

DÉPARTEMENT D'ÉCONOMIQUE

Faculté d'administration

Université de Sherbrooke

EFFICIENCE DES JOUEURS DANS LA LIGUE NATIONALE DE HOCKEY

Par

BRUNO ÉMOND

Mémoire présenté à

VALÉRIE VIERSTRAETE

Juin 2016

## Table des matières

Liste des tableaux.....	3
Liste des figures .....	4
Liste des abréviations .....	5
Résumé.....	6
Abstract.....	7
Problématique .....	8
Méthodologie et revue de littérature.....	10
Choix de la méthodologie .....	10
Choix des inputs.....	23
Choix des outputs .....	25
Hypothèses supplémentaires .....	29
Bases de données et échantillon .....	34
Résultats.....	38
Rachats de contrat .....	48
Limites.....	58
Conclusion.....	61
Bibliographie .....	63
Annexe 1: Calcul de la contribution d'un joueur (Ryder, 2003) .....	67
Annexe 2 : Corrélacion entre les variables utilisées dans les régressions .....	72
Annexe 3 : Variable entrées une à une dans les régressions .....	74

## Liste des tableaux

Tableau 1. Association entre le nombre de points au classement par saison et plusieurs statistiques de performance d'une équipe (saisons 2007-2008 à 2010-2011) .....	27
Tableau 2. Statistiques descriptives sur les inputs et les outputs .....	37
Tableau 3. Statistiques descriptives sur l'efficience .....	40
Tableau 4. Statistiques descriptives sur le score d'efficience après bootstrap par position .....	44
Tableau 5. Association entre le nombre de points au classement par saison et l'efficience moyenne des joueurs de l'équipe en contrôlant pour les statistiques de performance de l'équipe (saisons 2007-2008 à 2010-2011) .....	46
Tableau 6. Exploration des variables ayant un lien potentiel avec le rachat de contrat .....	51
Tableau 7. Prédicteurs de la probabilité d'un rachat de contrat <sup>1</sup> .....	55
Tableau 8. Corrélation entre les variables utilisées aux tableaux 1 et 5 .....	72
Tableau 9. Corrélation entre les variables utilisées au tableau 7 .....	73
Tableau 10. Variables entrées une à une dans les régressions aux tableaux 1 et 5 .....	74
Tableau 11. Variables entrées une à une dans la régression au tableau 7 .....	76

## Liste des figures

Figure 1. Estimation de la frontière de production par les méthodes <i>SFA</i> et <i>DEA</i> .....	15
Figure 2. Biais de surestimation avec la méthode <i>DEA</i> .....	18
Figure 3. Persistance de la production par minute de jeu entre la saison courante et la saison précédente.....	22
Figure 4. Orientation input vs orientation output .....	30
Figure 5. Rendements d'échelle variables ( <i>VRS</i> ) vs rendements d'échelle constants ( <i>CRS</i> ).....	32
Figure 6. Invariance du score d'efficacité à une translation positive des outputs en <i>DEA</i> avec une orientation input et des rendements d'échelle variables .....	34
Figure 7. Distribution du score d'efficacité après bootstrap .....	39
Figure 8. Distribution du score d'efficacité avant et après bootstrap avec intervalle de confiance de 95% pour la saison 2007-2008.....	42
Figure 9. Corrélation entre le nombre de points au classement et l'efficacité moyenne des joueurs d'une équipe ( $\rho = 0,43$ ).....	45

## Liste des abréviations

BA	Buts alloués
BAJ	Buts alloués lorsque le joueur est sur la patinoire
BFD	Buts dans un filet désert
BM	Buts marqués
BMA	Buts marginaux empêchés par les autres joueurs
BME	Buts marginaux empêchés
BMG	Buts marginaux empêchés par le gardien de but
BMM	Buts marginaux marqués
BP	Buts pour
BR	Buts dont le joueur est responsable
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
LNH	Ligue Nationale de Hockey
MCO	Moindres carrés ordinaires
NBA	National Basketball Association
PAAQT	Pourcentage d'arrêts ajusté pour la qualité des tirs au but
PC	Player Contribution
PCD	Player Contribution Defense
PCG	Player Contribution Goaltending
PCO	Player Contribution Offense
SFA	Stochastic Frontier Analysis
T	Temps de jeu
TAB	Tirs au but
TBA	Taux de buts alloués
TCB	Taux de création de buts
TR	Taux de responsabilité
VRS	Variable returns to scale

## Résumé

En l'absence de mesure précise et unique de l'efficacité pour les joueurs de hockey, la présente étude a pour objectifs d'évaluer l'efficacité des joueurs dans la Ligue Nationale de Hockey (LNH) et de montrer comment celle-ci peut affecter la décision de racheter le contrat d'un joueur. Pour ce faire, les statistiques individuelles des joueurs de la LNH pour les saisons 2007-2008 à 2010-2011 sont utilisées. Pour estimer l'efficacité, la méthode de l'enveloppement de données (DEA) avec bootstrap est utilisée. Les inputs incluent le salaire et le nombre de minutes de jeu, alors que les outputs incluent la contribution défensive et offensive de chaque joueur. Pour estimer l'association entre l'efficacité individuelle et la probabilité d'un rachat de contrat, une régression logistique est utilisée. L'analyse des données montre que parmi 3 159 observations, l'efficacité moyenne est de 0,635. L'efficacité moyenne est similaire pour toutes les positions et toutes les saisons. Un lien positif et fort est trouvé entre le nombre de points au classement général d'une équipe et l'efficacité moyenne des joueurs qui la compose (coefficient de corrélation=0,43, valeur- $p < 0,01$ ). Les joueurs avec une efficacité plus élevée ont une probabilité plus faible de voir leur contrat racheté (rapport des chances=0,01, valeur- $p < 0,01$ ). La présente étude conclut donc que la plupart des joueurs de hockey dans la LNH ont un degré d'inefficacité non négligeable, qu'une efficacité plus élevée est associée à une meilleure performance au niveau de l'équipe et que les joueurs efficaces ont une probabilité plus faible de voir leur contrat racheté.

## Abstract

The objectives of the current study are to evaluate the efficiency of players in the National Hockey League and estimate its association with the probability of a contract buyout. To meet these objectives, individual statistics of NHL players for the 2007-2008 through 2010-2011 regular seasons are used. Efficiency is estimated using data envelopment analysis, with salary and minutes played as inputs, and offensive and defensive contributions as outputs. The association between efficiency and the probability of a buyout is estimated using a logistic regression. The data shows that among 3,159 observations, the mean efficiency is 0.635. The correlation between a team's standing points and its mean player efficiency is positive and strong ( $\rho=0.43$ ,  $p\text{-value}<0.01$ ). Players with higher efficiency are associated with a lower probability of a contract buyout (odds ratio=0.01,  $p\text{-value}<0.01$ ). Therefore, the current study concludes that higher player efficiency is associated with better team performance and lower buyout probability.

## Problématique

L'imposition d'un plafond sur la masse salariale des joueurs dans la Ligue Nationale de Hockey (LNH) après le lockout de 2004-2005 a changé la dynamique d'allocation des ressources dans le monde du hockey (Staudohar, 2005). L'idée était en partie d'introduire une plus grande parité entre les équipes au niveau des ressources disponibles. Ce faisant, la compétition serait plus égale et intense, ce qui donnerait un meilleur spectacle aux partisans. Désormais, les équipes les plus riches ne seraient pas constamment meilleures par le simple fait qu'elles attirent les meilleurs talents avec des salaires plus élevés. Ainsi, avec le plafond salarial, l'allocation *efficiente* des ressources rares est devenue une priorité. Il faut plus que jamais choisir des joueurs efficaces, c'est-à-dire qui peuvent «produire» le plus possible (compter et empêcher des buts) en utilisant le moins de ressources possible (temps de jeu et argent). Si l'efficacité devient un critère de sélection important, il faut donc pouvoir la mesurer de façon précise. Malgré cet important besoin de mesurer l'efficacité des joueurs, il n'existe pas de statistique unique résumant cette information. Certains sites de statistiques reliées au hockey rapportent des mesures partielles de l'efficacité (buts/salaire, passes/salaire, buts/minutes de jeu...), qui ne prennent pas en compte toutes les dimensions de la performance d'un joueur et de sa consommation de ressources. Le score d'efficacité qui sera calculé dans ce mémoire permettra de résoudre ce problème, car il sera unique et tiendra compte de toutes ces dimensions. Il permettra aussi aux gestionnaires des équipes de faire des choix plus éclairés. Par exemple, en fonction de son efficacité, il serait plus facile de voir si un joueur est susceptible d'être échangé, de voir son contrat racheté avant l'échéance ou de voir son salaire baisser ou

augmenter lors de la négociation de son prochain contrat. Cela permettrait aussi de mieux évaluer la qualité des joueurs sur le marché des agents libres.

L'émergence des statistiques avancées au hockey (Walthert, 2014) rend possible aujourd'hui une mesure plus précise de l'efficacité qui tient compte des aspects offensifs et défensifs d'un joueur, ce qui n'était pas possible auparavant, notamment au moment de la publication de l'article de Leibenstein et Maital (1992). En effet, à ce moment, il n'existait aucune statistique au hockey qui pouvait vraiment quantifier la qualité du jeu défensif d'un joueur. C'est pourquoi Leibenstein et Maital (1992) se sont limités à mesurer l'efficacité du jeu offensif des joueurs de hockey, qui pouvait alors être estimée à partir des buts et des passes. Comme l'attaque représente au plus 50% du jeu de hockey (une équipe doit se défendre lorsque l'autre équipe attaque), cette évaluation est évidemment incomplète si le but est d'évaluer l'efficacité sportive globale d'un joueur. Il n'est pas surprenant alors de découvrir que tous les défenseurs dans l'échantillon proposé à l'époque étaient fortement inefficaces. Même au niveau de l'attaque, la mesure proposée n'était pas aussi raffinée que ce qu'il est possible d'obtenir aujourd'hui.

L'objectif de ce mémoire sera donc d'évaluer, avec l'aide des statistiques avancées, l'efficacité des joueurs de hockey dans la LNH pour les saisons 2007-2008 à 2010-2011 et de montrer comment cette efficacité peut affecter certaines décisions prises par les gestionnaires des équipes, en particulier la décision de racheter le contrat d'un joueur.

## Méthodologie et revue de littérature

### Choix de la méthodologie

Un joueur de hockey peut être vu comme une firme utilisant des inputs pour produire des outputs, ce qui rend possible l'évaluation de son efficacité. Dans la théorie néo-classique du producteur, l'hypothèse souvent posée est qu'une firme (appelée « unité de décision », de façon générale) transforme de manière pleinement efficace ses inputs en outputs, ce qui signifie que toute entreprise se retrouve sur la frontière des possibilités de production. Or, dans la réalité, la plupart des entreprises ne sont pas pleinement efficaces. Il existe toujours certains processus qui peuvent être améliorés. Ainsi, contrairement à ce que dit la théorie, l'efficacité est l'exception plutôt que la règle. Il est alors nécessaire de mesurer le degré d'inefficacité, afin de découvrir qui sont les meilleures et les pires entreprises d'un secteur donné.

Dans l'évaluation de l'efficacité des entreprises, deux méthodes sont fréquemment utilisées. La première est l'analyse de frontière stochastique (*Stochastic Frontier Analysis, SFA*) développée par Aigner, Lovell et Schmidt (1977) et par Meeusen et van den Broeck (1977), basée sur l'approche économétrique traditionnelle (paramétrique). Dans cette méthode, pour une unité de décision  $i$ , l'output ( $y_i$ ) est une fonction des inputs ( $x_i$ ), pour lesquels il faut estimer le coefficient  $\beta$  s'y rattachant (selon une fonction  $f(x_i, \beta)$ , par exemple Cobb-Douglas). Le terme d'erreur ( $\varepsilon_i$ ) est quant à lui séparé en deux : une partie aléatoire usuelle ( $v_i \sim iid(0, \sigma_v^2)$ ) qui représente les événements externes et hors du contrôle de l'unité de décision et une partie tenant compte de l'inefficacité ( $u_i \sim iid(0, \sigma_u^2), u_i \geq 0$ ). Cette dernière partie indique que

l'unité de décision peut se retrouver sur la frontière de production en cas de pleine efficience ( $f(x_i, \beta), u_i = 0$ ) ou en-dessous de la frontière à cause d'un certain degré d'inefficience dans le processus de production ( $f(x_i, \beta) - u_i, u_i > 0$ ). Ainsi, le modèle général, qui peut être estimé par la méthode du maximum de vraisemblance, est le suivant :

$$y_i = f(x_i, \beta) + \varepsilon_i$$

$$\varepsilon_i = v_i - u_i \text{ où } u_i \geq 0$$

Dans le cas d'une fonction Cobb-Douglas, quelques manipulations permettent alors de trouver que le degré d'efficience de l'unité de décision  $i$  est donné par  $e^{-u_i}$ .

La deuxième méthode est l'enveloppement de données (*Data Envelopment Analysis, DEA*), développée par Charnes, Cooper et Rhodes (1978). Cette méthode est dite non paramétrique, car elle estime l'efficience d'une unité de décision via un programme d'optimisation. L'idée de base est qu'une unité de décision efficiente produit plus d'unités d'output par unité d'input que les autres unités de décision. Cela peut se traduire, dans un monde avec un input ( $x$ ) et un output ( $y$ ), par un ratio  $\frac{y}{x}$ . Cependant, la plupart des unités de décision utilise plusieurs inputs ( $x_1, \dots, x_m$ ) pour produire plusieurs outputs ( $y_1, \dots, y_s$ ). Si l'efficience doit être exprimée avec un seul ratio outputs/inputs, il faut alors additionner les inputs ensemble et les outputs ensemble :

$$Ratio = \frac{y_1 + \dots + y_s}{x_1 + \dots + x_m}$$

Cependant, cela suppose que chaque input et output a la même importance dans le processus de production. Si ce n'est pas le cas, il faut alors pondérer ces inputs et outputs selon leur importance (les poids sont donnés par  $u_r$  pour les outputs et  $v_i$  pour les inputs):

$$\text{Ratio} = \frac{u_1 * y_1 + \dots + u_s * y_s}{v_1 * x_1 + \dots + v_m * x_m} = \frac{\sum_{r=1}^s u_r * y_r}{\sum_{i=1}^m v_i * x_i}$$

Le choix des poids peut être arbitraire ou se baser sur l'information disponible, par exemple les prix. Or, dans plusieurs cas, dont celui à l'étude ici, les prix ne sont pas disponibles et il n'y a pas d'arguments suffisamment forts pour justifier le choix d'un poids par rapport à un autre. Il est alors possible de laisser un algorithme d'optimisation déterminer les poids de façon à maximiser le ratio  $\frac{\sum_{r=1}^s u_r * y_r}{\sum_{i=1}^m v_i * x_i}$  de chacune des unités de décision. Les poids qui solutionnent l'algorithme varient donc selon l'unité de décision pour laquelle le ratio est maximisé. Par exemple, pour un joueur de hockey, plus de poids sera accordé au volet offensif si celui-ci marque beaucoup de buts. Ce système de poids permet d'additionner des inputs et des outputs qui ne sont pas mesurés avec les mêmes unités.

Si il faut maximiser le ratio pour une unité de décision  $j$  ( $j = 1, \dots, n$ ), il faut imposer certaines contraintes, notamment que le ratio des autres unités de décision soit inférieur ou égal à 1 (une unité de décision est pleinement efficiente si son ratio est de 1) et que tous les poids soient positifs. Ainsi, pour chaque unité de décision, le problème d'optimisation est le suivant :

$$\text{Max}_{u_r, v_i} \frac{\sum_r u_r y_{r0}}{\sum_i v_i x_{i0}} \quad u: \text{ poids attribué à chaque output}$$

$$\text{Sous contrainte} \quad v: \text{ poids attribué à chaque input}$$

$$\frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1, j = 1, \dots, n \quad r: \text{ indice numérotant chaque output}$$

$$u_r, v_i \geq 0 \text{ pour tout } r, i \quad i: \text{ indice numérotant chaque input}$$

$j$ : indice numérotant chaque unité de décision

(l'indice 0 représente l'unité de décision dont le score d'efficacité est à évaluer). Pour une année donnée, une unité de décision représente un joueur de hockey.

$n$ : nombre d'unités de décision dans l'échantillon

Il y a donc  $n$  problèmes d'optimisation à résoudre, soit un par unité de décision. Au terme de la résolution du problème pour l'unité de décision sous évaluation (l'unité de décision 0), celle-ci obtient son score d'efficacité. Cet algorithme maximise cependant une fraction, ce qui pose problème. En linéarisant le problème et en considérant la forme duale du problème de maximisation, il est possible d'obtenir un algorithme riche en information pour déterminer l'efficacité d'une unité de décision, soit le modèle d'enveloppement<sup>1</sup> :

$$\min_{\theta, \lambda} \left[ \theta - \varepsilon * \left( \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right) \right]$$

*Sous contrainte*

$$\sum_{j=1}^n x_{ij} * \lambda_j + s_i^- = \theta * x_{i0} \quad i = 1, \dots, m$$

$$\sum_{j=1}^n y_{rj} * \lambda_j - s_r^+ = y_{r0} \quad r = 1, \dots, s$$

---

<sup>1</sup> Les hypothèses reliées à l'orientation du modèle et aux rendements d'échelle seront discutées dans la section « Hypothèses supplémentaires ». Les équations ci-dessus sont reliées au modèle DEA avec orientation input et rendements d'échelle constants.

$$\lambda_j \geq 0 \quad j = 1, \dots, n$$

Ici, les  $\lambda_j$  sont des poids virtuel. Pour une unité de décision inefficente,  $\lambda_j > 0$  signifie que l'unité de décision évaluée devrait s'inspirer de l'unité de décision pleinement efficiente  $j$  et la taille de  $\lambda_j$  indique la proportion dans laquelle l'unité de décision inefficente devrait s'inspirer de  $j$  dans son atteinte de la pleine efficience (une unité de décision peut avoir plusieurs candidats pleinement efficients qu'elle peut prendre en exemple pour s'améliorer).  $\theta$  est le score d'efficience.  $\varepsilon > 0$  est une constante non-archimédienne plus petite que n'importe quel nombre réel positif qui permet de résoudre le problème d'optimisation d'abord pour  $\theta$ , puis pour les  $\lambda_j$  (en fixant  $\theta = \theta^*$ ). Finalement,  $s_i^-$  et  $s_r^+$  représentent les « slacks » en inputs et en outputs, respectivement. Par exemple, si  $\theta = 1$  mais que  $s_i^-$  est positif, alors l'unité de décision en question est faiblement efficiente, c'est-à-dire qu'elle se trouve sur la frontière efficiente, mais qu'il est possible d'utiliser moins de l'input  $i$  tout en maintenant les outputs constants.

La **figure 1** montre comment les méthodes *SFA* et *DEA* estiment la frontière de production dans un contexte de production où il y a un input et un output. Dans le cas de la méthode *DEA*, la frontière de production est formée par l'enveloppe convexe des unités de décision utilisant le moins d'unités d'input pour produire le plus d'unités d'output. Les unités de décision A, B et C définissent cette frontière et sont par conséquent pleinement efficients, c'est-à-dire qu'elles ont un score d'efficience de 1. Les unités de décision D et E sont à l'intérieur de l'ensemble de production délimité par la frontière de production, mais ne sont pas situées sur la frontière de production, ce qui signifie qu'elles sont inefficientes. Leur degré d'inefficience est donné par la distance qui les sépare de la frontière de production. Quant à elle, la méthode *SFA* n'enveloppe

pas complètement les données et permet à certaines unités de décision d'être au-dessus de la frontière de production estimée, grâce à une certaine dose de « chance » reliée à des facteurs externes au processus de production. Cette chance est représentée par le terme d'erreur  $v_i$  de la fonction de régression. La courbe de régression *SFA* passe « au-dessus » de la courbe de régression estimée par les moindres carrés ordinaires (MCO) à cause de la composition différente du terme d'erreur dans le modèle *SFA*.

