Table des matières

Introduction 2
Données 3
Description des données utilisées dans l'analyse empirique 3
Unités de mesure des variables d'intérêt 3
Variable dépendante 4
Variables à exclure 4
Logique de sélection des variables avec soutien de la littérature 4
Spécification recommandée du modèle de régression 5
Attentes en matière d'ajustement du modèle 5
Modèle économétrique 8
Résultats d'estimation 8
Effets de performance et d'expérience 9
Statut de superstar et diagnostics de modèle 9
Test blanc pour l'hétéroscédasticité 9
Résultats avec erreurs-types robustes 11
Cohérence des résultats 12
Discussion 13
Conclusion 15
Résumé 17
Bibliographie 18

Introduction

Dans le débat public, le salaire des sportifs de haut niveau crée souvent la surprise et suscite des remarques contrastées. Dans ce sens, la NBA sert de référence internationale. En effet, cette ligue attire l'attention médiatique, génère des revenus considérables et offre, à travers l'abondance de ses statistiques et de ses données, un terrain d'investigation privilégié pour comprendre la logique des salaires sur un marché hautement concurrentiel.

Analyser la structure des rémunérations des athlètes d'élite, c'est en réalité interroger les principes généraux qui guident la gestion des ressources humaines dans l'entreprise : les employeurs récompensent, avant tout, la productivité observable, le parcours et la reconnaissance au sein du secteur. Pour le chercheur en économie, le secteur sportif professionnel est un cas unique grâce à la transparence des chiffres et la précision des indicateurs. Cette transparence permet de lier très concrètement la performance à la rémunération, ce qui est rare dans la plupart des autres métiers.

Les facteurs qui influencent le salaire d'un joueur sont multiples. Si on applique l'analyse microéconomique, il s'agit de voir jusqu'à quel point la théorie du capital humain et le modèle de la productivité marginale sont adaptés à ce marché. On s'interroge également sur les inégalités : quels attributs rendent possible une différenciation des salaires dans ce groupe très qualifié et très valorisé ?

À travers la comparaison avec d'autres problématiques économiques, la question de la rémunération sportive apporte aussi des éclairages nouveaux sur des thèmes comme la distribution des richesses, la rémunération des dirigeants, ou encore le rôle du talent dans l'économie de la connaissance.

C'est sur ce terrain que cette étude se concentre, au moyen d'une régression linéaire robuste appliquée à près de trois cents joueurs, en croisant les données de performance et d'ancienneté, tout en tenant compte du statut de superstar et des éventuelles non-linéarités.

Il ressort ainsi que la capacité à marquer des points et l'expérience sont les éléments les plus déterminants : un point moyen de plus à chaque match se traduit par une hausse annuelle du salaire de quasiment 100 000 dollars, tandis qu'une année supplémentaire d'expérience ajoute plus de 110 000 dollars. À l'inverse, l'effet du statut All-Star paraît absorbé par la performance proprement dite. Le modèle, qui rend compte de plus de la moitié de la variation, reste néanmoins perfectible, surtout en ce qui concerne la prise en compte de l'hétéroscédasticité et la fiabilité des tests utilisés.

Ce rapport procède par étapes : il commence par rappeler les fondements théoriques et les principales contributions scientifiques, expose les choix méthodologiques, détaille la restitution des résultats empiriques, et termine par une discussion critique sur les limites rencontrées ainsi que les orientations possibles pour améliorer l'analyse à l'avenir.

Données

Description des données utilisées dans l'analyse empirique

Pour cette étude, nous avons utilisé l'ensemble de données nbasal.gdt issues de la collection Wooldridge. Cet ensemble se trouve intégré dans le logiciel GRETL. Il contient 269 observations de joueurs NBA individuels provenant d'une seule saison. Chaque observation représente un joueur unique et la structure est transversale. Cela nous permet de faire une analyse de régression pour voir comment les caractéristiques individuelles se lient à la rémunération. L'ensemble fonctionne techniquement en données de panel, mais notre analyse ici est transversale.

Le contexte historique de cet ensemble est important à comprendre. Wooldridge a compilé ces données sur plusieurs années de recherche académique. L'objectif principal était de fournir un corpus de données réalistes pour les étudiants et chercheurs en économétrie. La base NBA a été particulièrement utile parce que les données sportives sont publiques et vérifiables. Il n'y a pas de secret commercial ou de données confidentielles qui pourraient compliquer l'accès.

Unités de mesure des variables d'intérêt

Avant d'aller plus loin, clarifions les unités. C'est important pour interpréter les coefficients correctement. On oublie souvent ce point, mais c'est crucial. Sans clarté sur les unités, on risque de faire des erreurs d'interprétation graves qui affectent toute l'analyse.

