derby* Corse (GFC Ajaccio *vs.* SC Bastia, uniquement pour la saison 2015-16) et le *Cashico* (Paris Saint-Germain *vs.* AS Monaco).

de se faire rattraper par le(s) concurrent(s) du moment (notée  $I_{obs}^r$ ), calculée suivant (8), relativement à l'équipe classée juste en dessous avant le déroulement de la journée concernée ; (ii) la probabilité de terminer devant le(s) concurrent(s) attendu(s) en fin de saison (*cf.* section 2).

La significativité de  $I_{obs}^r$  indique que les spectateurs sont concernés par le risque que leur équipe favorite se fasse rattraper par les poursuivants dans un horizon temporel de courte ou moyenne période. Cette probabilité renseigne, en outre, sur la performance relative observée par rapport aux poursuivants immédiats. *A contrario*, lorsque les spectateurs s'intéressent aux places situées "au-dessus" dans le classement, ces derniers ne raisonnent plus sur la distribution de points observés mais plutôt sur la probabilité espérée de pouvoir terminer à une meilleure position à l'issue de la saison, soit cette fois-ci la performance relative en longue période de leur équipe par rapport aux concurrents attendus pour le classement final. Cette incertitude espérée à l'issue de la saison est notre principale variable d'étude puisqu'elle renseigne sur l'intérêt que porte les spectateurs à la structure concurrentielle du championnat.

Suivant les coefficients estimés des variables de contrôle (*cf.* tableau 2), la probabilité  $I_{obs}^r$  présente une influence négative sur l'affluence de spectateurs. Lorsque leur équipe favorite réalise des performances moindres par rapport à leur(s) poursuivant(s) du moment, les spectateurs sont moins incités à venir soutenir leur équipe. Cet effet est très homogène<sup>19</sup> quel que soit la position des équipes dans le championnat, nous choisissons dès lors de ne pas avoir un découpage plus détaillé de cette probabilité. Par contre, cette homogénéité ne se retrouve pas pour les chances de rattraper un concurrent à l'issue de la saison. Cette probabilité est donc décomposée comme l'indique le tableau (3). Les coefficients estimés de l'ensemble des facteurs explicatifs<sup>20</sup> sont ainsi synthétisés dans les tableaux 2 et 4. Ils sont déterminés à partir de la spécification (20), ci-dessous :

---

<sup>19</sup>Une légère diminution de significativité est uniquement observée au niveau du milieu de classement.

<sup>20</sup>Les variables explicatives ont été centrées et réduites afin de rendre nos estimations insensibles à l'étendue des facteurs considérés (Scherrer [1984]) et de faciliter les comparaisons inter-temporelles.

$$t_i = \beta_0 + \sum_{k=1}^{19} \beta_k * Prob_k + \beta_{20} |p_i^+ - p_i^-| + \beta_{21} |p_i^+ - p_i^-|^2 + \beta_{22} I_{obs}^r + \beta_{23} Température + \beta_{24} Rivalités + \epsilon_i \quad (20)$$

avec  $Prob_k$  la probabilité que l'équipe, censée terminer à la  $k-1^{ième}$  position, parvient finalement à être classée à la  $k^{ième}$  position à l'issue du championnat<sup>21</sup> ; l'ensemble des  $\beta$  représente les coefficients estimés et  $\epsilon_i$  les résidus de ce modèle.

**Tableau 2 : coefficients estimés des variables de contrôle (saisons 2014-16)**

Variables	2014-15	2015-16
$-2 \log(\text{vraisemblance})$	-356.4	-250.0
<i>AIC</i>	-354.4	-247.0
<i>Constante</i>	0.157*** (0.006)	0.159*** (0.007)
$ p_i^+ - p_i^- $	-0.223** (0.100)	-0.069* (0.048)
$ p_i^+ - p_i^- ^2$	0.233** (0.100)	0.041 (0.048)
$I_{obs}^r$	-0.041*** (0.006)	-0.032*** (0.007)
<i>Température</i>	0.016** (0.006)	0.022*** (0.008)
<i>Rivalités</i>	0.018*** (0.006)	0.036*** (0.008)
<i>Nombre de matchs</i>	380	380

Source : estimations de l'auteur

Suivant les résultats et les critères d'information<sup>22</sup>, les  $t_i$  sont mieux expliqués en

<sup>21</sup> Par exemple,  $Prob\ 1^{er}$  représente la probabilité que le deuxième attendu au classement rattrape finalement le premier.