**Figure 1. Estimation de la frontière de production par les méthodes *SFA* et *DEA***

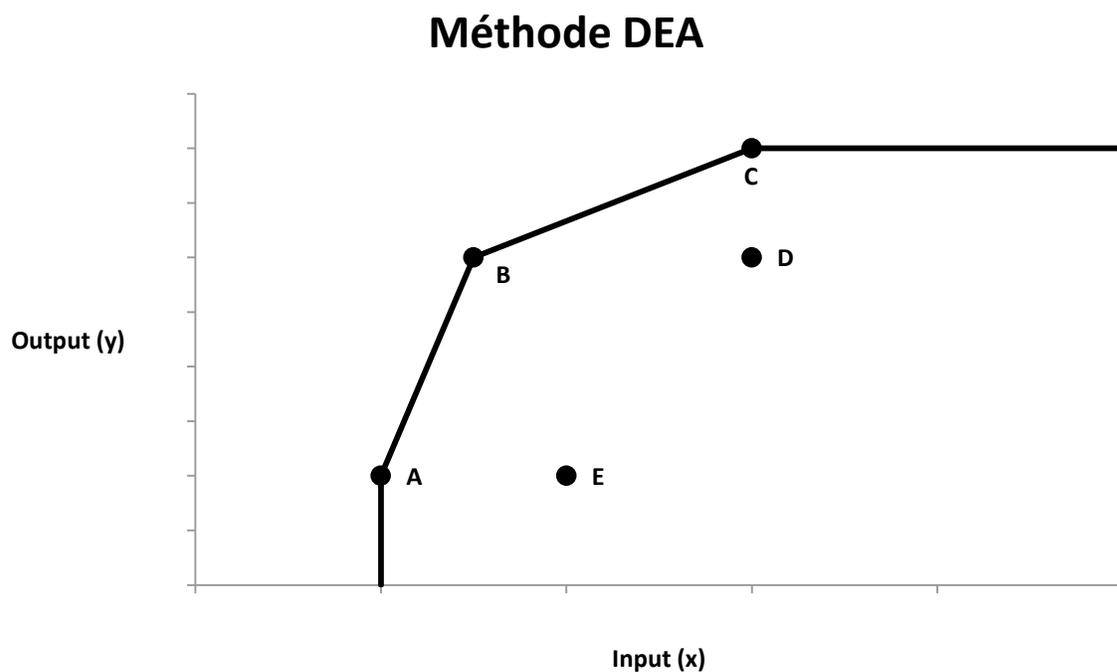
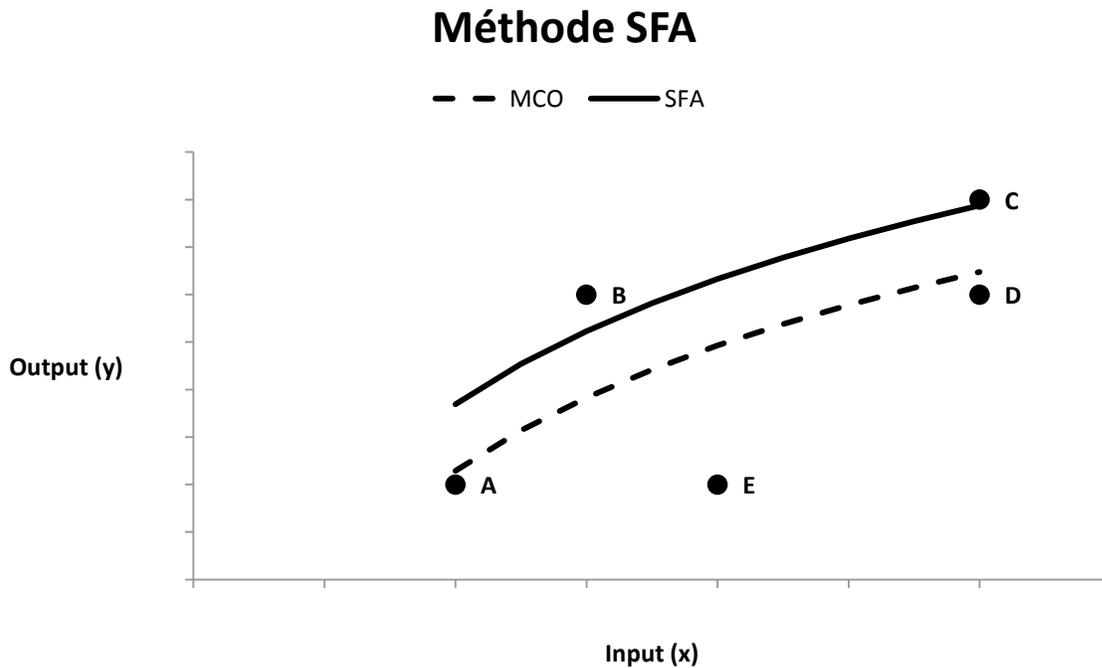


Figure 1 (suite)



Par rapport à la méthode *SFA*, la méthode *DEA* offre plus de flexibilité, car elle n'impose pas d'hypothèses sur la forme fonctionnelle de la fonction de production (hormis la libre disposition des inputs/outputs et la convexité de l'ensemble de production) et sur la distribution des termes d'erreur, car il n'y en a pas. Cette dernière s'adapte aussi plus facilement à un contexte multi-outputs, car en comparaison, la méthode *SFA* en situation multi-outputs nécessite l'utilisation de fonctions de distance, ce qui implique des hypothèses additionnelles (par exemple, que ces fonctions soient concaves en inputs et quasi-concaves en outputs) qui ne sont pas toujours respectées, en plus de possiblement créer un problème d'endogénéité (Coelli *et al.*, 2005). L'absence de terme d'erreur dans la méthode *DEA* signifie cependant que toute production ne peut pas être le fruit du hasard et est déterminée uniquement par le processus de production de l'unité de décision et la quantité d'inputs utilisée. Cela constitue une faiblesse

par rapport à la méthode *SFA*. La méthode *DEA* souffre également du problème de la « malédiction de la dimension », surtout dans les petits échantillons, c'est-à-dire que plus le nombre d'inputs et d'outputs est élevé, plus il est facile de trouver une certaine combinaison de poids qui montre l'unité de décision comme étant pleinement efficiente (Cooper, Seiford et Zhu, 2011). Elle a aussi de la difficulté à gérer les inputs et outputs négatifs ou indésirables (comme la pollution). Il existe cependant des solutions à ces problèmes (voir par exemple, Dyson *et al.*, 2001). Une autre difficulté liée à la méthode *DEA* est que l'efficacité estimée est une mesure relative qui dépend des unités de décision incluses dans l'échantillon. En effet, toutes les unités de décision sont évaluées par rapport aux unités de décision formant la frontière de production. De plus, l'ensemble de production estimé à partir d'un échantillon d'unités de décision est nécessairement un sous-ensemble du véritable ensemble de production, ce qui signifie que l'efficacité est surestimée. Heureusement, ce biais peut être corrigé par le bootstrap (Simar et Wilson, 2000; Bogetoft et Otto, 2011; Simm et Besstremyannaya, 2015). Ici, le bootstrap de Simm et Besstremyannaya (2015), basé sur l'algorithme de Simar et Wilson (2000), est utilisé<sup>2</sup>.

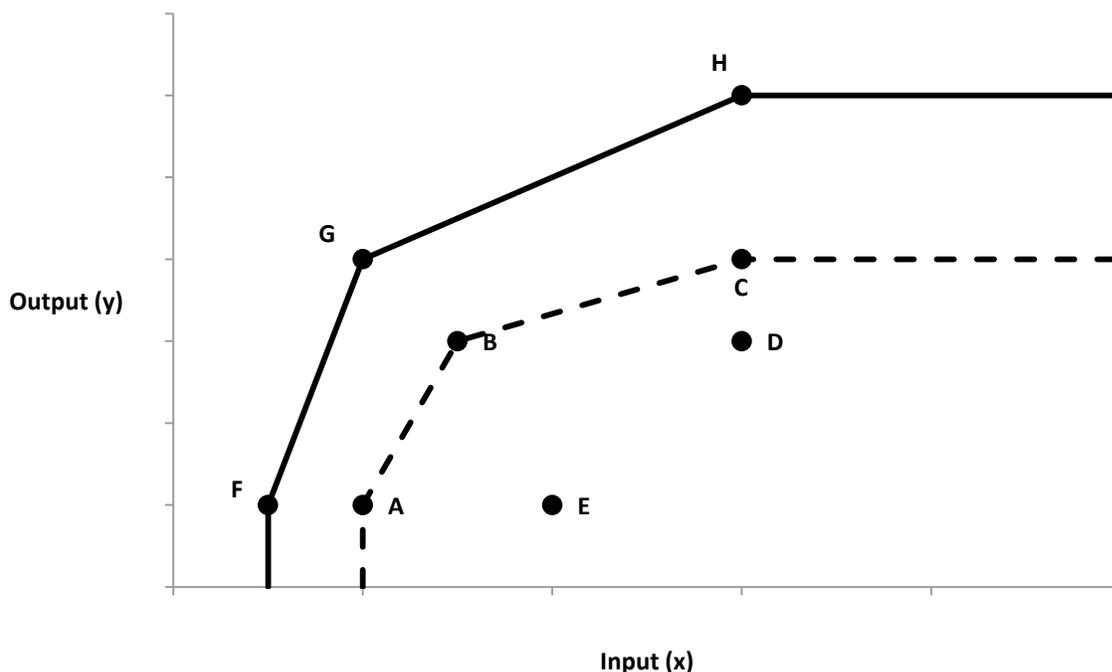
La **figure 2** illustre la problématique liée au problème du biais de surestimation dans un monde avec un input et un output. Si les unités de décision A à E sont incluses dans l'échantillon, la frontière de production estimée (en pointillée) est formée par les unités de décision A, B et C, qui sont les unités les plus efficaces de l'échantillon. Cependant, si la population totale inclut également les unités de décision F, G et H, qui ne sont pas incluses dans

---

<sup>2</sup> L'algorithme de Simar et Wilson (2000) n'a pas été utilisé, car il a été impossible d'obtenir la licence pour pouvoir utiliser le code *R* qui effectue la procédure de bootstrap. L'algorithme de Simm et Besstremyannaya (2015) utilise cependant la même procédure dans *R* et est accessible sans licence. L'algorithme de Bogetoft et Otto (2011) génère quant à lui des résultats erronés dans *R*.

l'échantillon pour différentes raisons (par exemple, parce que leur input et leur output sont difficilement mesurables), et que ces unités sont plus efficaces que les unités A à E, la frontière de production estimée avec les unités A à E est en fait située à l'intérieur de l'ensemble de production de la véritable frontière de production (ligne pleine) estimée avec les unités A à H. L'efficacité estimée seulement avec les unités A à E est alors surestimée par rapport à la situation où toutes les unités sont incluses. Par exemple, les unités A, B et C ont une efficacité de 1 dans l'échantillon original, mais un certain degré d'inefficacité lorsque les unités F, G et H sont incluses, car une certaine distance les sépare de la véritable frontière de production.

**Figure 2. Biases de surestimation avec la méthode DEA**



De façon simple et imagée, l'idée générale du bootstrap en DEA est donc de produire des unités de décision simulées à partir desquelles la véritable frontière de production est estimée et de

comparer les unités de décision de l'échantillon original par rapport à cette nouvelle frontière.

Les étapes précises (selon l'algorithme de Simm et Besstremyannaya, 2015) pour obtenir un score d'efficacité corrigé pour le biais de surestimation via la méthode du bootstrap sont les suivantes :

1. Estimer l'efficacité de toutes les unités de décision par la méthode DEA.  $\hat{\theta}_j$  est alors obtenu pour  $j = 1, \dots, n$  où  $n$  est le nombre d'unités de décision (joueurs de hockey) dans l'échantillon.
2. Tirer aléatoirement avec remise  $n$  scores d'efficacité de  $\hat{\theta}_1, \dots, \hat{\theta}_n$  (appelés  $\beta_1, \dots, \beta_n$ ).
3. Tirer  $n$  nombres d'une distribution normale standard (appelés  $\varepsilon_1, \dots, \varepsilon_n$ ).
4. Appliquer la méthode du lissage pour avoir une distribution imitant le plus possible une distribution continue :

$$\tilde{\theta}_j = \begin{cases} \beta_j + h * \varepsilon_j & \text{si } \beta_j + h * \varepsilon_j \leq 1 \\ 2 - \beta_j - h * \varepsilon_j & \text{sinon} \end{cases}$$

où  $h$  est le paramètre de lissage déterminé par la méthode de la validation croisée.

5. Ajuster  $\tilde{\theta}_j$  afin d'obtenir une variance asymptotiquement correcte :

$$\theta_j^* = \bar{\beta} + \frac{1}{\sqrt{1 + \frac{h^2}{\hat{\sigma}^2}}} * (\tilde{\theta}_j - \bar{\beta})$$

où  $\bar{\beta} = \frac{1}{n} * \sum_{j=1}^n \beta_j$ ,  $\hat{\sigma}^2 = \frac{1}{n} * \sum_{j=1}^n (\hat{\theta}_j - \bar{\theta})^2$  et  $\bar{\theta} = \frac{1}{n} \sum_{j=1}^n \hat{\theta}_j$

6. En orientation input<sup>3</sup>, ajuster les inputs selon l'efficacité ajustée  $\theta_j^*$  :

$$x_j^b = \frac{\hat{\theta}_j}{\theta_j^*} * x_j$$

---

<sup>3</sup> Les détails de l'orientation du modèle seront discutés dans la section « Hypothèses supplémentaires »

7. Estimer l'efficacité de toutes les unités de décision par la méthode *DEA* en utilisant les inputs ajustés  $x_j^b$  comme référence.
8. Répéter  $B$  fois les étapes 2 à 7. Chaque unité de décision a donc  $B$  scores d'efficacité simulés  $\theta_j^b, \dots, \theta_j^B$ . Dans le cadre du présent mémoire,  $B = 1000$ . Ce choix est en accord avec ce que suggèrent Simar et Wilson (1998).
9. Calculer l'efficacité corrigée par le biais de surestimation :

$$Biais_j = \left( \frac{1}{B} * \sum_{b=1}^B \theta_j^b \right) - \hat{\theta}_j$$

$$\hat{\theta}_j^* = \hat{\theta}_j - Biais_j$$

Comme Simar et Wilson (2000) le mentionnent, même si  $B$  tend vers l'infini,  $\hat{\theta}_j^*$  a une variance quatre fois plus grande que  $\hat{\theta}_j$ . Ainsi, pour chaque unité de décision, la correction pour le biais devrait être appliquée seulement si  $Var(\theta_j^b) < \frac{Biais_j^2}{3}$ .

La méthode *DEA* semble un choix plus approprié dans le contexte actuel, car la forme fonctionnelle de la fonction de production et la distribution des termes d'erreur sont inconnues pour les joueurs de hockey. De plus, deux inputs et trois outputs sont identifiés<sup>4</sup> et il est bien connu que la méthode *DEA* se prête mieux à un contexte multi-outputs. Dans le contexte actuel, le problème de la malédiction de la dimension est évité, car l'échantillon est suffisamment grand ( $n \approx 800$  joueurs par saison) par rapport au minimum proposé dans la littérature. En effet, celle-ci propose d'avoir  $n \geq 2 \times inputs \times outputs$  unités de décision (Dyson *et al.*, 2001) ou  $n \geq \max(inputs \times outputs, 3(inputs + outputs))$  (Cooper, Seiford et

---

<sup>4</sup> Le choix des inputs est détaillé dans la section « Choix des inputs », alors que celui des outputs est détaillé dans la section « Choix des outputs ».

Zhu, 2011). Il reste donc le problème de l'absence de terme aléatoire associé à la fonction de production de la méthode *DEA* (ce terme permettrait de tenir compte des évènements externes et hors de contrôle pour l'unité de décision). Il est clair que le facteur chance joue un rôle dans le processus de production d'un joueur de hockey. Cependant, l'importance de ce facteur est minimisée pour les joueurs ayant joué suffisamment de parties au cours d'une saison. Il faut donc un temps de jeu minimal pour s'assurer un minimum de persistance dans la production, laissant ainsi le moins de place possible au facteur chance. La plupart des articles dans la littérature (par exemple, Leibenstein et Maital, 1992, hockey et Cooper, Ruiz et Sirvent, 2009, basketball) imposent un temps de jeu minimum allant jusqu'à 50% de parties jouées sur le maximum possible en saison régulière. Ici, afin d'inclure le plus de joueurs possible, un minimum de 200 minutes de jeu dans une saison est imposé, ce qui représente 10 minutes par partie pour 20 parties (le maximum de parties dans une saison est 82 et le maximum de minutes par partie est 60). L'étude de la production (selon les trois outputs identifiés: *Player Contribution Offense [PCO]*, *Player Contribution Defense [PCD]* et *Player Contribution Goaltending [PCG]*<sup>5</sup>) de cet échantillon de joueurs à travers le temps montre qu'il existe une certaine persistance entre la production de la saison courante et la production de la saison précédente (**figure 3**). Cependant, il semble y avoir encore une certaine part d'aléatoire dans la production d'un joueur, comme le montrent les coefficients de corrélation obtenus (*PCO* :  $\rho = 0,69$ , *PCD* :  $\rho = 0,73$ , *PCG* :  $\rho = 0,26$ ). Malgré cette faiblesse, la méthode *DEA* demeure la plus attrayante à cause de sa meilleure capacité d'adaptation à un contexte multi-outputs et son absence d'hypothèses restrictives sur la forme fonctionnelle de la fonction de production.

---

<sup>5</sup> Les détails sont discutés dans la section « Choix des outputs ».

Figure 3. Persistance de la production par minute de jeu entre la saison courante et la saison précédente

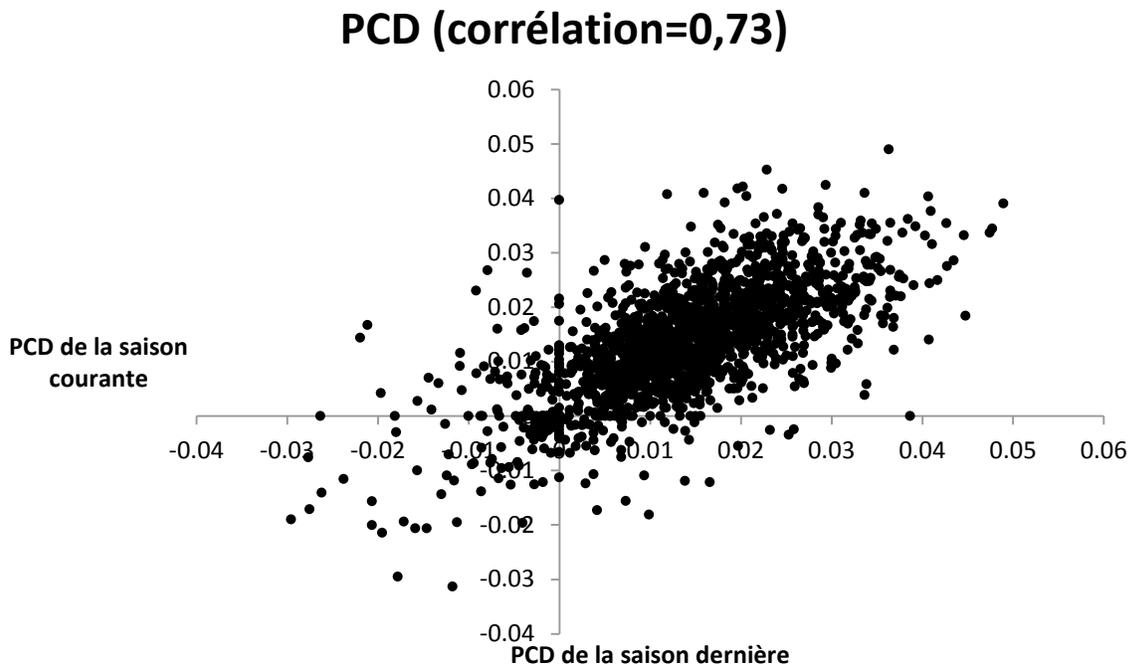
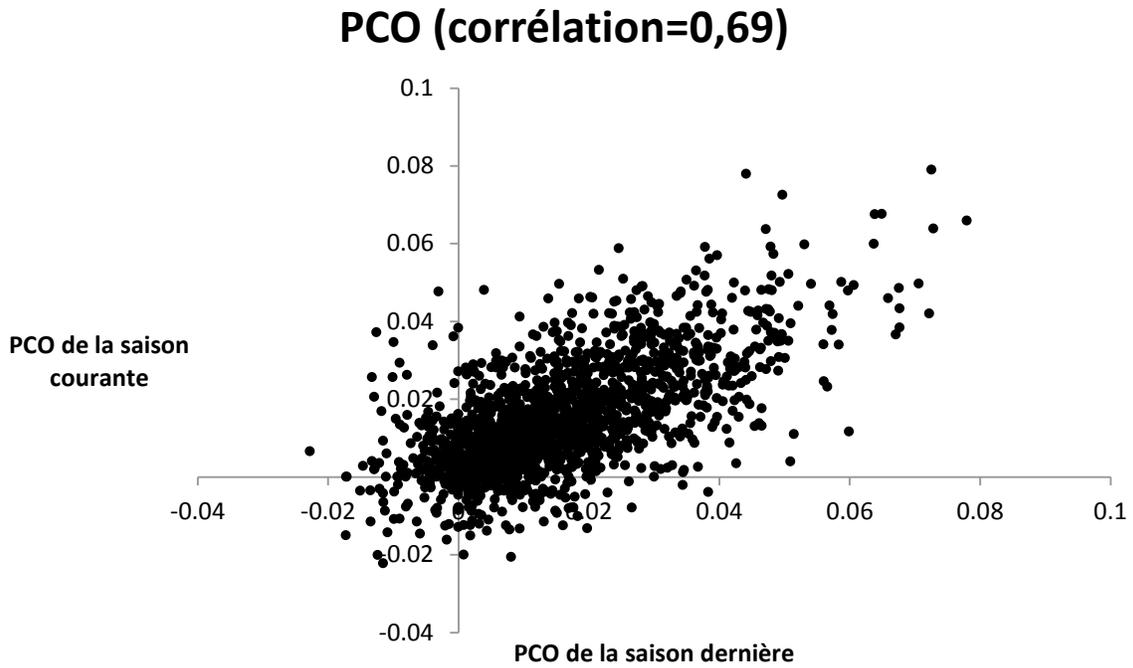
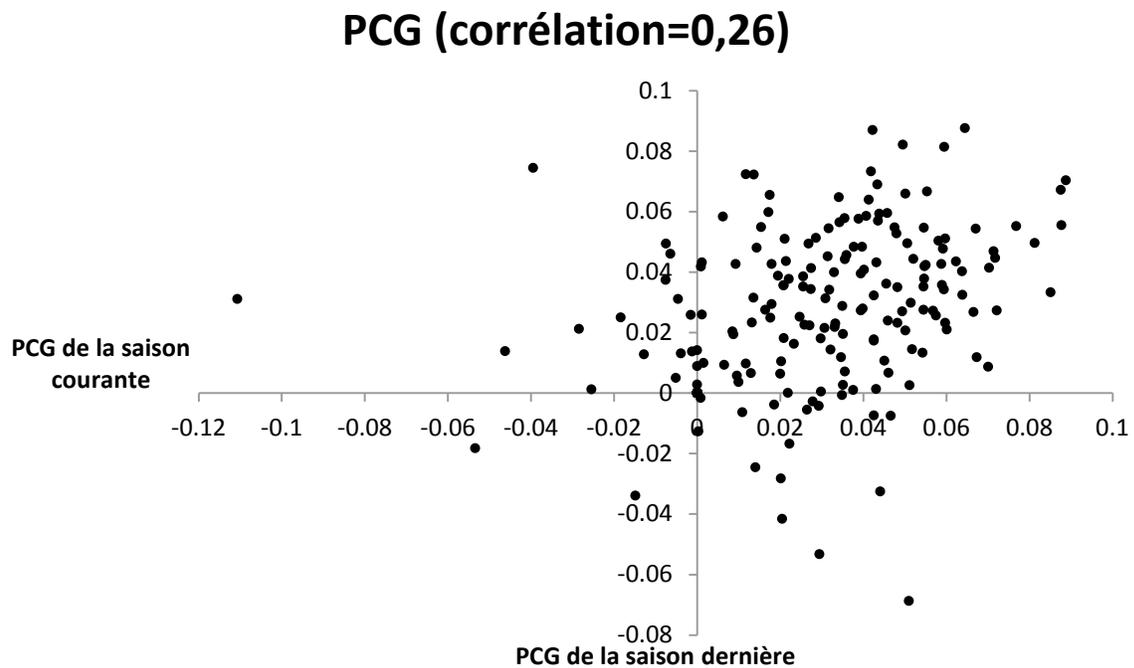


Figure 3 (suite)



### Choix des inputs

Dans la littérature sportive générale, deux inputs semblent importants lorsque vient le temps d'évaluer l'efficacité d'un joueur: le talent et l'opportunité de produire (en ce sens où pour avoir une production positive, un joueur doit avoir l'opportunité de produire, donc avoir un temps de jeu positif). Au hockey, plus un joueur a un niveau élevé de ces deux caractéristiques, plus il devrait compter et empêcher des buts.

La littérature diverge cependant sur la façon de mesurer ces éléments. Pour ce qui est de l'opportunité de produire, Halkos et Tzeremes (2012, tennis) croient qu'il faut considérer le nombre de parties jouées, alors que Radovanović *et al.* (2013, basketball) penchent plutôt pour

le nombre de minutes jouées et Leibenstein et Maital (1992, hockey) pour le nombre de tirs au but. Le choix approprié dépend du sport. Dans le cas du hockey, le nombre de tirs au but est une mesure incomplète, car elle ne mesure que l'opportunité de produire offensivement. Or, un joueur peut aussi produire défensivement en empêchant des buts, ce qui est tout aussi important que d'en marquer. Ensuite, le nombre de parties jouées constitue une mesure plus imprécise de l'opportunité de produire que le nombre de minutes de jeu, car deux joueurs peuvent jouer le même nombre de parties, mais pas le même nombre de minutes par partie. L'opportunité de produire est ainsi plus importante pour celui jouant un plus grand nombre de minutes par partie. Pour ces raisons, le nombre de minutes sera la mesure privilégiée dans ce mémoire pour quantifier l'opportunité de produire.

Pour mesurer le talent, Radovanović *et al.* (2013, basketball) ainsi que Leibenstein et Maital (1992, hockey) pensent qu'il faut utiliser le salaire. Cependant, le salaire pose problème car il est endogène, c'est-à-dire qu'il dépend de la performance passée, donc de l'output passé. Ce problème ne se pose pas pour tous les joueurs, car ceux qui ont signé des contrats à long terme voient leur salaire fixé pour plusieurs années, sans possibilité de variation suite à de bonnes ou mauvaises années. Par contre, les gestionnaires des équipes peuvent faire des erreurs dans l'évaluation du talent d'un joueur. Ainsi, le salaire n'est qu'une mesure approximative du talent réel. En fait, le salaire est davantage une mesure anticipée du talent et des habiletés d'un joueur. Une alternative au salaire serait de mesurer chaque composante du talent d'un joueur. Cependant, cela devient rapidement difficile, car il y a d'innombrables facettes au talent : la capacité d'anticipation, la précision du tir, la précision des passes, la vision du jeu, l'agressivité, etc. Cela pourrait donc entraîner une diminution considérable de la puissance statistique de

l'analyse effectuée ou un problème de surplus de dimensions dans un modèle non paramétrique. De plus, certaines de ces composantes sont très difficiles à mesurer. Une autre alternative au salaire pour mesurer le talent pourrait être le rang auquel un joueur a été sélectionné au repêchage, la logique étant qu'un joueur talentueux est repêché plus rapidement. Par contre, les joueurs de la LNH ne sont pas tous sélectionnés au repêchage. De plus, le choix au repêchage est fixe à travers le temps, alors que le talent peut varier avec les années. Par exemple, il est bien connu qu'avec l'âge, les capacités physiques d'un joueur, et par conséquent ses habiletés, diminuent. Bien qu'imparfait, le salaire réussit à mieux refléter cet aspect, car celui-ci varie avec le temps. Enfin, au moment du repêchage, l'incertitude quant au talent potentiel d'un joueur est à son maximum, ce qui signifie que le choix au repêchage est une mesure très imprécise du talent. Donc, malgré certaines faiblesses, le salaire semble la meilleure option parmi les choix possibles pour mesurer le talent d'un joueur. Si le salaire représente les anticipations quant au talent et aux habiletés d'un joueur, il faut alors inclure dans cette mesure tout ce qui est déterminé ex-ante dans le contrat. Cela signifie qu'il faut inclure, en plus du salaire de base, le bonus à la signature et le bonus à la performance. En effet, ces deux éléments sont fixés avant même le début de la saison régulière, comme le confirme Tolensky (2008).

### **Choix des outputs**

Dans l'optique où il faut évaluer l'efficacité sportive d'un joueur, il est important d'avoir le portrait le plus complet possible de sa production. Au hockey, cela signifie qu'il faut tenir compte autant de la contribution offensive que défensive d'un joueur. Cependant, les statistiques traditionnelles de hockey n'arrivent pas à bien capter la composante défensive, car

les tirs bloqués, les revirements ou les mises en échec (des mesures défensives traditionnelles) n'ont pas de lien statistique avec le nombre de points au classement d'une équipe ou la probabilité de gagner. En fait, un seul facteur est fortement associé avec le succès d'une équipe, soit le différentiel de buts pour et de buts contre (Ryder, 2008). Le **tableau 1** appuie cette affirmation. Celui-ci montre pour les saisons 2007-2008 à 2010-2011 une simple régression par les moindres carrés ordinaires du nombre de points au classement d'une équipe sur de nombreux facteurs pouvant y être lié comme le différentiel de buts pour/contre, les mises en échec, les tirs bloqués, etc. Clairement, il existe un lien fort et positif entre le différentiel de buts pour/contre et le nombre de points au classement d'une équipe (coefficient=0,3477, valeur- $p < 0,001$ ). Un pourcentage de mises en jeu gagnées élevé semble aussi être associé à un nombre plus élevé de points au classement (coefficient=0,432, valeur- $p = 0,054$ ), mais cette relation est si faible qu'on peut l'ignorer, car aucune équipe a un pourcentage de mises en jeu gagnées suffisamment élevé pour y associer un changement important dans le nombre de points au classement<sup>6</sup>.

---

<sup>6</sup> Gabriel Desjardins (2011) estime qu'il faudrait un différentiel de mises en jeu gagnées/perdus de +245 pour obtenir deux points de plus au classement, ce qui correspond à une victoire de plus. Seulement 18 équipes (7,5% des observations de l'échantillon utilisé) ont atteint un tel différentiel sur une saison complète.

**Tableau 1. Association entre le nombre de points au classement par saison et plusieurs statistiques de performance d'une équipe (saisons 2007-2008 à 2010-2011)**

**Nombre d'équipes: n=120**

<b>Variable</b>	<b>Coefficient</b>	<b>Erreur standard</b>	<b>Valeur-p</b>
Constante	80,1280	18,0181	<0,001*
Différentiel de buts pour/contre	0,3477	0,0169	<0,001*
% d'efficacité avantage numérique	-0,1558	0,1880	0,409
% d'efficacité désavantage numérique	-0,1236	0,1670	0,461
% d'efficacité mises en jeu	0,4320	0,2216	0,054
Mises en échec	0,0021	0,0016	0,183
Tirs bloqués	-0,0040	0,0032	0,215
Pertes de possession	0,0039	0,0028	0,156
Reprises de possession	0,0014	0,0038	0,717

\* indique une valeur-p < 0,05

Note: Les corrélations entre les variables utilisées et le détail des résultats en entrant les variables une à une dans la régression sont disponibles aux annexes 2 et 3.

Ainsi, pour voir quel joueur contribue le plus au succès d'une équipe (en termes de victoires ou de points au classement), il faut mesurer sa capacité à marquer et à prévenir des buts, car cela influence directement le différentiel de buts pour/contre au niveau de l'équipe. Les statistiques traditionnelles ont essayé, en vain, de mesurer cette capacité via la statistique du plus/moins, qui correspond au différentiel de buts pour/contre de l'équipe lorsque le joueur est sur la patinoire. Cette mesure souffre de plusieurs lacunes, mises en évidence par Lipson (2014) et Staples (2013), pour ne nommer que ceux-là. Entre autres, elle ne tient pas compte du degré de responsabilité du joueur dans le but compté ou alloué, de la situation de jeu (avantage numérique, désavantage numérique, jeu à cinq contre cinq) et de la qualité des coéquipiers, des adversaires et des gardiens de but lorsque le joueur est sur la glace. Les statistiques avancées de hockey peuvent mesurer la capacité à marquer et à prévenir des buts de façon

beaucoup plus précise que le plus/moins. Un bon exemple constitue la mesure de contribution du joueur (*Player Contribution, PC*) d'Alan Ryder (2003). L'idée est de comparer chaque joueur à un joueur marginal, tout juste assez bon pour jouer dans la LNH. Tout ce qu'un joueur produit en excès des capacités du joueur marginal correspond à sa contribution à la victoire (la contribution du joueur marginal est nulle). Chaque joueur (autre que le gardien de but) peut contribuer de deux façons: marquer des buts et prévenir des buts. Pour marquer des buts, il est possible de compter le but ou d'initier le jeu qui mène au but (passes). Un joueur peut prévenir des buts en minimisant la quantité et la qualité des tirs par un bon positionnement sur la patinoire, alors qu'un gardien de but peut prévenir des buts en arrêtant les tirs au but. Par conséquent, trois mesures caractérisent la contribution de chaque joueur au nombre de victoires d'une équipe : contribution offensive (*Player Contribution Offense, PCO*), contribution défensive (*Player Contribution Defense, PCD*) et contribution en tant que gardien (*Player Contribution Goaltending, PCG*). Ces trois éléments constituent les outputs d'un joueur de hockey. Le détail des calculs pour obtenir *PCO*, *PCD* et *PCG* est donné à l'annexe 1.

Un joueur autre que le gardien de but peut contribuer seulement de deux façons, soit via sa contribution offensive (*PCO*) et défensive (*PCD*). Pour ces joueurs,  $PCG = 0$  en tout temps. Par contre, le gardien de but peut contribuer selon les trois mesures. En effet, un gardien peut à l'occasion faire une passe qui mène à un but ou même compter un but, ce qui est très rare. La mesure *PCO* sera alors faiblement positive. De plus, celui-ci peut prendre des pénalités, ce qui affecte négativement sa contribution défensive (*PCD*)<sup>7</sup>. Ainsi, la mesure *PCD* pour un gardien

---

<sup>7</sup> Ryder (2003) montre comment il est possible de décomposer la mesure *PCD* en ses composantes, dont les pénalités. Prendre des pénalités donne l'occasion à l'autre équipe de compter plus facilement un but, car l'équipe

de but ne peut pas être positive, car les pénalités représentent le seul élément défensif qu'il peut apporter. Finalement, la majeure partie de la contribution d'un gardien provient de sa capacité à arrêter des rondelles, ce qui est capturé par la mesure *PCG*.

Les mesures de contribution du joueur présentent cependant certaines lacunes. En particulier, un joueur peut contribuer à la victoire grâce à son leadership auprès de ses pairs. Le fait d'avoir une bonne chimie entre les joueurs, combinée à un système de jeu mettant en valeur les forces des joueurs, peut stimuler la contribution de ceux-ci. Le concept d'équipe est donc complètement absent de cette analyse. Les joueurs obtiennent également tout le crédit de leur performance, alors que les entraîneurs peuvent y être pour quelque chose. De plus, les mesures de contribution se limitent à la saison régulière et ne tiennent pas compte des séries éliminatoires. Or, certains joueurs sont reconnus pour leur contribution supérieure durant ces séries.

En résumé, un joueur utilise son talent, mesuré par le salaire, et l'opportunité qui s'offre à lui, mesurée par le temps de jeu, pour créer et prévenir des buts, ces éléments étant mesurés par les mesures de contribution *PCO*, *PCD* et *PCG*.

## Hypothèses supplémentaires

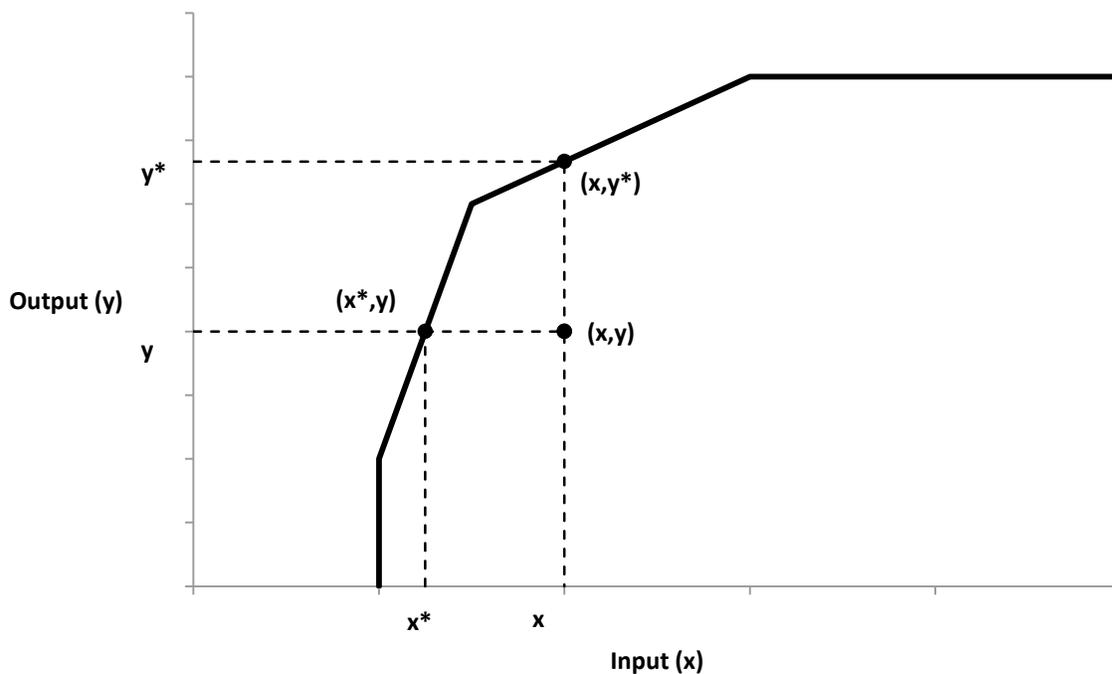
Bien que flexible, la méthode *DEA* nécessite tout de même certains choix. En particulier, l'orientation peut être input ou output, ce choix affectant l'interprétation du score d'efficacité et la façon de le calculer. Dans une orientation input, une unité de décision peut améliorer son efficacité par une diminution proportionnelle de ses inputs, tout en maintenant un niveau de

---

prise en défaut doit se débrouiller avec un joueur en moins pour deux minutes ou plus. Cela signifie que les pénalités ont un impact négatif sur la mesure *PCD*.

production constant. Dans un exemple avec un input et un output (voir **figure 4**), cela signifie que le degré d'inefficience est évalué par rapport à l'axe des inputs, soit l'axe des abscisses. Ainsi, le degré d'inefficience de l'unité de décision utilisant  $x$  unités d'input et  $y$  unités d'output est donné par la distance qui sépare les coordonnées  $(x, y)$  des coordonnées  $(x^*, y)$  situées sur la frontière de production. Si  $x^*$  représente 80% de  $x$ , alors il est possible d'affirmer que l'unité de décision pourrait utiliser seulement 80% de ses inputs actuels et réussir à maintenir sa production au niveau  $y$ . Par conséquent, dans une orientation input, tous les scores d'efficacité sont compris entre 0 et 1.