<sup>22</sup> Cf. le critère *AIC* et le logarithme de la vraisemblance du modèle.

2014-15 que pour l'année suivante. Concernant les résultats des deux saisons, les coefficients estimés sont de signes identiques et attendus, témoignant d'une certaine robustesse de notre spécification économétrique. La constante correspond au remplissage moyen "au-dessus" du minimum observé<sup>23</sup> dans la distribution des  $t_i$  (*cf.* annexe 7.4), ce dernier étant plus important en 2015-16. La température présente le signe espéré avec une augmentation significative de l'affluence hors périodes hivernales (Baranzini, *et al.* [2008]) et les "rivalités" entre équipes semblent également susciter un intérêt plus important des spectateurs (Hogan *et al.* [2013]).

La différence de probabilités de victoire entre l'équipe recevant et visiteuse,  $|p_i^+ - p_i^-|$ , indique que plus le résultat de la rencontre est prévisible moins les spectateurs sont incités à venir au stade. Contrairement à la récente étude de Andreff et Scelles [2015], qui ne met pas en évidence de lien significatif entre incertitude d'une rencontre et affluence dans les stades de Ligue 1, nous retrouvons ici un résultat cohérent relativement à la vision initiale de Rottenberg [1956] et Neale [1964], avec une corrélation positive entre incertitude *a priori* d'un match et affluence de spectateurs<sup>24</sup>. Pour approfondir ce lien et étudier une potentielle non-linéarité de la relation entre affluence et incertitude d'une rencontre, une forme quadratique de cette variable a été introduite. Si certitude et nombre de spectateurs sont corrélés négativement, le coefficient estimé de ce terme au carré révèle que cet effet diminue pour les plus importantes différences de probabilités. Autrement dit, dans le cas d'un match fortement déséquilibré, les spectateurs sont incités à venir observer l'opposition face à une équipe composée probablement de joueurs "vedettes" de la Ligue 1 (Hill *et al.* [1982] ; Jones [1984] et Scully [1974]) ou d'assister à une large victoire de leur club favori.

---

<sup>23</sup>Les variables explicatives étant centrées et réduites, l'ordonnée à l'origine représente la moyenne des  $t_i$ .

<sup>24</sup>Certains travaux empiriques (*cf.* par exemple, Czarnitzki et Stadtmann, [2002] ; Pawlowski et Anders [2012]) parviennent à un résultat inverse, avec une corrélation négative entre incertitude et affluence.

**Tableau 3 : coefficients estimés de la décomposition de la variable d'intérêt**

Variables	2014-15	2015-16	Variables	2014-15	2015-16
<i>Prob 1<sup>er</sup></i>	0.057*** (0.009)	0.022** (0.011)	<i>Prob 11<sup>ième</sup></i>	0.029*** (0.009)	0.026*** (0.010)
<i>Prob 2<sup>ième</sup></i>	0.026*** (0.009)	0.018* (0.010)	<i>Prob 12<sup>ième</sup></i>	0.037*** (0.009)	0.033*** (0.010)
<i>Prob 3<sup>ième</sup></i>	0.034*** (0.009)	0.032*** (0.009)	<i>Prob 13<sup>ième</sup></i>	0.026*** (0.009)	0.024** (0.010)
<i>Prob 4<sup>ième</sup></i>	0.040*** (0.009)	0.055*** (0.010)	<i>Prob 14<sup>ième</sup></i>	0.041*** (0.009)	0.021** (0.010)
<i>Prob 5<sup>ième</sup></i>	0.051*** (0.009)	0.028*** (0.010)	<i>Prob 15<sup>ième</sup></i>	0.045*** (0.009)	0.016* (0.009)
<i>Prob 6<sup>ième</sup></i>	0.033*** (0.009)	0.021** (0.010)	<i>Prob 16<sup>ième</sup></i>	0.042*** (0.009)	0.028*** (0.010)
<i>Prob 7<sup>ième</sup></i>	0.029*** (0.009)	0.036*** (0.010)	<i>Prob 17<sup>ième</sup></i>	0.041*** (0.009)	0.006 (0.010)
<i>Prob 8<sup>ième</sup></i>	0.035*** (0.009)	0.016* (0.009)	<i>Prob 18<sup>ième</sup></i>	0.043*** (0.009)	0.058*** (0.010)
<i>Prob 9<sup>ième</sup></i>	0.052*** (0.009)	0.036*** (0.010)	<i>Prob 19<sup>ième</sup></i>	0.027*** (0.009)	0.018* (0.010)
<i>Prob 10<sup>ième</sup></i>	0.040*** (0.008)	0.029*** (0.010)	-	-	-