**Figure 4. Orientation input vs orientation output**



Dans une orientation output, une unité de décision peut améliorer son efficacité par une augmentation proportionnelle de ses outputs, tout en maintenant ses inputs au même niveau.

En reprenant l'exemple précédent (voir **figure 4**), cela signifie que le degré d'inefficience est

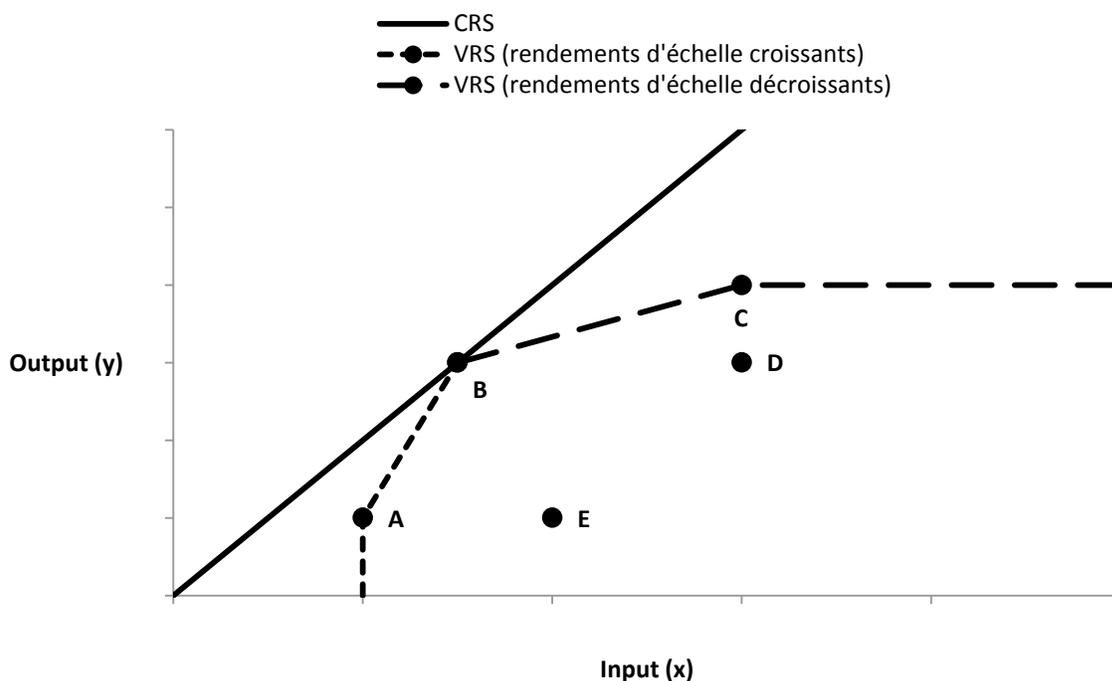
évalué par rapport à l'axe des outputs, soit l'axe des ordonnées. Ainsi, le degré d'inefficience de l'unité de décision utilisant  $x$  unités d'input et  $y$  unités d'output est donné par la distance qui sépare les coordonnées  $(x, y)$  des coordonnées  $(x, y^*)$  situées sur la frontière de production. Si  $y^*$  est 1,2 fois plus grand que  $y$ , alors il est possible d'affirmer que l'unité de décision pourrait produire 1,2 fois plus, tout en utilisant la même quantité d'inputs. Par conséquent, dans une orientation output, tous les scores d'efficience sont supérieurs ou égaux à 1. Dans les deux orientations, une unité de décision est efficiente si son score d'efficience est égal à 1. Plus le score s'éloigne de 1, plus le degré d'inefficience de l'unité de décision est élevé. Coelli *et al.* (2005) suggèrent de choisir l'orientation selon ce que les gestionnaires d'une entreprise peuvent modifier le plus facilement. Au hockey, l'entraîneur peut modifier le temps de jeu d'un joueur et pratiquer à l'entraînement certains aspects de son jeu, améliorant ainsi ses habiletés. Le directeur général peut modifier le salaire d'un joueur lors de la négociation d'un nouveau contrat selon sa perception de son talent potentiel et de ses habiletés. Ainsi, l'orientation choisie est input, car les gestionnaires ont davantage de contrôle sur les inputs.

Par ailleurs, les rendements d'échelle peuvent être constants (modèle *CRS, constant returns to scale*) ou variables (modèle *VRS, variable returns to scale*). Ici, des rendements d'échelle variables sont supposés, car il n'y pas de raisons *a priori* de restreindre le modèle seulement à des rendements d'échelle constants. Par exemple, en doublant le temps de jeu et le talent d'un joueur qui joue déjà beaucoup, la production augmentera, mais il est impossible de savoir de combien. Cette augmentation peut être plus faible que le double, car la fatigue peut entrer en ligne de compte, ou plus forte que le double, car il peut y avoir un effet d'apprentissage. La **figure 5** présente la différence dans l'estimation de la frontière de production en supposant des

rendements d'échelle constants ou variables dans une situation avec un input et un output.

Avec des rendements d'échelle constants, la frontière est une droite passant par l'origine. Avec des rendements d'échelle variables, la frontière présente une convexité « plus prononcée » de façon à mieux « coller » aux données. Cela a pour impact de diminuer la distance de toutes les unités de décision par rapport à la frontière. Un résultat important est donc que l'efficacité estimée avec des rendements d'échelle variables est toujours plus grande ou égale par rapport à l'efficacité estimée avec des rendements d'échelle constants. Par exemple, dans la **figure 5**, seul B est efficient selon avec une technologie CRS (ligne pleine), alors que A, B et C sont efficaces selon une technologie VRS (ligne avec des traits). Avec le modèle VRS, il est possible de noter également qu'à droite de B, sur la ligne avec de longs traits, les rendements d'échelle sont décroissants, alors qu'à gauche de B, sur la ligne avec des traits courts, ils sont croissants.

**Figure 5. Rendements d'échelle variables (VRS) vs rendements d'échelle constants (CRS)**



L'hypothèse *VRS* introduit une contrainte supplémentaire au modèle d'enveloppement présenté plus haut :

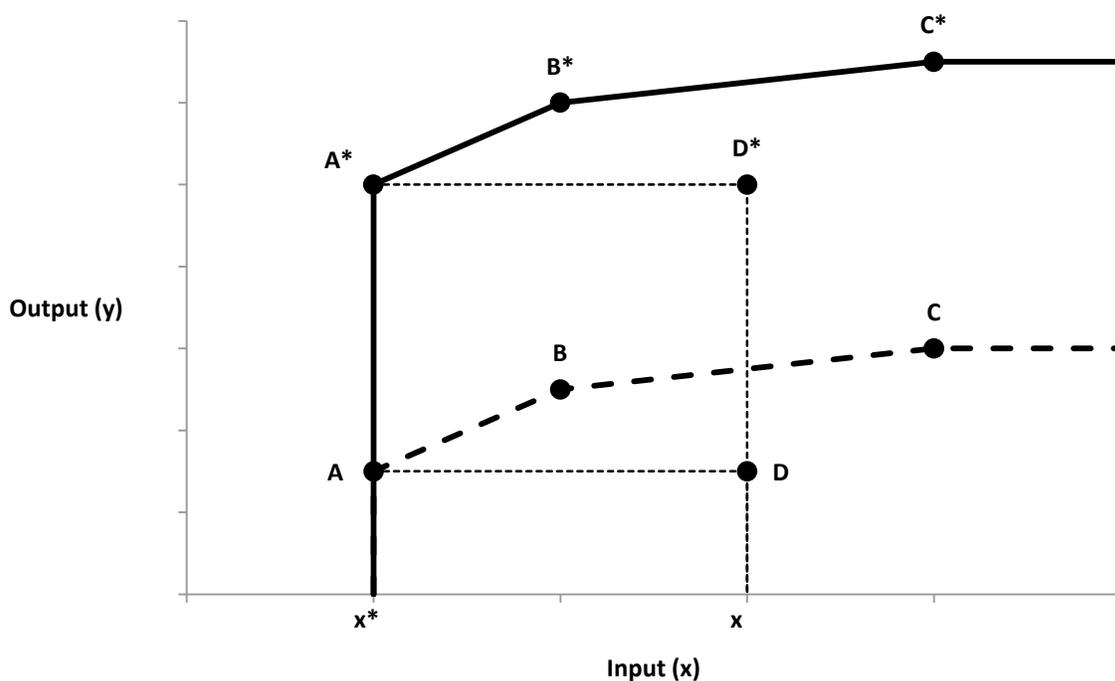
$$\sum_{j=1}^n \lambda_j = 1$$

Ce modèle a été introduit par Banker, Charnes et Cooper (1984). Jumelé à cette dernière contrainte, les équations du modèle d'enveloppement présenté plus haut forment le modèle utilisé dans la présente analyse, soit un modèle *DEA* avec une orientation input et des rendements d'échelle variables.

Un problème avec l'utilisation des mesures *PCO*, *PCD* et *PCG* est qu'un joueur peut être pire que le joueur marginal dans certaines dimensions. Ainsi, ces mesures peuvent prendre des valeurs négatives, car la contribution du joueur marginal est fixée à zéro. La méthode *DEA* ne pourra pas trouver de solution pour ces joueurs, car une des hypothèses du modèle est que les inputs et les outputs doivent être non négatifs. Cependant, il est possible d'ajouter une quantité positive suffisamment grande à tous les outputs de façon à ce que la plus petite valeur soit positive (c'est une translation positive des outputs). Cette opération ne modifie pas le score d'efficacité dans la présente analyse, car selon la proposition de Zhu et Cook (2007), en supposant des rendements d'échelle variables et une orientation input, le score d'efficacité obtenu pour une unité de décision est invariant à une translation positive des outputs. La **figure 6** illustre l'idée derrière cette proposition dans un monde avec un input et un output. Au départ, la frontière de production est donnée par la ligne pointillée et A, B et C sont pleinement efficaces. D est inefficace et son degré d'inefficacité est donné par le ratio  $\frac{x^*}{x}$  si l'orientation

est input. En faisant une translation positive des outputs, la frontière de production se déplace vers le haut et est donnée par la ligne pleine. A, B, et C sont pleinement efficaces et D est inefficent, comme auparavant. Le degré d'inefficience de D est également le même, car il est donné par le ratio  $\frac{x^*}{x}$  et ces deux éléments n'ont pas changé par rapport à la situation initiale.

**Figure 6. Invariance du score d'efficacité à une translation positive des outputs en DEA avec une orientation input et des rendements d'échelle variables**



## Bases de données et échantillon

Plusieurs bases de données ont été utilisées pour réaliser les différentes parties de l'analyse. Une grande partie des données provient du site officiel de la LNH (NHL.com). Depuis la saison 1997-1998, la ligue rapporte des statistiques sur la production des joueurs (buts, passes, points, tirs bloqués, revirements, temps de jeu, mises en échec, plus/moins, pénalités, mises au jeu, nombre de parties jouées...), des données biographiques (position, équipe, date et lieu de

naissance, poids, grandeur, année et rang de repêchage...) et des statistiques sur la performance des équipes (classement, victoires, buts pour, buts contre, pénalités, mises en échec, efficacité en avantage numérique, efficacité en désavantage numérique...). Les données sur les contrats des joueurs (durée, salaire, bonus, type de contrat et historique des équipes pour lesquelles le joueur a joué) sont tirées de Capgeek.com et NHLnumbers.com. Ces deux sites utilisent les rapports médiatiques, les annonces officielles des équipes et leurs propres informateurs pour compiler ces statistiques. Les données contractuelles complètes sont disponibles à partir de la saison 2007-2008. Les statistiques sur la contribution du joueur (*PCO*, *PCD* et *PCG*) sont tirées de hockeyanalytics.com et sont disponibles pour les saisons 2002-2003 à 2010-2011. Alan Ryder, l'auteur des mesures de contribution du joueur, en est le propriétaire et calcule lui-même les statistiques disponibles sur ce site. Finalement, Forbes.com compile des données sur la valeur monétaire et les résultats financiers des équipes de la LNH depuis la saison 2005-2006. Ces données seront utilisées comme variables de contrôle lorsque viendra le temps d'estimer l'association entre le score d'efficacité et la décision de racheter un contrat.

Avec l'information disponible dans ces bases de données, il est possible de calculer un score d'efficacité pour chaque joueur ayant joué durant les saisons 2007-2008 à 2010-2011. 1 449 joueurs ont joué dans au moins une de ces quatre saisons. Certains ont joué plus d'une saison, ce qui porte le total à 3 855 observations. 1 091 joueurs (3 159 observations) ont joué au moins 200 minutes dans une de ces quatre saisons: 784 en 2007-2008, 784 en 2008-2009, 783 en 2009-2010 et 808 en 2010-2011. Le **tableau 2** présente des statistiques descriptives sur les inputs et les outputs de ces joueurs. Comme mentionné précédemment, ce tableau révèle que les mesures de contribution (*PCO*, *PCD* et *PCG*) peuvent prendre des valeurs négatives, car un

joueur peut être pire que le joueur marginal, dont la contribution est fixée à zéro, dans certaines dimensions. Une quantité de 100 est donc ajoutée à chacune de ces mesures de contribution afin de permettre à la méthode *DEA* de trouver un score d'efficacité pour tous les joueurs dans l'échantillon. De plus, comme l'efficacité d'un joueur peut varier d'une saison à l'autre, le score d'efficacité est estimé pour chaque saison. Ainsi, un joueur qui a joué dans les quatre saisons incluses dans l'échantillon obtient quatre scores d'efficacité. La méthode *DEA* est donc appliquée en *window analysis* (Charnes *et al.*, 1984; Yang et Chang, 2009; Halkos et Tzeremes, 2009).

Il est important de noter que les gardiens de but sont inclus dans cet échantillon et sont donc évalués en même temps que les autres joueurs. Par construction des mesures de contribution du joueur (voir **tableau 2**), ajouter les gardiens de but n'affecte pas le score d'efficacité des autres joueurs, car  $PCG = 0$  pour ces derniers et  $PCG$  est la principale composante par laquelle les gardiens peuvent contribuer à la victoire. De plus, comme les gardiens peuvent avoir une contribution selon les trois mesures ( $PCO$ ,  $PCD$  et  $PCG$ ), le fait de les inclure avec les autres joueurs les empêche d'avoir un score d'efficacité élevé à cause d'une contribution offensive positive, mais insignifiante par rapport aux attaquants. En effet, il n'est pas souhaitable d'affirmer qu'un gardien de but est pleinement efficace car il a réussi à contribuer 1 victoire à son équipe par la qualité de son jeu offensif. En comparant les gardiens aux autres joueurs, en particulier les attaquants qui ont une mesure  $PCO$  beaucoup plus élevée, cela rend les contributions offensives et défensives des gardiens beaucoup plus insignifiantes relativement aux autres joueurs. Cela reflète mieux les priorités d'un gestionnaire lors du choix d'un gardien de but, c'est-à-dire que le seul véritable élément important pour un gardien, c'est sa

contribution à la victoire via les tirs au but arrêtés (*PCG*). Pour faire ressortir l'importance de cette caractéristique et réduire le poids accordé aux mesures *PCO* et *PCD* pour les gardiens, il faut inclure les gardiens avec les autres joueurs lors de l'évaluation du score d'efficacité.

Deux logiciels sont utilisés pour effectuer les analyses. Pour faire le traitement de base des données et les régressions (nombre de points au classement et probabilité d'un rachat de contrat), le logiciel SAS version 9.3 est utilisé. Pour calculer le score d'efficacité des joueurs de hockey et corriger le biais de surestimation grâce au bootstrap, le logiciel R version 3.1.1 est utilisé avec le module supplémentaire *rDEA* de Simm et Besstremyannaya (2015).

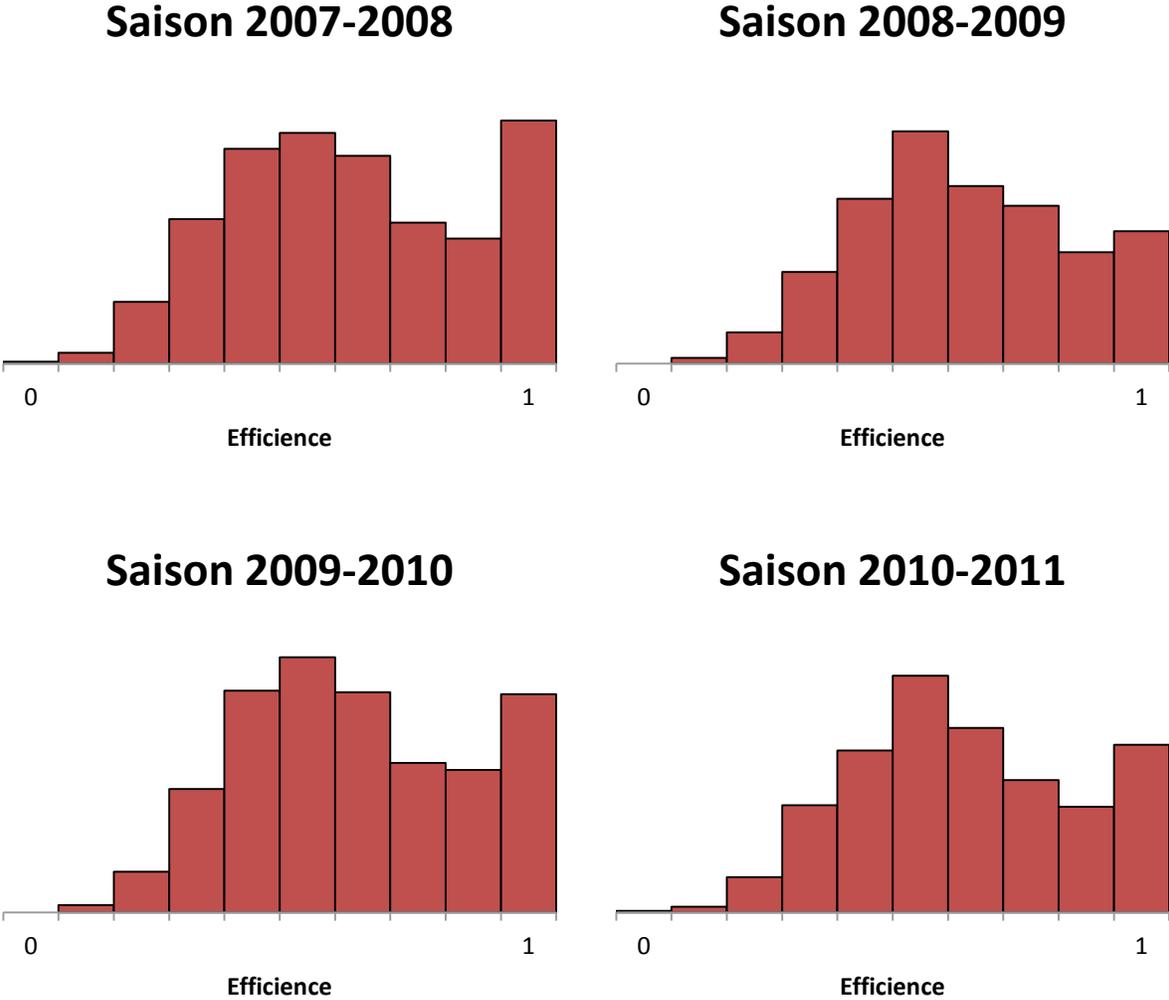
**Tableau 2. Statistiques descriptives sur les inputs et les outputs**

Input/output	Moyenne	Écart-type	Minimum	Médiane	Maximum
<b>Tous les joueurs</b>			<b>n = 3 159</b>		
Minutes de jeu	1 088	668	200	1 026	4 930
Salaire (M\$)	2,1	1,8	0,5	1,4	10,0
PCO	14,4	18,8	-14,0	8,0	138,0
PCD	13,6	14,1	-24,0	11,0	79,0
PCG	6,5	30,7	-42,0	0,0	316,0
<b>Gardiens de buts</b>			<b>n = 300</b>		
Minutes de jeu	1 969	1 265	208	1 717	4 930
Salaire (M\$)	2,2	2,0	0,5	1,2	10,0
PCO	1,6	1,8	-1,0	1,0	9,0
PCD	-1,1	1,3	-7,0	-1,0	0,0
PCG	68,5	75,3	-42,0	46,0	316,0
<b>Autres joueurs</b>			<b>n = 2 859</b>		
Minutes de jeu	996	486	200	1 001	2 252
Salaire (M\$)	2,1	1,8	0,5	1,4	10,0
PCO	15,8	19,3	-14,0	10,0	138,0
PCD	15,1	14,0	-24,0	12,0	79,0
PCG	0,0	0,0	0,0	0,0	0,0

## Résultats

La **figure 7** montre la distribution du score d'efficacité après correction du biais de surestimation pour les quatre saisons. Puisque la méthode *DEA* a été appliquée en *window analysis*, un joueur ayant joué durant ces quatre saisons apparaît dans les quatre graphiques. Le **tableau 3** présente la même information, mais sous forme de statistiques descriptives et inclut aussi l'information sur l'efficacité avant bootstrap et sur la taille du biais de surestimation. Puisque l'orientation est input, le score d'efficacité varie de 0 à 1. Le premier constat est que la distribution du score d'efficacité est approximativement la même d'une saison à l'autre. Cela est confirmé par les chiffres du **tableau 3**, notamment l'efficacité moyenne qui, à chaque saison, tourne autour de 0,635 après bootstrap. Il est possible de voir également que l'échantillon inclut des joueurs pleinement efficaces (score de 1), mais aussi des joueurs fortement inefficaces, comme en témoigne le minimum qui oscille autour de 0,1 pour chaque saison. De plus, hormis quelques exceptions (1,3% des joueurs ont une correction de plus de 0,1), la correction pour le biais est très faible, la moyenne oscillant autour de 0,015. Cela peut s'expliquer par la taille importante de l'échantillon de joueurs de hockey utilisée. En effet, en incluant 75,3% ( $\frac{1091}{1449}$ ) des joueurs disponibles, la frontière de production estimée à chaque saison ne peut pas être très éloignée de la véritable frontière.

Figure 7. Distribution du score d'efficience après bootstrap



**Tableau 3. Statistiques descriptives sur l'efficience**

	Moyenne	Écart-type	Minimum	Médiane	Maximum
<b>Efficience avant bootstrap</b>	<b>0,6498</b>	<b>0,2092</b>	<b>0,0872</b>	<b>0,6263</b>	<b>1,0000</b>
2007-2008	0,6482	0,2207	0,0872	0,6264	1,0000
2008-2009	0,6525	0,1990	0,1470	0,6323	1,0000
2009-2010	0,6534	0,2089	0,1389	0,6329	1,0000
2010-2011	0,6455	0,2079	0,0981	0,6226	1,0000
<b>Efficience après bootstrap</b>	<b>0,6350</b>	<b>0,2062</b>	<b>0,0872</b>	<b>0,6140</b>	<b>1,0000</b>
2007-2008	0,6330	0,2188	0,0872	0,6133	1,0000
2008-2009	0,6335	0,1941	0,1470	0,6092	1,0000
2009-2010	0,6390	0,2057	0,1326	0,6238	1,0000
2010-2011	0,6345	0,2056	0,0981	0,6128	1,0000
<b>Biais de surestimation</b>	<b>0,0149</b>	<b>0,0267</b>	<b>0,0000</b>	<b>0,0061</b>	<b>0,4146</b>
2007-2008	0,0152	0,0262	0,0000	0,0092	0,3018
2008-2009	0,0190	0,0294	0,0000	0,0104	0,3074
2009-2010	0,0144	0,0283	0,0000	0,0054	0,4146
2010-2011	0,0109	0,0217	0,0000	0,0000	0,3250

Il faut noter que 48,3% des joueurs n'ont pas de correction du biais, car ceux-ci ne respectent pas le critère de variance de Simar et Wilson (2000), qui stipule que pour appliquer la correction,  $Var(\theta_j^b) < \frac{Biais_j^2}{3}$ . À titre d'exemple, la **figure 8** illustre la distribution du score d'efficience pour les joueurs de la saison 2007-2008. Dans les deux parties de la figure, les joueurs sont classés en ordre croissant selon leur score d'efficience avant bootstrap. La première partie montre l'estimé du score d'efficience avant bootstrap et l'intervalle de confiance de 95% estimé par la procédure du bootstrap. Il est possible de voir que certains joueurs ont un intervalle de confiance très large, surtout pour ceux avec un score d'efficience

élevé. La deuxième partie de la figure montre l'estimé du score d'efficacité après bootstrap avec l'intervalle de confiance de 95% pour les joueurs dont la correction du biais peut être appliquée selon le critère de variance. Pour les autres joueurs, la deuxième partie de la figure rapporte l'estimé du score d'efficacité avant bootstrap sans intervalle de confiance. C'est pourquoi la plupart des grands intervalles de confiance observés dans la première partie de la figure ont disparu (car ces joueurs ne respectent pas le critère de variance). Des distributions similaires sont observées pour les autres saisons.

D'autres résultats intéressants ressortent de l'analyse du score d'efficacité. En particulier, au **tableau 4**, l'analyse après bootstrap révèle des scores d'efficacité très similaires entre les joueurs jouant à différentes positions (gardiens, centres, ailiers et défenseurs). Cela signifie que le choix des inputs et des outputs est cohérent et ne favorise pas un groupe de joueurs en particulier. Ainsi, tous les joueurs ont la chance de bien paraître.

Figure 8. Distribution du score d'efficacité avant et après bootstrap avec intervalle de confiance de 95% pour la saison 2007-2008

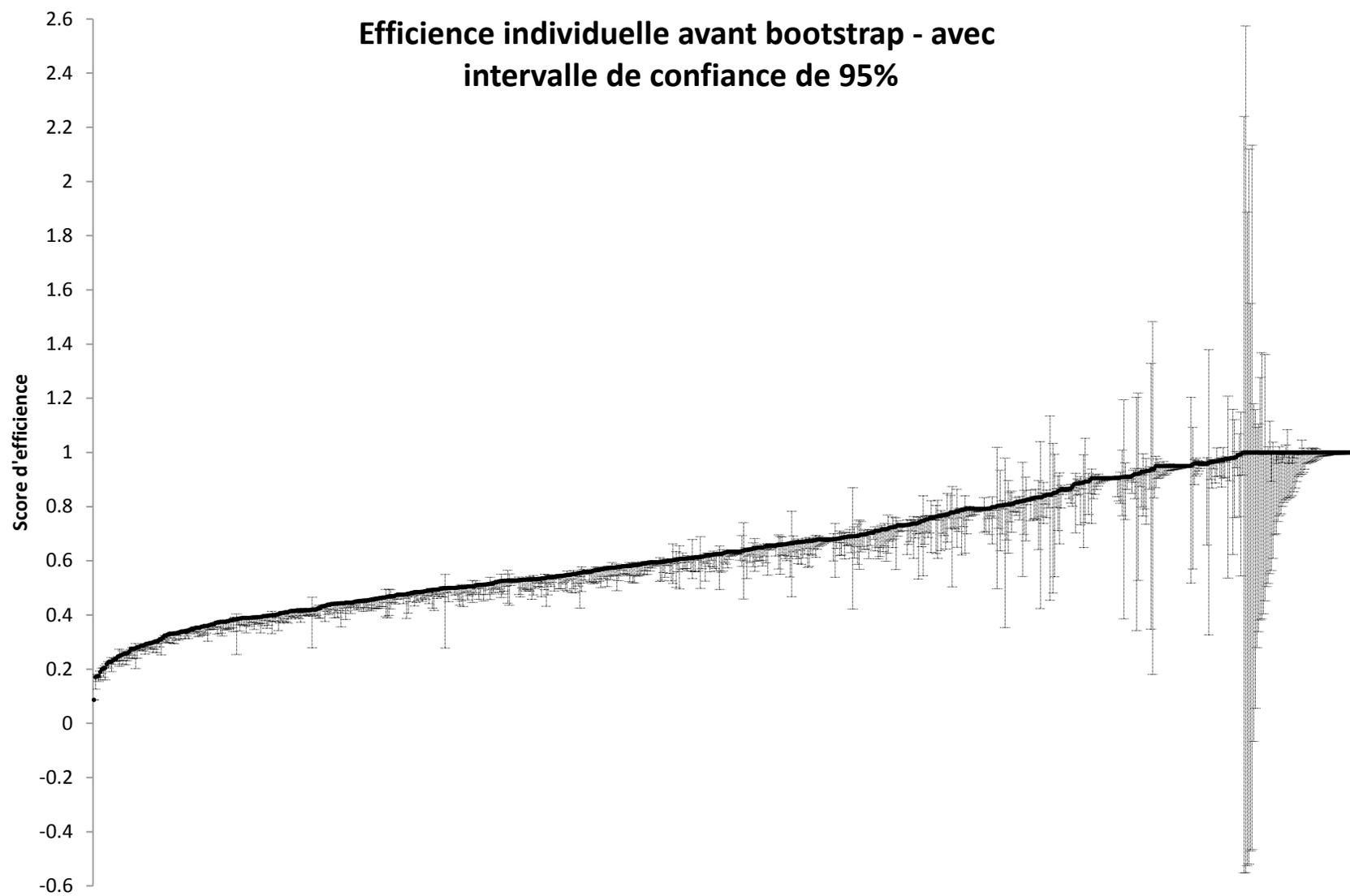
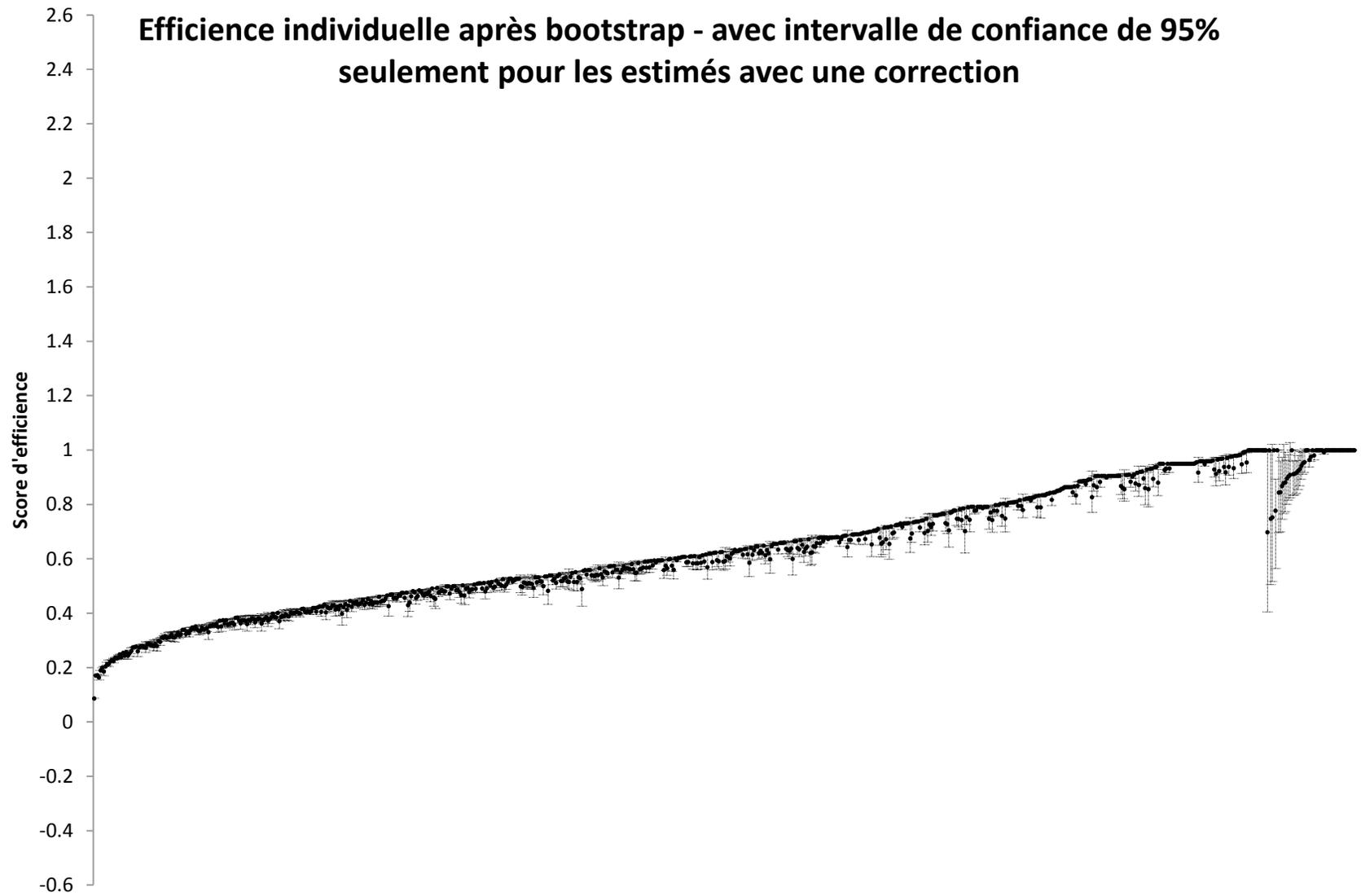


Figure 8 (suite)

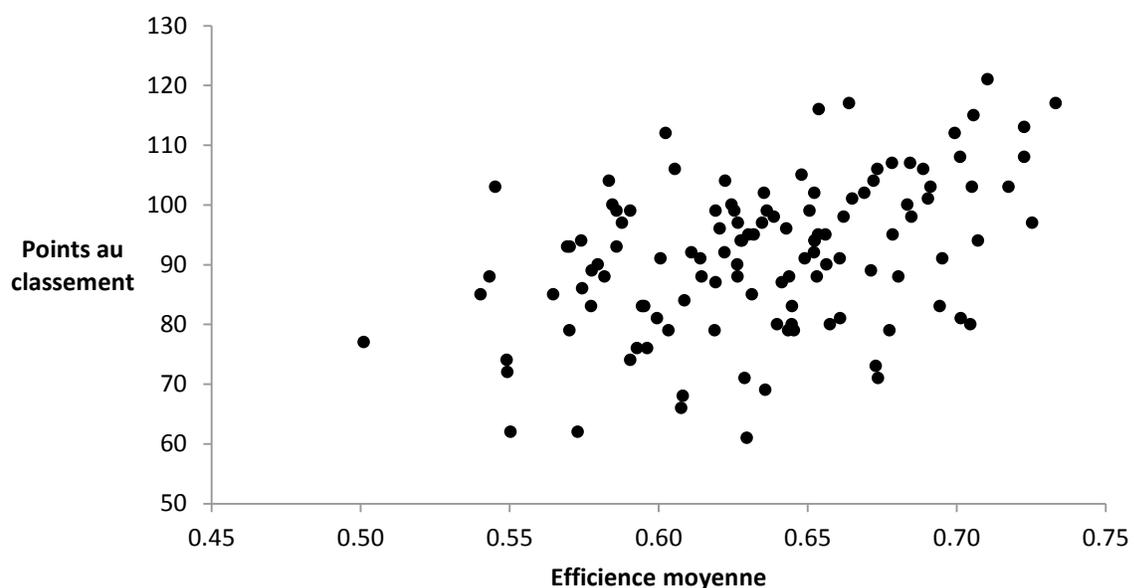


**Tableau 4. Statistiques descriptives sur le score d'efficience après bootstrap par position**

Position	Nombre de joueurs	Moyenne	Écart-type	Minimum	Médiane	Maximum
<b>Toutes les saisons</b>						
Tous les joueurs	3 159	0,63	0,21	0,09	0,61	1,00
Centre	735	0,62	0,21	0,19	0,60	1,00
Ailier gauche	552	0,63	0,20	0,17	0,60	1,00
Ailier droit	546	0,65	0,20	0,19	0,64	1,00
Défenseur	1 026	0,64	0,20	0,16	0,61	1,00
Gardien de but	300	0,64	0,25	0,09	0,64	1,00
<b>Saison 2007-2008</b>						
Tous les joueurs	784	0,63	0,22	0,09	0,61	1,00
Centre	175	0,61	0,22	0,19	0,59	1,00
Ailier gauche	148	0,64	0,21	0,21	0,62	1,00
Ailier droit	133	0,64	0,22	0,19	0,61	1,00
Défenseur	255	0,64	0,21	0,16	0,62	1,00
Gardien de but	73	0,63	0,25	0,09	0,61	1,00
<b>Saison 2008-2009</b>						
Tous les joueurs	784	0,63	0,19	0,15	0,61	1,00
Centre	172	0,61	0,20	0,25	0,59	1,00
Ailier gauche	137	0,63	0,19	0,21	0,60	1,00
Ailier droit	144	0,66	0,19	0,27	0,65	1,00
Défenseur	252	0,63	0,18	0,18	0,60	1,00
Gardien de but	79	0,65	0,23	0,15	0,66	1,00
<b>Saison 2009-2010</b>						
Tous les joueurs	783	0,64	0,21	0,13	0,62	1,00
Centre	179	0,63	0,20	0,21	0,60	1,00
Ailier gauche	142	0,62	0,21	0,17	0,59	1,00
Ailier droit	136	0,64	0,20	0,28	0,63	1,00
Défenseur	254	0,65	0,20	0,21	0,64	1,00
Gardien de but	72	0,65	0,23	0,13	0,68	1,00
<b>Saison 2010-2011</b>						
Tous les joueurs	808	0,63	0,21	0,10	0,61	1,00
Centre	209	0,65	0,20	0,21	0,62	1,00
Ailier gauche	125	0,62	0,19	0,25	0,59	1,00
Ailier droit	133	0,67	0,20	0,24	0,66	1,00
Défenseur	265	0,62	0,19	0,24	0,60	1,00
Gardien de but	76	0,62	0,28	0,10	0,60	1,00

La **figure 9** montre un nuage de points mettant en relation le nombre de points au classement pour une équipe et l'efficacité moyenne des joueurs dans cette équipe. Le coefficient de corrélation trouvé ( $\rho=0,43$ , valeur- $p<0,0001$ ) suggère qu'il existe un lien relativement important et positif, compte tenu d'un contexte de données microéconomiques, entre la combinaison de joueurs efficaces et la performance d'une équipe au classement général. En ajoutant comme variable indépendante l'efficacité moyenne des joueurs d'une équipe dans le modèle MCO des points au classement du **tableau 1**, la corrélation positive et significative de l'efficacité moyenne (coefficient= $0,3761$ , valeur- $p<0,001$ ) est détectée (voir **tableau 5**). Dans ce nouveau modèle, et comme au **tableau 1**, un différentiel de buts positif et élevé (coefficient= $0,3379$ , valeur- $p<0,001$ ) ainsi qu'un pourcentage d'efficacité plus élevé pour les mises en jeu (coefficient= $0,5153$ , valeur- $p=0,015$ ) sont associés à un nombre plus élevé de points au classement pour une équipe.