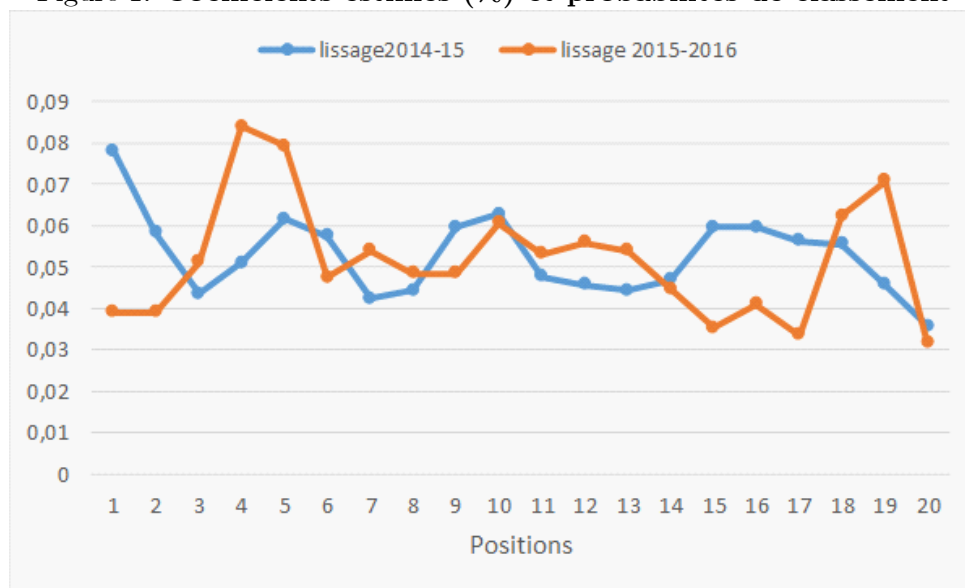
Source : estimations de l'auteur

Le tableau 3 renseigne sur la décomposition de la probabilité de rattraper un concurrent à l'issue de la saison. Elle a été réalisée pour l'ensemble des écarts, concernant toutes les positions envisageables du classement. Tous les coefficients estimés étant significatifs et positifs, ils indiquent que les spectateurs sont intéressés par une probabilité (ou incertitude) plus importante de rattraper des concurrents à l'issue de la saison. Globalement, le déroulement de la compétition en 2014-15 semble avoir été plus attractif, puisque la

somme des coefficients estimés est supérieure à celle de la saison suivante. Cet intérêt supplémentaire provient notamment d’une lutte plus serrée pour le titre en 2014-15 (*cf.* Prob 1<sup>er</sup>), avec un titre de champion décroché par le Paris Saint-Germain (PSG) à la 37<sup>ième</sup> journée ; contrairement à la saison 2015-16, où cette équipe a remporté le championnat à l’issue de la 30<sup>ième</sup> journée. La lutte pour la première place peut ainsi représenter une potentielle source majeure de l’attractivité globale d’une saison.

Pour faciliter l’interprétation de ces résultats, la figure 1 représente les coefficients estimés en pourcentage de l’ensemble des valeurs du tableau 3. En outre, nous effectuons un lissage en calculant la moyenne du coefficient estimé de la position donnée et celui situé immédiatement en-dessous<sup>25</sup> puisqu’une équipe est toujours concernée par deux écarts (excepté pour la première et dernière place).