**Figure 9. Corrélation entre le nombre de points au classement et l'efficacité moyenne des joueurs d'une équipe ( $\rho = 0,43$ )**



**Tableau 5. Association entre le nombre de points au classement par saison et l'efficacité moyenne des joueurs de l'équipe en contrôlant pour les statistiques de performance de l'équipe (saisons 2007-2008 à 2010-2011)**

**Nombre d'équipes: n=120**

Variable	Coefficient	Erreur standard	Valeur-p
Constante	51,2490	18,2681	0,006*
Différentiel de buts pour/contre	0,3379	0,0160	<0,001*
% d'efficacité avantage numérique	-0,3214	0,1805	0,078
% d'efficacité désavantage numérique	-0,0873	0,1565	0,578
% d'efficacité mises en jeu	0,5153	0,2083	0,015*
Mises en échec	0,0008	0,0015	0,588
Tirs bloqués	0,0007	0,0032	0,825
Pertes de possession	0,0046	0,0026	0,078
Reprises de possession	-0,0028	0,0037	0,453
Efficacité moyenne des joueurs	0,3761	0,0917	<0,001*

\* indique une valeur-p < 0,05

Note: Les corrélations entre les variables utilisées et le détail des résultats en entrant les variables une à une dans la régression sont disponibles aux annexes 2 et 3.

En termes d'allocation des ressources, les résultats aux **figures 7 et 8** ainsi qu'au **tableau 3** montrent qu'il y a encore place à amélioration, car une quantité importante de joueurs se retrouve avec un score d'efficacité faible. En effet, 28% des observations ont un score d'efficacité après bootstrap inférieur à 0,5. Cela peut s'expliquer en partie par une certaine rigidité dans la composition des équipes, c'est-à-dire que chaque équipe doit avoir au minimum 12 attaquants, 6 défenseurs et 2 gardiens de but à sa disposition pour bien fonctionner. Pour remplir ces positions, un gestionnaire peut être prêt à donner des contrats à des joueurs un peu moins efficaces. Cela s'ajoute au fait que la profondeur à certaines positions est moindre (par exemple, les gardiens de but), ce qui fait qu'au final, le gestionnaire peut se retrouver avec un choix assez limité pour combler les postes vacants. De plus, la liquidité du marché des joueurs

est encore plus faible si on considère seulement les joueurs vedettes. Cela fait en sorte que certaines équipes sont prêtes à payer une prime pour essayer d'attirer une vedette dans un marché de hockey peu attirant (à cause du manque de partisans, des impôts plus élevés<sup>8</sup>...), réduisant ainsi le score d'efficience pour ce joueur. Tous ces éléments, en plus de l'impact sur les ventes de produits dérivés, peuvent expliquer pourquoi au final, un gestionnaire peut se résigner à garder des joueurs inefficients dans son équipe. L'intérêt du score d'efficience développé ici pourrait donc être de diminuer l'asymétrie d'information sur le marché des joueurs de hockey (excluant les nouveaux joueurs, car il n'existe pas de données historiques sur leur performance dans la LNH), ce qui améliorerait sa fluidité et amènerait les gestionnaires à mieux cerner les problèmes dans l'équipe. Il est important de noter aussi que le score d'efficience calculé ici ne tient pas compte des séries éliminatoires. Cela peut expliquer dans certains cas pourquoi les gestionnaires conservent année après année des joueurs relativement inefficients en saison régulière. En effet, ces joueurs peuvent changer de comportement et devenir beaucoup plus efficaces en séries éliminatoires, ce qui leur vaut l'étiquette de « joueurs de séries ». Cependant, les séries éliminatoires étant un tournoi à la ronde où seulement 16 des 30 équipes participent et jouent jusqu'à sept parties de suite contre la même équipe, la dynamique est complètement différente. Le fait de ne pas en tenir compte correspond néanmoins à une limite de l'analyse effectuée ici.

---

<sup>8</sup> Les salaires dans la présente étude sont évalués en USD et avant impôts.

## Rachats de contrat

Après la fin des séries éliminatoires s'ouvre une période d'environ deux semaines durant laquelle les gestionnaires de chaque équipe peuvent racheter le contrat des joueurs dans leur équipe. Racheter un contrat termine prématurément l'entente entre le joueur et l'équipe. Le joueur devient alors un agent libre et peut signer une entente avec une autre équipe. L'équipe doit alors payer au joueur  $1/3$  ou  $2/3$  (selon l'âge du joueur) du salaire restant pour les saisons du contrat qui sont rachetées, mais sur une période deux fois plus longue (National Hockey League, 2005). Cela permet de libérer de l'espace en-dessous du plafond salarial. Par exemple, une équipe rachète le contrat d'un joueur dont il reste 3M de dollars à payer par année pour deux ans (donc il reste un total de 6M au contrat). Si le joueur est âgé de plus de 26 ans, alors il recevra de l'équipe qui a racheté son contrat 1M par année pour 4 ans, ce qui correspond à  $2/3$  de 6M (donc 4M) réparti sur le double du temps restant au contrat<sup>9</sup>.

Il est connu qu'en général, les gestionnaires rachètent les contrats de joueurs dont la performance est décevante par rapport aux attentes. Ce sont donc des joueurs inefficients. Si la mesure d'efficacité construite est bonne, alors la probabilité de rachat de contrat devrait être plus élevée pour un joueur inefficace. La validité de la mesure d'efficacité sera donc partiellement confirmée si le lien entre l'efficacité et la probabilité de rachat est négatif et statistiquement significatif.

Le rachat de contrat est un événement du type 0 (pas de rachat) ou 1 (rachat), ce qui signifie qu'une régression logistique peut être utilisée. Cependant, dans le cas d'événements rares (il y

---

<sup>9</sup> Ce montant est versé même si le joueur signe une entente avec une autre équipe.

a une dizaine de rachats de contrat par saison), les coefficients estimés sont biaisés (King et Zeng, 2001). Il y a trop peu de 1 pour avoir une estimation juste de  $P(Y = 1|X)$ . Il existe deux solutions à ce problème : corriger l'estimateur directement (King et Zeng, 2001) ou corriger la fonction de vraisemblance du modèle logit avec une fonction de vraisemblance pénalisée (Firth, 1993; Heinze et Schemper, 2002). Celles-ci donnent des résultats très similaires (King et Zeng, 2001). Ici, la méthode de Firth (1993) sera utilisée.

Outre l'efficacité individuelle, d'autres variables peuvent avoir un lien avec la probabilité d'un rachat de contrat. Par exemple, un nombre élevé de parties non jouées par un joueur au cours de la saison (en raison de blessures ou autres), des revenus élevés pour une équipe (ce qui implique plus d'argent disponible pour un rachat) et un faible nombre de points au classement pour une équipe pourraient augmenter la probabilité d'un rachat de contrat. Un nombre élevé d'années restant à un contrat pourrait à l'inverse diminuer cette probabilité, car bien que le rachat de contrat libère de l'espace sur la masse salariale, l'opération demeure coûteuse pour une équipe. L'exploration préliminaire des données disponibles en faisant des comparaisons simples entre les deux groupes (rachat contre pas de rachat) donne une première idée des variables de contrôle potentielles (voir **tableau 6**). Il est important de noter que l'échantillon ne contient que les joueurs pour lesquels il reste au moins une année au contrat au moment d'entrer dans la période de rachat de contrat. Ainsi, il reste 1562 observations: 35 rachats de contrat et 1527 joueurs dont le contrat n'a pas été racheté. Les comparaisons ont été effectuées grâce au test du khi-carré de Pearson (Pearson, 1900) pour les variables dichotomiques et au test de Wilcoxon-Mann-Whitney (Wilcoxon, 1945; Mann et Whitney, 1947) pour les variables continues. Le **tableau 6** montre que la moyenne d'âge est plus élevée

pour le groupe de joueurs dont le contrat a été racheté (31,6 vs 28,0, valeur- $p < 0,0001$ ). De plus, la proportion de Nord-Américains (88,6% vs 71,4%, valeur- $p = 0,0260$ ) et la proportion de joueurs avec un contrat dit « standard » (94,3% vs 78,0%, valeur- $p = 0,0207$ ) sont tous les deux plus élevées dans le groupe des rachats<sup>10</sup>. Le nombre moyen de parties non jouées est aussi plus élevé pour les joueurs dont le contrat a été racheté, comparé aux joueurs dont le contrat n'a pas été racheté (32,4 vs 18,1, valeur- $p < 0,0001$ ). De plus, le nombre moyen d'années restant au contrat (1,3 vs 2,0, valeur- $p = 0,0052$ ) et le nombre moyen de points au classement de l'équipe sont plus faibles (85,9 vs 91,5, valeur- $p = 0,0119$ ) pour les joueurs avec un rachat de contrat. Finalement, l'efficacité moyenne après bootstrap est significativement plus faible pour les joueurs dont le contrat a été racheté (0,47 vs 0,60, valeur- $p < 0,0001$ ). Cela semble confirmer l'intuition initiale d'une probabilité de rachat plus élevée pour les joueurs fortement inefficients.

---

<sup>10</sup> Le contrat dit « standard » est le plus commun dans la LNH. Cependant, les jeunes joueurs peuvent avoir un contrat d'entrée (*Entry Level*) et les joueurs de 35 ans et plus peuvent avoir un contrat « 35 ans et plus ». Ces deux types de contrat diffèrent sur quelques points par rapport au contrat « standard ».

**Tableau 6. Exploration des variables ayant un lien potentiel avec le rachat de contrat**

Caractéristiques	Joueurs rachetés		Autres joueurs		Valeur-p
	N=35		N=1 527		
<b>Age, moyenne ± écart-type [médiane]</b>	31,6 ± 3,9 [31,4]		28,0 ± 4,6 [27,8]		<0,0001*
<b>Continent d'origine, n (%)</b>					
Amérique du Nord	31	(88,6)	1 091	(71,4)	0,0260*
Europe	4	(11,4)	425	(27,8)	0,0316*
Autre	0	(0,0)	11	(0,0)	0,6143
<b>Position, n (%)</b>					
Centre	4	(11,4)	375	(24,6)	0,0732
Ailier gauche	7	(20,0)	254	(16,6)	0,5976
Ailier droit	9	(25,7)	252	(16,5)	0,1487
Défenseur	11	(31,4)	510	(33,4)	0,8069
Gardien de but	4	(11,4)	136	(8,9)	0,6055
<b>Côté du tir, n (%)</b>					
Gauche	24	(68,6)	1 022	(66,9)	0,8381
Droite	11	(31,4)	503	(32,9)	0,8507
<b>Parties non jouées, moyenne ± écart-type [médiane]</b>	32,4 ± 22,5 [30,0]		18,1 ± 19,1 [11,0]		<0,0001*
<b>Efficiéce après bootstrap, moyenne ± écart-type [médiane]</b>	0,47 ± 0,23 [0,39]		0,60 ± 0,19 [0,57]		<0,0001*
<b>Taille et poids, moyenne ± écart-type [médiane]</b>					
Taille (pouces)	72,9 ± 2,1 [73,0]		73,2 ± 2,1 [73,0]		0,5074
Poids (livres)	203,9 ± 16,5 [200,0]		204,9 ± 15,5 [205,0]		0,6339
<b>Repêchage, moyenne ± écart-type [médiane]</b>					
Ronde de repêchage	2,7 ± 2,1 [2,0]		3,0 ± 2,5 [2,0]		0,7660
Position de repêchage	63,6 ± 59,6 [41,0]		74,8 ± 76,1 [44,0]		0,9933

\* indique une valeur-p < 0,05

Tableau 6 (suite)

Caractéristiques	Joueurs rachetés		Autres joueurs		Valeur-p
	N=35		N=1 527		
<b>Information contractuelle</b>					
Valeur restante au contrat (M\$), moyenne ± écart-type [médiane]	3,5	± 2,3 [3,1]	6,9	± 10,5 [3,0]	0,7034
Années restantes au contrat, moyenne ± écart-type [médiane]	1,3	± 0,6 [1,0]	2,0	± 1,6 [1,0]	0,0052*
Valeur totale du contrat (M\$), moyenne ± écart-type [médiane]	9,6	± 7,0 [8,6]	11,9	± 13,9 [6,8]	0,5795
Années totales du contrat, moyenne ± écart-type [médiane]	3,3	± 1,2 [3,0]	3,6	± 1,7 [3,0]	0,6940
Contrat d'entrée (entry-level), n (%)	0	(0,0)	283	(18,5)	-
Contrat standard, n (%)	33	(94,3)	1 191	(78,0)	0,0207*
Contrat pour les 35 ans et plus, n (%)	2	(5,7)	53	(3,5)	0,4765
<b>Statistiques de l'équipe, moyenne ± écart-type [médiane]</b>					
Points au classement	85,9	± 11,4 [86,0]	91,5	± 12,5 [92,0]	0,0119*
Nombre de victoires	38,3	± 6,2 [38,0]	40,9	± 6,6 [41,0]	0,0236*
<b>Statistiques financières de l'équipe, moyenne ± écart-type [médiane]</b>					
Valeur totale (M\$)	238,6	± 111,5 [192,5]	241,2	± 90,4 [213,1]	0,2506
Revenus annuels (M\$)	102,6	± 32,9 [91,8]	103,5	± 27,8 [97,0]	0,3673
Revenus d'opération (M\$)	5,0	± 22,2 [-3,0]	4,4	± 16,3 [-0,2]	0,1632
Ratio dette/valeur (%)	52,6	± 27,7 [53,0]	45,2	± 29,3 [46,0]	0,0841
<b>Saison, n (%)</b>					
2007-2008	13	(37,1)	372	(24,4)	0,0828
2008-2009	5	(14,3)	403	(26,4)	0,1070
2009-2010	9	(25,7)	368	(24,1)	0,8253
2010-2011	8	(22,9)	384	(25,1)	0,7573

\* indique une valeur-p < 0,05

L'exploration préliminaire des données suggère donc que l'âge, le continent d'origine, le fait d'avoir un contrat dit « standard », l'efficacité après bootstrap, le nombre de parties non jouées, le nombre d'années restant au contrat et le nombre de points au classement peuvent avoir un lien avec la probabilité d'un rachat de contrat. Ces variables seront donc ajoutées au modèle de prédiction de la probabilité d'un rachat de contrat. Cependant, d'un point de vue logique, deux autres variables de contrôle doivent être ajoutées, bien que les comparaisons simples entre les deux groupes n'aient pas détecté de différence significative. Premièrement, il se peut que les joueurs jouant à certaines positions soient plus susceptibles de voir leur contrat racheté. La plus grande différence risque d'être entre les gardiens de but et les autres joueurs, car c'est entre ces deux groupes que les différences de contribution à l'équipe sont les plus grandes (voir **tableau 2**). Ainsi, une variable dichotomique prenant la valeur 1 pour un gardien de but et 0 sinon est par conséquent ajoutée au modèle. Deuxièmement, puisque le rachat de contrat demeure une opération coûteuse pour une équipe, les équipes les plus riches sont probablement les plus susceptibles de s'engager dans cette voie. Pour mesurer cet effet, les revenus bruts annuels de l'équipe, tirés de Forbes.com (*The Business of Hockey*) et ajustés pour l'inflation de ceux-ci à travers les saisons, sont ajoutés comme variable au modèle.

Le **tableau 7** rapporte les résultats de la régression logistique en utilisant la méthode de Firth et les variables mentionnées au paragraphe précédent. Comme anticipé, un score d'efficacité élevé est associé à une forte diminution de la probabilité d'un rachat de contrat (rapport des chances<sup>11</sup> = 0,01, valeur-p < 0,001). Cela permet de partiellement confirmer la validité de la

---

<sup>11</sup> Un rapport des chances supérieur à 1 indique que la probabilité d'un rachat est plus élevée pour les joueurs ayant un score d'efficacité élevé, alors qu'un rapport des chances inférieur à 1 indique que la probabilité d'un rachat est plus faible pour les joueurs avec un score d'efficacité élevé.

mesure d'efficacité utilisée. Comme mentionné auparavant, les directeurs généraux sont prêts à garder des joueurs inefficaces dans leur équipe lorsqu'ils sont incapables d'aller chercher mieux sur le marché des joueurs. Cependant, lorsque le degré d'inefficacité est trop fort, il est plus probable qu'ils optent pour le rachat de contrat. Ce résultat montre également que les directeurs généraux ont instinctivement une idée générale de l'efficacité des joueurs à leur disposition. En effet, ils ont peut-être de la difficulté à cerner la différence d'efficacité entre deux joueurs dont les scores sont légèrement différents (par exemple, un joueur avec un score de 0,6 comparé à un joueur avec un score de 0,7), mais ils savent détecter une importante différence, comme en témoigne le fait qu'ils rachètent davantage les contrats de joueurs hautement inefficaces. Ainsi, en ayant une mesure précise de l'efficacité des joueurs, les directeurs généraux pourront encore mieux comparer les joueurs, ce qui mènera à des décisions plus judicieuses dans la gestion de leur équipe et dans des contextes plus complexes que les rachats de contrat.

**Tableau 7. Prédicteurs de la probabilité d'un rachat de contrat<sup>1</sup>****Rachat de contrat: n=35****Pas de rachat de contrat: n=1 527**

Variable	Rapport des chances (odds ratio)	Intervalle de confiance (95%)	Valeur-p
Âge	1,15	(1,05 , 1,26)	0,002*
Originaire d'Amérique du Nord	3,12	(1,17 , 8,34)	0,023*
Gardien de but	0,24	(0,07 , 0,80)	0,020*
Efficienc e après bootstrap	0,01	(0,00 , 0,11)	<0,001*
Le contrat est de type «Standard» <sup>2</sup>	3,46	(0,99 , 12,06)	0,051
Nombre d'années restant au contrat	0,53	(0,33 , 0,86)	0,010*
Revenus de l'équipe (M\$)	1,00	(0,99 , 1,02)	0,461
Parties non jouées	1,05	(1,03 , 1,08)	<0,001*
Points de l'équipe au classement	0,98	(0,95 , 1,01)	0,133

\* indique une valeur-p &lt; 0,05

Notes:

1. Les corrélations entre les variables utilisées et le détail des résultats en entrant les variables une à une dans la régression sont disponibles aux annexes 2 et 3.
2. La variable dichotomique pour le type de contrat prend la valeur 1 si le contrat est «standard» et 0 sinon.

D'autres résultats intéressants ressortent de la régression au **tableau 7**. En particulier, les joueurs plus âgés ont une plus grande probabilité de voir leur contrat racheté (rapport des chances=1,15, valeur-p=0,002). Cela est probablement attribuable au fait qu'avec l'âge, le talent potentiel du joueur diminue, mais il y a une certaine incertitude quant au moment exact où cette diminution se produit et la rapidité avec laquelle le talent diminue. Dans ce cas de figure, les directeurs généraux sont plus proactifs et optent plus rapidement pour le rachat de contrat.

De plus, les joueurs originaires d'Amérique du Nord (Canada et États-Unis) ont une probabilité 3,12 fois plus élevée (valeur-p=0,023) de voir leur contrat racheté par rapport aux autres

joueurs. Une explication possible pour cette forme de « discrimination » est que les gestionnaires des équipes de la LNH peuvent avoir une préférence pour les joueurs locaux (Nord-américains), ce qui peut attirer plus de spectateurs aux matchs. Ce choix se fait cependant au détriment du talent du joueur. La probabilité de rachat peut alors être plus élevée lorsque les gestionnaires d'une équipe réalisent que les performances du favori de la foule sont en-dessous des attentes. Ce phénomène est probablement plus fréquent chez les joueurs de soutien que chez les meilleurs joueurs. En effet, le bassin de joueurs de soutien est beaucoup plus grand, ce qui signifie que les gestionnaires ont beaucoup plus de flexibilité dans leur choix et peuvent se permettre d'avoir une préférence pour des joueurs locaux (nord-américains). À l'opposé, le bassin de joueurs vedettes est tellement petit qu'un gestionnaire ne peut pas se permettre de se limiter aux joueurs locaux dans ses décisions. Ainsi, selon cette logique, le simple fait d'être Américain n'augmente pas réellement la probabilité d'un rachat de contrat. Une variable d'interaction entre le continent d'origine et le statut du joueur (soutien vs vedette) serait plus judicieuse. Cependant, toutes les estimations pour prendre en compte cet effet n'ont pas été concluantes. Une réflexion plus approfondie sur cet aspect serait nécessaire pour tester rigoureusement cette hypothèse.

Par ailleurs, les gardiens de but ont une probabilité beaucoup plus faible de voir leur contrat racheté (rapport des chances=0,24, valeur-p=0,02), par rapport aux autres joueurs. Cela peut être en raison du manque de profondeur à cette position par rapport aux autres positions. En effet, l'échantillon de gardiens est plus faible pour chaque saison (environ 70, soit un peu moins de 10% de l'échantillon de joueurs pour une saison). Obtenir un gardien de but de premier plan est donc très difficile. Les directeurs généraux sont alors plus réticents à se départir d'un

gardien de but via un rachat. Ryder (2003) partage aussi cette opinion. Un autre argument réduisant la probabilité de rachat est que les gardiens de but sont plus faciles à évaluer que les autres joueurs, car leur contribution est presque unidimensionnelle via la mesure *PCG*. Par conséquent, le degré d'erreur dans l'évaluation d'un gardien est plus faible, ce qui engendre moins de rachats de contrat à cette position. Cet argument est appuyé par Lavoie, Grenier et Coulombe (1987).

Les joueurs ayant un contrat dit « standard » ont une probabilité de rachat 3,46 fois plus élevée (valeur- $p=0,051$ ) que les autres joueurs. La valeur- $p$  est toutefois légèrement supérieure à 0,05, étant donnée la forte variance reliée à ce coefficient. Cela provient d'un « problème de séparation », en ce sens où les joueurs dont le contrat est racheté ont presque tous un contrat standard, ce qui signifie que la variable dichotomique « contrat standard » prend presque uniquement la valeur 1 dans le groupe des rachats de contrat. Le modèle a alors de la difficulté à trouver un estimé précis du coefficient associé à cette variable. Bien qu'il soit possible d'avoir un rachat de contrat pour les autres types de contrat, la probabilité est plus faible. Par exemple, pour les jeunes joueurs ayant un contrat d'entrée, cela peut s'expliquer par la patience des dirigeants, l'espoir étant qu'ils développent leur plein potentiel avec le temps.

Un joueur avec un nombre plus élevé d'années restant au contrat a une probabilité de rachat plus faible (rapport des chances=0,53, valeur- $p=0,01$ ). Cela confirme l'intuition initiale reliée au fait que le rachat de contrat est une opération coûteuse où les directeurs généraux préfèrent racheter une ou deux années de contrat plutôt que plusieurs années. La même relation est

trouvée si la variable « nombre d'années restant au contrat » est remplacée par la variable « valeur restante au contrat » (rapport des chances=0,869, valeur-p=0,0083).

Les joueurs avec un nombre plus élevé de parties non jouées au cours d'une saison ont une probabilité plus élevée de voir leur contrat racheté (rapport des chances=1,05, valeur-p<0,001).

Cela est normal, car un directeur général ne sera pas satisfait d'un joueur qui joue une faible proportion du nombre de parties totales sur une saison, indépendamment de son degré d'efficacité. Idéalement, celui-ci veut une production positive à tous les matchs.

Finalement, le nombre de points au classement (rapport des chances=0,98, valeur-p=0,133) et les revenus bruts de l'équipe (rapport des chances=1,00, valeur-p=0,461) ne semblent pas avoir de lien avec la décision de racheter le contrat d'un joueur. Cela suggère que c'est l'évaluation individuelle du joueur qui domine le processus de décision de rachat de contrat et non les performances sportive et financière de l'équipe<sup>12</sup>.

## Limites

L'analyse de l'efficacité des joueurs de hockey effectuée ici comporte évidemment certaines limites. Tout d'abord, elle ne tient pas compte de la qualité des entraîneurs et suppose que toute production offensive et défensive d'un joueur provient uniquement de son talent et de son temps de jeu. Il se peut donc qu'il manque un input important dans la fonction de production estimée par la méthode *DEA*. Dans ce mémoire, la qualité des entraîneurs ne peut pas être ajoutée, car cet élément est difficile à mesurer. Leur salaire pourrait être une option

---

<sup>12</sup> Une analyse de sensibilité avec des variables dichotomiques pour les années révèle des résultats similaires à ce qui a été présenté au **tableau 7**. Les coefficients de ces variables dichotomiques ne sont pas statistiquement différents de 0.

pour l'approximer, mais cette information n'est pas disponible pour tous les entraîneurs et lorsqu'elle l'est, elle est imprécise. De plus, plusieurs articles mettent en doute l'importance de la qualité de l'entraîneur sur la performance d'une équipe ou d'un joueur. Par exemple, Berri *et al.* (2009) montrent que dans la *National Basketball Association (NBA)* la plupart des entraîneurs n'ont pas d'impact significatif sur la performance des joueurs ou de l'équipe. Audas, Goddard et Rowe (2006) montrent même que les équipes de la LNH qui ont changé d'entraîneur en milieu de saison performant moins bien *après* le changement. Cela provient du fait que les entraîneurs dans le sport professionnel ont des aptitudes similaires et prennent donc le même genre de décisions (Berri, 2012). Autrement dit, le degré d'homogénéité des entraîneurs est élevé.

Ensuite, un joueur ne produit pas seulement sur la glace. Par exemple, les qualités de leader d'un joueur peuvent être un output important qui peut permettre à une équipe de mieux performer au même titre que créer ou empêcher des buts. Cependant, cet aspect du jeu est difficile à mesurer. En élargissant la perspective d'étude, un joueur peut aussi générer des revenus de publicité, de vente de marchandises et de billets pour une équipe. Ainsi, une perspective plus large de l'efficacité des joueurs de hockey pourrait tenir compte de la performance sportive *et* financière d'un joueur. Encore une fois, il est difficile de mesurer directement ces éléments, car les équipes de hockey ne sont pas tenues de dévoiler leurs résultats financiers, car elles sont membres de la LNH, qui est un organisme sans but lucratif (Sportsdocuments.com, 2013).

La mesure d'efficacité développée ici ne tient pas compte non plus de l'importance du jeu d'équipe dans la performance individuelle. En effet, une bonne chimie entre les joueurs peut stimuler la production offensive et défensive individuelle, tout comme un bon système de jeu tirant profit des forces et des faiblesses de chacun. La qualité des adversaires et la qualité des coéquipiers peuvent aussi avoir un impact sur la production individuelle. Par exemple, si un joueur joue constamment contre les meilleurs joueurs adverses, sa production offensive et défensive pourrait en souffrir. Si un joueur joue constamment avec les meilleurs de son équipe, sa production pourrait être artificiellement gonflée et ne pas refléter sa véritable capacité à marquer et empêcher des buts. Cependant, Tulsky (2012) montre qu'au cours d'une saison, la qualité des coéquipiers et des adversaires est très similaire à travers les joueurs des différentes équipes, ce qui signifie que le biais d'omission de ces deux caractéristiques est probablement négligeable.

Les séries éliminatoires sont également ignorées dans cette analyse, alors qu'il est connu que certains joueurs performant mieux durant cette période. Ici encore, le biais d'omission est probablement faible, car 16 équipes sur 30 participent à ce tournoi et de ce nombre, 8 sont éliminées au premier tour, ce qui signifie que 50% des équipes jouent au maximum 7 parties durant les séries éliminatoires (une équipe avance au prochain tour en gagnant 4 parties contre son adversaire sur un maximum de 7 parties). Ainsi, l'impact des séries sur le temps de jeu total et la contribution totale (*PCO*, *PCD* et *PCG*) est plutôt faible, comparé à l'impact de la saison régulière, qui est de 82 parties.

Une autre limite se situe au niveau de la disponibilité des données. Pour l'analyse, seules les saisons 2007-2008 à 2010-2011 ont été étudiées. La disponibilité des données sur le salaire des joueurs empêche de remonter davantage dans le temps. En ayant un historique des salaires plus grand, il aurait été intéressant de vérifier si l'introduction du plafond salarial suite au lock-out de 2004-2005 a amené une hausse généralisée de l'efficacité des joueurs de hockey dans la LNH.

Malgré ces limites, un bon bout de chemin a été parcouru avec ce mémoire. En particulier, les analyses effectuées dans ce mémoire ont permis de montrer comment l'efficacité des joueurs de hockey peut être estimée en utilisant les outils économiques et statistiques récemment développés et comment cette efficacité peut être associée au nombre de points au classement d'une équipe et à la décision de racheter un contrat.

## **Conclusion**

L'analyse de l'efficacité des joueurs de hockey dans la LNH pour les saisons 2007-2008 à 2010-2011 par la méthode *DEA* montre que beaucoup de joueurs sont hautement inefficients et pourraient avoir un niveau de production équivalent tout en utilisant une quantité moindre d'inputs. Ceci est probablement le résultat d'un manque de profondeur à certaines positions sur le marché des joueurs. Le score d'efficacité moyen est similaire pour toutes les positions (gardien, défenseur,...), ce qui signifie qu'aucune position n'est favorisée par le choix des inputs et des outputs. L'analyse montre également que le score d'efficacité moyen des joueurs d'une équipe est fortement lié avec les points de cette équipe au classement. De plus, l'efficacité individuelle permet de mieux cerner les individus à risque de voir leur contrat racheté, car il a

été montré qu'une efficacité plus faible est reliée à une probabilité de rachat plus élevée.

L'analyse des rachats de contrat a aussi révélé d'autres facteurs pouvant avoir un lien avec la probabilité de rachat, comme l'âge, la position du joueur, le continent d'origine, le nombre de parties non jouées, le type de contrat et le nombre d'années restant au contrat.

## Bibliographie

### Articles de périodiques

Aigner, D., Lovell, C.A.K. et P. Schmidt (1977), Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, 6, 20-37.

Audas, R., Goddard, J. et W.G. Rowe (2006), Modelling Employment Durations of NHL Head Coaches: Turnover and Post-succession Performance, *Managerial and Decision Economics*, 27, 293-306.

Banker, R.D., Charnes, A., et W.W. Cooper (1984), Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, 30 (9), 1078-1092

Berri, D.J., Leeds, M.A., Leeds, E.M. et M. Mondello (2009), The Role of Managers in Team Performance, *International Journal of Sport Finance*, 4, 75-93.

Charnes, A., Clark, C.T., Cooper, W.W. et B. Golany (1984), A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces, *Annals of Operations Research*, 2 (1), 95-112.

Charnes, A., Cooper, W.W. et E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2, 429-444.

Cooper, W.W., Ruiz, J.L. et I. Sirvent (2009), Selecting non-zero weights to evaluate effectiveness of basketball players with DEA, *European Journal of Operational Research*, 195, 563-574.

Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S. et E.A. Shale (2001), Pitfalls and protocols in DEA, *European Journal of Operational Research*, 132, 245-259.

Firth, D. (1993), Bias reduction of maximum likelihood estimates, *Biometrika*, 80 (1), 27-38.

Halkos, G. et N. Tzeremes (2009), Exploring the existence of Kuznets curve in countries' environmental efficiency using DEA window analysis, *Ecological Economics*, 68 (7), 2168-2176

Heinze, G. et M. Schemper (2002), A solution to the problem of separation in logistic regression, *Statistics in Medicine*, 21, 2409-2419.

Lavoie, M., Grenier, G., et S. Coulombe (1987), Discrimination and Performance Differentials in the National Hockey League, *Canadian Public Policy*, 13 (4), 407-422.

Leibenstein, H. et S. Maital (1992), X-inefficiency after a Quarter of a Century, *American Economic Review*, 82 (2), 428-433.

Mann, H.B. et D.R. Whitney (1947), On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other, *The Annals of Mathematical Statistics*, 18 (1), 50-60.

Meeusen, W. et J. van den Broeck (1977), Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18(2), 435-444.

Pearson, K. (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine Series 5*, 50 (302), 157–175.

Radovanović, S., Radojičić, M., Jeremić, V. et G. Savić (2013), A Novel Approach in Evaluating Efficiency of Basketball Players, *Journal for Theory and Practice Management*, 67.

Simar, L. et P.W. Wilson (1998), Sensitivity Analysis of Efficiency Scores : How to Bootstrap in Nonparametric Frontier Models, *Management Science*, 44 (1), 49-61.

Simar, L. et P.W. Wilson (2000), A general methodology for bootstrapping in non-parametric frontier models, *Journal of Applied Statistics*, 27 (6), 779-802.

Staudohar, P.D. (2005), The hockey lockout of 2004-05, *U.S. Bureau of Labor Statistics: Monthly Labor Review*, December 2005, 23-29

Wilcoxon, F. (1945), Individual comparisons by ranking methods, *Biometrics Bulletin*, 1 (6), 80-83.

Yang, H.H. et C.Y. Chang (2009), Using DEA window analysis to measure efficiencies of Taiwan's integrated telecommunication firms, *Telecommunications Policy*, 33 (1-2), 98-108.

## **Articles électroniques**

### **Articles scientifiques**

King, G. et L. Zeng (2001), Logistic Regression in Rare Events Data, *Society for Political Methodology* [En ligne] <http://gking.harvard.edu/files/0s.pdf> [Page consultée le 18 novembre 2014].

Halkos, G. et N. Tzeremes (2012), Evaluating professional tennis players' career performance: A Data Envelopment Analysis approach, *Munich Personal RePEc Archive* [En ligne] [https://mpra.ub.uni-muenchen.de/41516/1/MPRA\\_paper\\_41516.pdf](https://mpra.ub.uni-muenchen.de/41516/1/MPRA_paper_41516.pdf) [Page consultée le 18 novembre 2014].

Ryder, A. (2003), *Player Contribution* [En ligne] [http://www.hockeyanalytics.com/Research\\_files/Player\\_Contribution\\_System.pdf](http://www.hockeyanalytics.com/Research_files/Player_Contribution_System.pdf) [Page consultée le 18 novembre 2014].

Ryder, A. (2004), *Shot Quality* [En ligne] [http://hockeyanalytics.com/Research\\_files/Shot\\_Quality.pdf](http://hockeyanalytics.com/Research_files/Shot_Quality.pdf) [Page consultée le 18 novembre 2014].

### **Articles non scientifiques**

- Berri, D. (2012), *Is Changing the Coach Really the Answer?* [En ligne] <http://freakonomics.com/2012/12/21/is-changing-the-coach-really-the-answer/> [Page consultée le 8 Juin 2015].
- Desjardins, G. (2011), *Impact of Winning an Offensive Zone Faceoff, Even-Strength vs Power-Play* [En ligne] <http://www.arcticicehockey.com/2011/10/18/2491154/impact-of-winning-an-offensive-zone-faceoff-even-strength-vs-power> [Page consultée le 20 Mai 2015].
- Lipson, D. (2014), *Why Plus/Minus is the Worst Statistic in Hockey* [En ligne] <http://www.arcticicehockey.com/2014/6/5/5602668/why-plus-minus-is-the-worst-statistic-in-hockey> [Page consultée le 18 novembre 2014].
- Ryder, A. (2008), *The Ten Laws of Hockey Analytics* [En ligne] <http://hockeyanalytics.com/2008/01/the-five-laws-of-hockeynomics/> [Page consultée le 4 avril 2015].
- Staples, D. (2013), *Just how “horse shit” is the NHL’s official plus/minus stat?* [En ligne] <http://blogs.edmontonjournal.com/2013/05/13/just-how-horse-shit-is-the-nhls-official-plus-minus-stat/> [Page consultée le 18 novembre 2014].
- Tolensky, D. (2008), *Inside Malkin’s Contract: An In-Depth Look at Entry Level SPCs* [En ligne] [http://www.hockeybuzz.com/blog.php?post\\_id=16855](http://www.hockeybuzz.com/blog.php?post_id=16855) [Page consultée le 3 mars 2015].
- Tulsky, E. (2012), *The Importance of Quality of Competition* [En ligne] <http://nhlnumbers.com/2012/7/23/the-importance-of-quality-of-competition> [Page consultée le 11 juin 2015].
- Walthert, M. (2014), *How Hockey Got Moneyballed* [En ligne] <http://motherboard.vice.com/read/nhl-hockey-stats-have-finally-been-moneyballed> [Page consultée le 18 novembre 2014].

## Documents légaux

National Hockey League (2005), *Collective Bargaining Agreement Between National Hockey League and National Hockey League Players’ Association*.

Sportsdocuments.com (2013), *NHL Constitution*, art. 2.2. [En ligne] <http://sportsdocuments.com/2013/11/nhl-constitution/> [Page consultée le 12 octobre 2015].

## Logiciels

SAS Institute, Cary, NC, SAS software, version 9.3

R Foundation for Statistical Computing (2015), *R: A Language and Environment for Statistical Computing*, [En ligne] <http://www.R-project.org>

Simm, J. et G. Besstremyannaya (2015). *rDEA: Robust Data Envelopment Analysis (DEA) for R*, R package version 1.1-1.

## Monographies

Bogetoft, P. et L. Otto (2011), *Benchmarking with DEA, SFA, and R*. Springer New York.

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J. et G.E. Battese (2005), *An Introduction to Efficiency and Productivity Analysis*. Springer, New York.

Cooper, W.W., Seiford, L.M. et J. Zhu (2011), *Handbook on Data Envelopment Analysis*. Springer, New York.

Zhu, J. et W.D. Cook (2007), *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*. Springer, New York.

## Annexe 1: Calcul de la contribution d'un joueur (Ryder, 2003)

Pour construire la mesure de contribution du joueur pour une saison donnée, il faut commencer par déterminer le nombre de buts marginaux marqués et empêchés pour chaque équipe, c'est-à-dire le nombre de buts marqués/empêchés au-dessus d'un certain seuil. Ces éléments sont utilisés par la suite pour ajuster la contribution individuelle de chaque joueur.