Figure 1: **Coefficients estimés (%) et probabilités de classement**



Source : estimations de l’auteur

Les deux courbes présentent une forme sinusoïdale quelle que soit la saison. La décomposition obtenue est relativement remarquable puisque la figure 1 fait apparaître un

<sup>25</sup>L’effet négatif associé à la crainte de se faire rejoindre, en courte période, étant considéré comme uniforme quel que soit la position du classement.



groupe d'équipes luttant pour la ligue des champions (*cf.* les effets sur les trois premières positions) et un ensemble d'équipes en concurrence pour les places qualificatives de l'*Europa League* (4<sup>ième</sup> et 5<sup>ième</sup>). La lutte pour la 7<sup>ième</sup> et 8<sup>ième</sup> position suscite un intérêt moindre chez le spectateur : les équipes se situant à ce niveau sont probablement rapidement décrochées de la course aux places européennes. Plus surprenant, il semble que les spectateurs attachent une importance à ce que leur équipe termine dans la première partie du classement (*cf.* les effets entre la 9<sup>ième</sup> et 11<sup>ième</sup> position) pour les deux saisons. Une potentielle explication de ce résultat est que, outre l'enjeu sportif, les supporters accordent également un intérêt non négligeable à l'aspect symbolique : ils souhaitent, malgré une non qualification pour les compétitions européennes, que leur club favori appartienne aux dix meilleures équipes françaises.

Si les deux courbes ont des tendances similaires, elles présentent également deux mouvements contradictoires : la lutte pour la première position a suscité un intérêt relativement important en 2014-15 alors que pour la saison 2015-2016 les spectateurs ont préféré regarder la compétition pour les places de l'*Europa League*. Les trois premières positions du championnat français présentent donc un effet très volatile selon le déroulement de la saison, les nouvelles capacités financières du PSG ayant déformé la structure concurrentielle en haut de classement. La deuxième rupture se situe au niveau des équipes mal classées puisque pour la saison 2015-16 les spectateurs se sont intéressés à la lutte pour éviter la relégation, avec une importance plus forte concernant les clubs classés 18<sup>ième</sup> et 19<sup>ième</sup>, alors que pour la saison 2014-15 cette lutte contre la relégation a concerné plus d'équipes puisqu'un intérêt significatif est repéré à partir de la 15<sup>ième</sup> position. Le format de ligues ouvertes produit donc mécaniquement une attractivité plus importante en comparaison aux ligues fermées des Etats-Unis, avec une incertitude supplémentaire provenant du risque de relégation.

Pour résumer les résultats communs aux deux saisons, l'intérêt des spectateurs est motivé par la lutte pour les places qualificatives de l'*Europa League*. Une saison de Ligue 1 peut ainsi s'avérer très attractive lorsqu'on observe une compétition intense au niveau de la 4<sup>ième</sup> et 5<sup>ième</sup> place, qui peut dans certains cas susciter un intérêt plus important que

l'affrontement pour le titre de champion (saison 2015-16). Le "ventre mou" semble engendrer une moindre (et plus uniforme) attention du spectateur, à l'exception de l'aspect symbolique concernant la 9<sup>ième</sup> et 10<sup>ième</sup> place. La saison peut également présenter une attractivité plus importante si la compétition est intense dans le bas du classement, exceptée l'équipe terminant dernière qui est généralement plus rapidement décrochée et qui génère, par conséquent, une incertitude relativement faible dans la structure du championnat.

## 6 Conclusion

Cette étude propose une nouvelle approche pour modéliser l'incertitude d'une rencontre de football. A partir des croyances initiales sur les forces intrinsèques des équipes et des résultats observés durant la saison, nous avons inféré de manière aléatoire des différences relatives de qualité d'équipes. Suivant cette méthodologie, et contrairement au résultat sur la Ligue 1 de Andreff et Scelles [2015], nous trouvons premièrement un lien significatif entre l'incertitude d'une rencontre et l'affluence : les spectateurs sont moins intéressés par les matchs dont l'aléa de résultat est plus faible. En approfondissant l'analyse de ce lien, un effet non-linéaire est deuxièmement mis en évidence : pour des rencontres fortement déséquilibrées, la baisse d'affluence est moindre, indiquant que les spectateurs désirent également assister aux oppositions face à des équipes composées de joueurs "vedettes" et d'espérer un éventuel exploit de leur club favori.

Contrairement aux précédents travaux, nous avons raisonné à partir d'un espace probabiliste pour appréhender la structure concurrentielle du championnat de France. Ce type d'approche a, notamment, permis de distinguer la probabilité de se faire rattraper en courte et moyenne période des chances de rejoindre un concurrent à l'issue de la saison, apportant une nouvelle interprétation du lien entre incertitude et attractivité.