Les formules sont les suivantes :

$$\text{Buts marginaux marqués} = BMM = BP - \overline{BP} * \text{importance de la défense au hockey}$$

$$\text{Buts marginaux empêchés} = BME$$

$$= \overline{BP} * (1 + \text{importance de la défense au hockey}) - BA$$

où  $BP$  = buts marqués par l'équipe (buts pour),  $BA$  = buts alloués par l'équipe et  $\overline{BP}$  = buts marqués en moyenne par équipe.

Les cinq joueurs sur la glace (sauf le gardien) passent en moyenne la moitié de leur temps à défendre leur zone et à attaquer la zone adverse, alors que le gardien passe tout son temps à défendre son territoire. Donc, pour les six joueurs sur la glace, l'importance de l'attaque est de

$$\frac{5}{12} = \frac{1}{6} * 0 + \frac{5}{6} * 0,5 \text{ et celle de la défense est de } \frac{7}{12} = \frac{1}{6} * 1 + \frac{5}{6} * 0,5. \text{ Les buts marginaux}$$

empêchés peuvent être décomposés de la façon suivante :

$$BME = \text{buts marginaux empêchés par le gardien de but (BMG)}$$

$$+ \text{buts marginaux empêchés par les autres joueurs (BMA)}$$

Un joueur peut prévenir des buts en minimisant la quantité et la qualité des tirs par un bon positionnement sur la patinoire, ce qui est très difficile à mesurer, alors qu'un gardien de but

peut prévenir des buts en arrêtant les tirs au but, ce qui est beaucoup plus simple à mesurer.

Ainsi, après avoir trouvé  $BME$  et  $BMG$ ,  $BMA$  sera trouvé de façon résiduelle :  $BMA = BME - BMG$ . La formule suivante détermine  $BMG$  :

$$BMG = (TAB - BFD) * (1 - Seuil) - (TAB(1 - PAAQT) - BFD)$$

où  $TAB$  = tirs au but,  $BFD$  = buts dans un filet désert,  $PAAQT$  = pourcentage d'arrêts ajusté pour la qualité des tirs au but<sup>13</sup> et  $Seuil$  = pourcentage d'arrêts d'un gardien jugé tout juste assez bon pour jouer dans la LNH (ex : pour la saison 2002-2003,  $Seuil = 89,3\%$ ).

Essentiellement,  $BMG$  compare le nombre de buts alloués par les gardiens d'une équipe (ajusté pour la qualité des tirs reçus) contre le nombre de buts alloués si l'équipe avait des gardiens tout juste assez bons pour jouer dans la LNH. La différence entre les deux donne le nombre de buts empêchés par les gardiens de but.

Après avoir obtenu les mesures  $BMM$ ,  $BMG$  et  $BMA$ , il est possible d'allouer les buts marqués/empêchés aux joueurs d'une équipe. Tout d'abord, un joueur peut marquer des buts en comptant le but ou faisant la passe menant au but. En supposant que les passes menant à un but sont aussi importantes que le but lui-même (donc, ces deux éléments ont un poids de 0,5), les buts marqués ( $BM$ ) par un joueur sont ainsi donnés par la formule suivante :

$$BM = \left( 0,5 * \text{buts comptés par le joueur} + 0,5 * \frac{\text{passes du joueur menant à un but}}{\text{nombre moyen de passes par but de l'équipe}} \right)$$

---

<sup>13</sup>  $PAAQT$  est estimé à partir d'une mesure de la qualité des tirs reçus par une équipe, construite dans un autre papier de Ryder (2004).

Il faut ensuite comparer cette mesure à un seuil correspondant au nombre de buts marqués par un joueur tout juste assez bon pour jouer dans la LNH (le joueur marginal). Pour ce faire, il faut contrôler pour le temps de jeu du joueur ( $T$ ), l'importance de la défense au hockey ( $\frac{7}{12}$ ) et le taux de création de buts par position ( $TCB$ ), car un attaquant a un taux de création de buts plus élevé qu'un défenseur :

$$BMM_{\text{joueur}} = BM - \text{Seuil} = BM - \frac{7}{12} * TCB * T$$

$$\text{où } TCB \text{ par position} = \frac{\text{Buts marqués par tous les joueurs jouant à cette position}}{\text{Temps de jeu total des joueurs jouant à cette position}}$$

Finalement, il faut contrôler pour le nombre de buts marginaux nécessaires pour obtenir une victoire, car ce nombre varie d'une équipe à l'autre (à la marge, une équipe faible a besoin de moins de buts supplémentaires pour obtenir une victoire de plus par rapport à une équipe forte) :

$$\text{Player Contribution Offense} = PCO = \frac{BMM_{\text{joueur}}}{\left( \frac{BMM_{\text{équipe}} + BME_{\text{équipe}}}{\text{Victoires}_{\text{équipe}}} \right)}$$

La mesure finale obtenue ( $PCO$ ) donne le nombre de victoires qu'un joueur a contribué à son équipe par sa capacité à marquer des buts. Par exemple, si  $PCO = 4$ , un joueur a contribué l'équivalent de 4 victoires à son équipe grâce aux buts qu'il a marqués. Il est important de noter que Ryder (2003) multiplie cette valeur par 20, simplement pour avoir des nombres plus gros et plus faciles à comparer.

Un joueur peut également empêcher des buts. Dans le même esprit que la mesure  $PCO$ , la mesure  $PCD$  (*Player Contribution Defense*) donne le nombre de victoires qu'un joueur contribue à l'équipe grâce au nombre de buts que celui-ci a empêché en surplus d'un seuil minimal. Le seuil minimal dépend de l'importance de la défense au hockey ( $\frac{7}{12}$ ), du temps de jeu ( $T$ ), du nombre moyen de buts alloués par équipe ( $\overline{BA}$ ) et du taux de buts alloués selon la position du joueur ( $TBA$ ) :

$$Seuil = (1 + importance\ de\ la\ defense) * \overline{BA} * \frac{T * TBA}{\sum_{équipe} (T * TBA)}$$

Le nombre de buts dont le joueur est responsable ( $BR$ ) dépend du taux de responsabilité du joueur ( $TR$ , en moyenne, un défenseur est davantage coupable d'un but alloué qu'un attaquant), du nombre de buts alloués lorsque le joueur est sur la patinoire ( $BAJ$ ) et du nombre de buts alloués par l'équipe ( $BA$ ) :

$$BR = BA * \frac{BAJ * TR}{\sum_{équipe} (BAJ * TR)}$$

Ainsi, si  $\frac{BMG}{BME}$  représente le pourcentage de buts empêchés par les gardiens de but d'une équipe,

la formule suivante donne la mesure  $PCD$  pour un joueur :

$$PCD = \frac{\left(1 - \frac{BMG}{BME}\right) * (Seuil - BR)}{\left(\frac{BMM_{équipe} + BME_{équipe}}{Victoires_{équipe}}\right)}$$

La formule utilise  $(Seuil - BR)$  et non  $(BR - Seuil)$ , car  $Seuil > BR$  dans la plupart des cas, c'est-à-dire qu'un joueur alloue généralement moins de buts à l'adversaire ( $BR$ ) lorsqu'il est sur la patinoire qu'un joueur marginal ( $Seuil$ ).

Finalement, un gardien de buts empêche des buts au-dessus d'un seuil minimal en ayant un pourcentage d'arrêts plus élevé que le pourcentage d'arrêts d'un gardien de but tout juste assez bon pour jouer dans la LNH :

$$Player\ Contribution\ Goaltending = PCG = \frac{TAB(PAAQT - Seuil)}{\left(\frac{BMM_{\text{équipe}} + BME_{\text{équipe}}}{Victoires_{\text{équipe}}}\right)}$$

## Annexe 2 : Corrélation entre les variables utilisées dans les régressions

Tableau 8. Corrélation entre les variables utilisées aux tableaux 1 et 5

**Nombre d'équipes: n=120**

Variables	Points au classement	Différentiel de buts pour/contre	% d'efficacité avantage numérique	% d'efficacité désavantage numérique	% d'efficacité mises en jeu	Mises en échec	Tirs bloqués	Pertes de possession	Reprises de possession	Efficienne moyenne des joueurs
Points au classement	1,00	0,94*	0,48*	0,40*	0,37*	0,02	-0,32*	-0,06	-0,17	0,43*
Différentiel de buts pour/contre	0,94*	1,00	0,52*	0,43*	0,32*	-0,01	-0,30*	-0,10	-0,20*	0,34*
% d'efficacité avantage numérique	0,48*	0,52*	1,00	-0,03	0,21*	-0,05	-0,01	0,23*	-0,04	0,31*
% d'efficacité désavantage numérique	0,40*	0,43*	-0,03	1,00	0,23*	0,05	-0,24*	-0,17	-0,19*	0,06
% d'efficacité mises en jeu	0,37*	0,32*	0,21*	0,23*	1,00	0,03	-0,33*	-0,06	0,00	0,15
Mises en échec	0,02	-0,01	-0,05	0,05	0,03	1,00	0,26*	-0,14	-0,02	0,07
Tirs bloqués	-0,32*	-0,30*	-0,01	-0,24*	-0,33*	0,26*	1,00	0,18*	0,34*	-0,29*
Pertes de possession	-0,06	-0,10	0,23*	-0,17	-0,06	-0,14	0,18*	1,00	0,28*	-0,02
Reprises de possession	-0,17	-0,20*	-0,04	-0,19*	0,00	-0,02	0,34*	0,28*	1,00	0,08
Efficienne moyenne des joueurs	0,43*	0,34*	0,31*	0,06	0,15	0,07	-0,29*	-0,02	0,08	1,00

\* indique une valeur-p < 0,05

**Tableau 9. Corrélation entre les variables utilisées au tableau 7**

**Rachat de contrat: n=35**

**Pas de rachat de contrat: n=1527**

<b>Variables</b>	Rachat de contrat	Efficience après bootstrap	Âge	Nombre d'années restant au contrat	Le contrat est de type «Standard»	Originaire d'Amérique du Nord	Revenus de l'équipe	Parties non jouées	Points de l'équipe au classement	Gardien de but
Rachat de contrat	1,00	-0,10*	0,11*	-0,06*	0,06*	0,06*	0,00	0,11*	-0,07*	0,01
Efficience après bootstrap	-0,10*	1,00	-0,27*	-0,11*	-0,09*	0,06*	-0,03	0,20*	0,12*	0,01
Âge	0,11*	-0,27*	1,00	0,03	0,44*	-0,07*	0,03	-0,13*	0,08*	0,08*
Nombre d'années restant au contrat	-0,06*	-0,11*	0,03	1,00	0,18*	-0,07*	0,04	-0,13*	0,07*	0,02
Le contrat est de type «Standard»	0,06*	-0,09*	0,44*	0,18*	1,00	-0,01	0,03	-0,21*	0,06*	0,00
Originaire d'Amérique du Nord	0,06*	0,06*	-0,07*	-0,07*	-0,01	1,00	-0,06*	0,01	0,00	-0,06*
Revenus de l'équipe	0,00	-0,03	0,03	0,04	0,03	-0,06*	1,00	-0,02	0,19*	-0,02
Parties non jouées	0,11*	0,20*	-0,13*	-0,13*	-0,21*	0,01	-0,02	1,00	-0,09*	0,37*
Points de l'équipe au classement	-0,07*	0,12*	0,08*	0,07*	0,06*	0,00	0,19*	-0,09*	1,00	0,00
Gardien de but	0,01	0,01	0,08*	0,02	0,00	-0,06*	-0,02	0,37*	0,00	1,00

\* indique une valeur-p < 0,05

## Annexe 3 : Variables entrées une à une dans les régressions

Tableau 10. Variables entrées une à une dans les régressions aux tableaux 1 et 5

<b>Nombre d'équipes: n=120</b>			
<b>Variable</b>	<b>Coefficient</b>	<b>Erreur standard</b>	<b>Valeur-p</b>
Constante	91,6024	0,3987	<0,001*
Différentiel de buts pour/contre	0,3474	0,0118	<0,001*
Constante	92,5944	3,0927	<0,001*
Différentiel de buts pour/contre	0,3498	0,0139	<0,001*
% d'efficacité avantage numérique	-0,0545	0,1684	0,747
Constante	98,1651	15,1416	<0,001*
Différentiel de buts pour/contre	0,3530	0,0164	<0,001*
% d'efficacité avantage numérique	-0,0766	0,1790	0,669
% d'efficacité désavantage numérique	-0,0632	0,1681	0,708
Constante	77,7576	16,9821	<0,001*
Différentiel de buts pour/contre	0,3468	0,0162	<0,001*
% d'efficacité avantage numérique	-0,1190	0,1760	0,500
% d'efficacité désavantage numérique	-0,1177	0,1661	0,480
% d'efficacité mises en jeu	0,5128	0,2083	0,015*
Constante	76,3307	17,1001	<0,001*
Différentiel de buts pour/contre	0,3470	0,0162	<0,001*
% d'efficacité avantage numérique	-0,1140	0,1764	0,520
% d'efficacité désavantage numérique	-0,1237	0,1665	0,459
% d'efficacité mises en jeu	0,5073	0,2087	0,017*
Mises en échec	0,0012	0,0015	0,422
Constante	81,9139	18,0249	<0,001*
Différentiel de buts pour/contre	0,3431	0,0167	<0,001*
% d'efficacité avantage numérique	-0,0800	0,1798	0,657
% d'efficacité désavantage numérique	-0,1323	0,1667	0,429
% d'efficacité mises en jeu	0,4455	0,2180	0,043*
Mises en échec	0,0017	0,0016	0,288
Tirs bloqués	-0,0030	0,0030	0,329

\* indique une valeur-p < 0,05

**Tableau 10 (suite)**

<b>Variable</b>	<b>Coefficient</b>	<b>Erreur standard</b>	<b>Valeur-p</b>
Constante	80,2671	17,9441	<0,001*
Différentiel de buts pour/contre	0,3475	0,0169	<0,001*
% d'efficacité avantage numérique	-0,1624	0,1864	0,386
% d'efficacité désavantage numérique	-0,1290	0,1657	0,438
% d'efficacité mises en jeu	0,4473	0,2167	0,041*
Mises en échec	0,0021	0,0016	0,191
Tirs bloqués	-0,0037	0,0030	0,234
Pertes de possession	0,0042	0,0027	0,123
Constante	80,1280	18,0181	<0,001*
Différentiel de buts pour/contre	0,3477	0,0169	<0,001*
% d'efficacité avantage numérique	-0,1558	0,1880	0,409
% d'efficacité désavantage numérique	-0,1236	0,1670	0,461
% d'efficacité mises en jeu	0,4320	0,2216	0,054
Mises en échec	0,0021	0,0016	0,183
Tirs bloqués	-0,0040	0,0032	0,215
Pertes de possession	0,0039	0,0028	0,156
Reprises de possession	0,0014	0,0038	0,717
Constante	51,2490	18,2681	0,006*
Différentiel de buts pour/contre	0,3379	0,0160	<0,001*
% d'efficacité avantage numérique	-0,3214	0,1805	0,078
% d'efficacité désavantage numérique	-0,0873	0,1565	0,578
% d'efficacité mises en jeu	0,5153	0,2083	0,015*
Mises en échec	0,0008	0,0015	0,588
Tirs bloqués	0,0007	0,0032	0,825
Pertes de possession	0,0046	0,0026	0,078
Reprises de possession	-0,0028	0,0037	0,453
Efficiencce moyenne des joueurs	0,3761	0,0917	<0,001*

\* indique une valeur-p < 0,05

**Tableau 11. Variables entrées une à une dans la régression au tableau 7**

**Nombre d'équipes: n=120**

Variable	Rapport des chances (odds ratio)	Intervalle de confiance (95%)	Valeur-p
Efficiencia après bootstrap	0,02	(0,00 , 0,13)	<0,001*
Efficiencia après bootstrap	0,05	(0,01 , 0,39)	0,004*
Âge	1,14	(1,05 , 1,23)	<0,001*
Efficiencia après bootstrap	0,03	(0,00 , 0,27)	0,001*
Âge	1,12	(1,04 , 1,21)	0,003*
Nombre d'années restant au contrat	0,52	(0,32 , 0,84)	0,008*
Efficiencia après bootstrap	0,03	(0,00 , 0,28)	0,002*
Âge	1,13	(1,03 , 1,23)	0,007*
Nombre d'années restant au contrat	0,50	(0,31 , 0,81)	0,005*
Le contrat est de type «Standard»	3,30	(0,92 , 11,87)	0,068
Efficiencia après bootstrap	0,03	(0,00 , 0,24)	<0,001*
Âge	1,14	(1,04 , 1,24)	0,004*
Nombre d'années restant au contrat	0,50	(0,31 , 0,81)	0,005*
Le contrat est de type «Standard»	3,48	(0,96 , 12,63)	0,058
Originaire d'Amérique du Nord	3,36	(1,25 , 9,03)	0,016*
Efficiencia après bootstrap	0,03	(0,00 , 0,24)	<0,001*
Âge	1,14	(1,04 , 1,24)	0,004*
Nombre d'années restant au contrat	0,50	(0,31 , 0,81)	0,004*
Le contrat est de type «Standard»	3,46	(0,96 , 12,48)	0,058
Originaire d'Amérique du Nord	3,38	(1,27 , 9,03)	0,015*
Revenus de l'équipe	1,00	(0,99 , 1,01)	0,726
Efficiencia après bootstrap	0,02	(0,00 , 0,12)	<0,001*
Âge	1,14	(1,04 , 1,25)	0,004*
Nombre d'années restant au contrat	0,57	(0,36 , 0,91)	0,019*
Le contrat est de type «Standard»	3,69	(1,03 , 13,25)	0,045*
Originaire d'Amérique du Nord	3,19	(1,19 , 8,55)	0,021*
Revenus de l'équipe	1,00	(0,99 , 1,01)	0,731
Parties non jouées	1,04	(1,03 , 1,06)	<0,001*

\* indique une valeur-p < 0,05

**Tableau 11 (suite)**

<b>Variable</b>	<b>Coefficient</b>	<b>Intervalle de confiance (95%)</b>	<b>Valeur-p</b>
Efficiencce après bootstrap	0,02	(0,00 , 0,17)	<0,001*
Âge	1,15	(1,05 , 1,25)	0,003*
Nombre d'années restant au contrat	0,58	(0,36 , 0,91)	0,019*
Le contrat est de type «Standard»	3,62	(1,02 , 12,82)	0,046*
Originaire d'Amérique du Nord	3,14	(1,18 , 8,36)	0,022*
Revenus de l'équipe	1,00	(0,99 , 1,02)	0,487
Parties non jouées	1,04	(1,02 , 1,06)	<0,001*
Points de l'équipe au classement	0,98	(0,95 , 1,00)	0,080
Efficiencce après bootstrap	0,01	(0,00 , 0,11)	<0,001*
Âge	1,15	(1,05 , 1,26)	0,002*
Nombre d'années restant au contrat	0,53	(0,33 , 0,86)	0,010*
Le contrat est de type «Standard»	3,46	(0,99 , 12,06)	0,051
Originaire d'Amérique du Nord	3,12	(1,17 , 8,34)	0,023*
Revenus de l'équipe	1,00	(0,99 , 1,02)	0,461
Parties non jouées	1,05	(1,03 , 1,08)	<0,001*
Points de l'équipe au classement	0,98	(0,95 , 1,01)	0,133
Gardien de but	0,24	(0,07 , 0,80)	0,020*

\* indique une valeur-p < 0,05