Suivant cette approche, les spectateurs de Ligue 1 présentent un comportement antagoniste : ils sont démotivés si leur équipe favorite perd du terrain face à ses concurrents au cours du championnat mais ils désirent, cependant, une incertitude plus importante (soit une diminution des écarts), concernant la détermination du classement à l'issue de la saison. Or, ceci n'est envisageable que si certaines équipes perdent du terrain face à

leurs poursuivants. Nous avons également observé que l'intérêt pour les trois premières places est très volatile dans l'actuelle structure du championnat. L'incertitude concernant la victoire finale peut s'avérer un facteur prépondérant, puisque l'attractivité plus importante de la saison 2014-15 provient essentiellement d'une lutte plus serrée pour le titre. Un résultat similaire est constaté dans l'étude de Scelles [2017], reposant sur les audiences télévisuelles du championnat anglais, qui trouve également une corrélation positive entre concurrence pour le titre et attractivité de la saison.

Il semble, par contre, se dégager de manière plus robuste que : (i) les premières places européennes (4<sup>ième</sup> et 5<sup>ième</sup> positions) ont suscité un vif intérêt chez le spectateur (Scelles *et al.* [2016]), (ii) malgré un "ventre mou" relativement moins attractif, le spectateur semble attacher une importance à l'aspect symbolique d'appartenir au groupe des dix meilleurs clubs de la saison, avec un intérêt plus important pour la 9<sup>ième</sup> et 10<sup>ième</sup> place et, enfin, (iii) la lutte contre la relégation pour les équipes situées en bas de classement (pouvant remonter jusqu'à la 15<sup>ième</sup> position, suivant la saison considérée) a produit un certain engouement du public. L'intérêt du spectateur n'est donc pas linéairement décroissant avec la hiérarchie du classement mais plutôt de forme sinusoïdale, suivant les précédents résultats discutés.

Il apparaît sur les deux saisons étudiées que les nouvelles capacités financières du PSG ont réduit l'incertitude pour l'attribution du titre. Pourtant, notre travail montre que la lutte pour la première place peut être un important déterminant de l'attractivité d'une saison, si elle est relativement serrée. Dès lors, un potentiel risque pour le championnat français peut être une domination récurrente du PSG tant qu'il sera sous propriété quatarie, avec un manque de concurrence au niveau national si d'autres clubs ne parviennent pas à plus amplement développer leurs capacités financières.

Ce travail sera certainement à reproduire à chaque changement structurel concernant l'incertitude de classement final. Un prolongement naturel serait dès lors de dépasser le cadre français, en effectuant des comparaisons avec des championnats étrangers. La première Ligue anglaise serait un bon candidat puisque cette compétition est considérée comme le championnat européen le plus attractif et il devrait constituer, par conséquent, un cas de référence. Le championnat espagnol pourrait être également envisagé puisque ce

dernier présente une structure concurrentielle différente, avec une domination historique de deux ou trois clubs.

## **Remerciements**

L'auteur est reconnaissant envers les participants au colloque "Society, Choices and Sports" (Caen [2017]) pour leurs diverses suggestions. Il tient également à remercier C. Burczyk pour sa patience lors des week-ends consacrés aux calculs bayésiens.

## **7 Annexes**

### **7.1 Espérance de points des clubs avant le début de la saison**

**Tableau A1 : espérance de classement avant le début de la saison**

Clubs	Espérance de points (2014-15)	Clubs	Espérance de points (2015-16)
PSG	112	PSG	112
OL	83	OL	84
ASM	83	ASM	62
OM	72	OM	62
LOSC	63	Bordeaux	61
Bordeaux	57	LOSC	55
ASSE	53	ASSE	53
Rennes	51	Montpellier	47
Evian	49	Nice	45
Toulouse	47	Lorient	45
Montpellier	46	Toulouse	45
Nantes	40	Bastia	43
Bastia	40	Caen	43
Reims	39	Nantes	42
Nice	38	Rennes	42
Lorient	38	Reims	41
Guingamp	38	Ajaccio	39
Metz	32	Guingamp	39
Caen	30	Angers	39
Lens	27	Troyes	36

Source : LFP & DNCG

## 7.2 Méthode IS appliquée dans un cadre de calcul bayésien

Soit  $f$  une densité *a posteriori* de l'écart de forces  $\delta$  entre deux équipes, sachant le résultat

$x$  de la rencontre  $[\delta|x]$  alors :  $U = \int u(\delta)[\delta|x]d\delta$

Pour  $g$ , on choisit le *prior*  $[\delta]$  :  $w_i = \frac{[x|\delta_i][\delta_i]}{[\delta_i][x]} = \frac{[x|\delta_i]}{[x]}$ .

L'estimateur IS approché est donc :  $\tilde{U} = \sum_{i=1}^{i=n} u(\delta_i)\tilde{w}_i$  avec  $\tilde{w}_i = \frac{w_i}{\sum_{i=1}^{i=n} w_i} = \frac{[x|\delta_i]}{\sum_{i=1}^{i=n} [x|\delta_i]}$ .

### 7.3 Exemple d'estimation de probabilités d'une journée

**Tableau A2 : espérance des probabilités avec *Prior* informatif (1<sup>ère</sup> journée, 2015-16)**

Buts	Domicile	Probabilités	<i>Prior</i> $\delta_{ij}$	ESS	Extérieur	Buts
1	OGC Nice	(0.01 ; 0.08 ; 0.91)	2.41	4.14	AS Monaco	2
2	Toulouse FC	(0.74 ; 0.20 ; 0.06)	1.28	2.49	AS Saint-Etienne	1
0	Montpellier Hérault SC	(0.04 ; 0.19 ; 0.77)	1.42	4.16	Angers SCO	2
1	FC Nantes	(0.81 ; 0.16 ; 0.03)	1.60	2.82	EA Guingamp	0
0	Olympique Lyonnais	(0.75 ; 0.20 ; 0.05)	1.35	3.15	FC Lorient	0
0	ESTAC Troyes	(0.06 ; 0.20 ; 0.74)	1.30	2.07	GFC Ajaccio	0
0	LOSC	(0.04 ; 0.18 ; 0.78)	1.46	4.94	Paris Saint-Germain	1
0	Olympique de Marseille	(0.73 ; 0.21 ; 0.06)	1.24	2.36	SM Caen	1
2	SC Bastia	(0.80 ; 0.17 ; 0.03)	1.57	3.44	Stade Rennais FC	1
1	Girondins de Bordeaux	(0.72 ; 0.22 ; 0.06)	1.19	2.19	Stade de Reims	2

Source : estimations de l'auteur

### 7.4 Statistiques descriptives des variables

**Tableau A3 : statistiques des variables (expliquée et explicatives)**

Variables (saison 2014-15)	Moyenne	$\sigma$	Min.	Max.
$t_i$	0.157	0.137	0	0.778
$ p_i^+ - p_i^- $	0.765	0.067	0.552	0.988
Température	8.289	5.076	-1.4	17.9
Rivalités	0.037	0.189	0	1
$I_{obs}^r$	0.922	0.133	0	1
Probabilité espérée de rattraper un concurrent	0.966	0.004	0.956	0.971

Variables (saison 2015-16)	Moyenne	$\sigma$	Min.	Max.
$t_i$	0.163	0.159	0	1
$ p_i^+ - p_i^- $	0.562	0.338	0.439	1
Température	8.289	3.831	4	16.6
Rivalités	0.042	0.201	0	1
$I_{obs}^r$	0.906	0.116	0	1
Probabilité espérée de rattraper un concurrent	0.959	0.002	0.957	0.965

Source : LFP & DNCG

## 8 Références bibliographiques

- Andreff, W. [2009], Équilibre Compétitif et Contrainte Budgétaire dans une Ligue de Sport Professionnel, *Revue Économique*, 60, 591-633.
- Andreff, W., Scelles, N. [2015], Walter C. Neale fifty years after: Beyond Competitive Balance, the League Standing Effect Tested with French Football Data, *Journal of Sports Economics*, 16, 819-34.
- Baranzini, A., Ramirez, J., Weber, S. [2008], The Demand for Football in Switzerland: an Empirical Estimation, *Cahier de Recherche, Haute École de Gestion*, Genève.
- Bourg, J-F. [2008], Les Très Hauts Revenus des Superstars du Sport. Un état des Approches Théoriques et Empiriques, *Revue d'Économie Politique*, 118, 375-94.
- Bradley, R.A., Terry M.E. [1952], Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons, *Biometrika*, 39, 324-45.
- Bouvet, P. [1996], Les Salaires des "Vedettes" du Sport Professionnel par Équipe : une Application Privilégiée de la Théorie du Salaire d'Efficiency ?, *Revue Française d'Économie*, 11, 119-44.
- Brandes, L., Franck, E. [2007], Who Made Who? An Empirical Analysis of Competitive Balance in European Soccer Leagues, *Eastern Economic Journal*, 33, 379-403.
- Czarnitzki, D., Stadtmann, G. [2002], Uncertainty of Outcome versus Reputation: Empirical Evidence for the First German Football Division, *Empirical Economics*, 27, 101-12.
- Davidson, R.R. [1970], On Extending the Bradley-Terry Model to Accomodate Ties in

Paired Comparison Experiments, *Journal of the American Statistical Association*, 65, 317-28.

Falter, J-M., Perignon, C. [2000], Demand for Football and Intramatch Winning Probability: an Essay on the Glorious Uncertainty of Sports, *Applied Economics*, 32, 1757-65.

Forrest, D., Simmons, R. [2002], Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer, *Journal of the Royal Statistical Society, Series D*, 51, 229-41.

Gayant, J-P., Le Pape, N. [2015], Mesure de la *Competitive Balance* dans les Ligues de Sports Professionnels, *Revue Économique*, 66, 427-48.

Gayant, J-P., Le Pape, N. [2016], Increasing Downside or Outer Risk? The Challenge of Measuring Competitive Imbalance in Closed and Open Leagues, *Southern Economic Journal*, 83, 774-95.

Goldstick, J. [2009], *Introduction to Statistical Computing*, University of Michigan.

Hill, J.R., Madura, J., Zuber, R.A. [1982], The Short Run Demand for Major League Baseball, *Atlantic Economic Journal*, 10, 31-35.

Hogan, V., Massey, P., Massey, S. [2013], Competitive Balance and Match Attendance in European Rugby Union Leagues, *Economic and Social Review*, 44, 425-46.

Jones, J.C.H. [1984], Winners, Losers and Hosers: Demand and Survival in the National Hockey League, *Atlantic Economic Journal*, 10, 54-63.

Kuper, S., Szymanski, S. [2016], *Les Attaquants les Plus Chers ne sont pas ceux qui Marquent le Plus*, De Boeck, Louvain-la-Neuve.

Kuypers, T. [1996], The Beautiful Game? An Econometric Study of Why People Watch English Football, *Discussion Paper in Economics*, 96-01, University College London.

Neale, W. [1964], The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition, *Quarterly Journal of Economics*, 78, 1-14.

Parent, E., Bernier, J. [2007], *Le Raisonnement Bayésien : Modélisation et Inférence*, Springer-Verlag, Paris.

Pawlowski, T., Anders, C. [2012], Stadium Attendance in German Professional Football: The (un)Importance of Uncertainty of Outcome Reconsidered, *Applied Economics Letters*, 19, 1553-56.



- Peel, D., Thomas, D. [1992], The Demand for Football: Some Evidence on Outcome Uncertainty, *Empirical Economics*, 17, 323-31.
- Rao, P.V., Kupper, L.L. [1967], Ties in Paired Comparison Experiments: A Generalization of the Bradley-Terry Model, *Journal of the American Statistical Association*, 62, 194-204.
- Robert, C.P., Casella, G. [2009], *Introducing Monte Carlo Methods with R*, Springer.
- Rottenberg, S. [1956], The Base-Ball Player's Labour Market, *Journal of Political Economy*, 64, 242-58.
- Scelles, N. [2017], Star Quality and Competitive Balance? Television Audience Demand for English Premier League Football Reconsidered, *Applied Economics Letters*, 24, 1399-402.
- Scelles, N., Durand, C., Bonnal, L., Goyeau, D., Andreff, W. [2013], Competitive Balance versus Competitive Intensity before a Match: is one of these two Concepts more Relevant in Explaining Attendance? The Case of the French Football Ligue 1 over the Period 2008-2011, *Applied Economics*, 45, 4184-92.
- Scelles, N., Durand, C., Bonnal, L., Goyeau, D., Andreff, W. [2016], Do all Sporting Prizes have a Significant Positive Impact on Attendance in a European National Football League? Competitive Intensity in the French Ligue 1, *Ekonomicheskaya Politika/Economic Policy*, 11, 82-107.
- Scherrer, B. [1984], *Biostatistiques*, G. Morin (Ed.), Chicoutimi, Québec.
- Scully, G.W. [1974], Pay and Performance in Major League Baseball, *American Economic Review*, 64, 915-30.
- Scully, G.W. [1989], *The Business of Major League Baseball*, University of Chicago Press, Chicago.