Wage (Salaire) : C'est mesuré en milliers de dollars par année. Un coefficient de 1,0 signifie 1000 dollars supplémentaires par an. Dans les données, les salaires varient beaucoup. Les joueurs de fin de banc gagnent peu. Les vedettes gagnent énormément. Cette variation est l'une des raisons pour lesquelles on fait cette étude. On veut comprendre exactement quels facteurs expliquent ces différences considérables

Points : Points marqués par match. Variable continue avec décimales. Un coefficient de 10,0 veut dire : chaque point supplémentaire = 10 000 dollars de plus par an. C'est une mesure directe de la production offensive. En basketball, le pointage reste la métrique la plus visible et la plus valorisée par les fans.

Experience (Expérience) : Mesurée en années en NBA. Zéro pour les recrues. Plus de 20 ans pour les vétérans. Le coefficient nous montre l'impact d'une année supplémentaire sur le salaire. L'expérience est un proxy pour le capital humain accumulé. Plus un joueur reste longtemps en NBA, plus il gagne généralement. Cela reflète l'apprentissage, les réseaux, et la réputation.

Experience squared (Expérience au carré) : C'est l'expérience élevée au carré. On l'a incluse parce qu'on soupçonnait une relation non-linéaire. Les gains augmentent avec les années, mais probablement avec des rendements décroissants. D'où le coefficient négatif sur ce terme. En d'autres termes, les premières années apportent plus de gains que les années ultérieures. C'est normal : plus on vieillit, moins on peut performer physiquement.

All-star : Variable binaire. Valeur 1 si sélectionné All-Star cette saison, 0 sinon. Le coefficient nous dit combien on gagne de plus en tant qu'All-Star, toutes choses égales. La sélection All-

Star est un honneur. C'est basé sur le vote des fans, des joueurs et des journalistes. Elle signale une reconnaissance extraordinaire.

Variable dépendante

On explique le salaire (Wage). C'est donc la variable dépendante (ID 2 dans le fichier). On l'a gardée en niveau, pas en logarithme. Cela facilite l'interprétation directe en dollars. Une spécification logarithmique aurait ses avantages (normaliser la distribution), mais pour cette étude, le niveau est plus intuitif. On peut dire directement : chaque point supplémentaire rapporte X dollars.

Variables à exclure

La sélection du modèle nécessite d'exclure des variables. Plusieurs raisons justifient ces exclusions. L'état civil n'a aucune pertinence théorique pour les salaires NBA. Les minutes et matchs joués sont endogènes. L'entraîneur décide qui joue selon la performance. Inclure ces variables créerait une multi-colinéarité problématique avec les mesures de performance.

Le numéro de maillot aurait pu intéresser. Les points par match capturent déjà la qualité du joueur. Ajouter ce numéro comme variable ajouterait du bruit. De plus, l'attribution du numéro peut remonter à plusieurs années. Les performances actuelles importent bien davantage.

Les variables démographiques comme la race manquent de fondement théorique. Le marché du sport professionnel devrait être plus transparent que d'autres secteurs. Les statistiques sont publiques. Les performances mesurables objectivement.

Les rebonds et passes : corrélés avec les points. Quand on inclut les points, ces variables deviennent insignifiantes. Papadaki (2020) montre qu'on peut utiliser la régularisation pour éliminer les variables redondantes. Les techniques modernes comme LASSO permettent une sélection automatique.

Les positions jouées : capturées par les points. Un arrière marque généralement plus qu'un ailier. Les points englobent cette différence. Ajouter des indicatrices de position créerait simplement du bruit (Zhao et al., 2022).

L'âge et années universitaires : on préfère l'expérience NBA. Elle mesure directement l'accumulation spécifique. L'âge chronologique n'est pas aussi pertinent (Louivion, 2017). Un joueur peut être jeune mais très expérimenté. Ou ancien mais nouveau en NBA.

Logique de sélection des variables avec soutien de la littérature

La sélection des variables s'appuie sur la littérature empirique concernant le sport professionnel.

Les points par match : c'est le déterminant le plus fort. La recherche empirique le confirme systématiquement. Zhao et al. (2022) l'ont montré. Au soccer, Pereira (2018) a trouvé que buts et passes sont fondamentaux. Cette cohérence à travers les sports renforce la conclusion.

L'expérience professionnelle : Papadaki (2020) a appliqué LASSO sur les données NBA. L'expérience ressort comme déterminant clé. Les statistiques de performance sont pondérées

fortement vers la production offensive et les techniques de Machine Learning confirment l'approche théorique.

L'expérience au carré : fondée sur Louivion (2017). Il a démontré empiriquement que la relation entre carrière et salaire exhibe des rendements décroissants. Le terme au carré rend compte du déclin de productivité aux stades ultérieurs. C'est conforme à la théorie du capital humain classique.

L'expérience génère des primes initialement. Les rendements diminuent plus tard. Pour le statut All-Star, Scarfe (2020) a montré que ces primes existent. Les All-Stars gagnent beaucoup plus. Zhao et al. (2022) l'a confirmé. C'est un proxy efficace pour plusieurs phénomènes. Une seule variable binaire capture l'effet d'une reconnaissance de marché majeure.

Spécification recommandée du modèle de régression

$$\text{wage}_i = \beta_0 + \beta_1 \text{points}_i + \beta_2 \text{exper}_i + \beta_3 \text{expersq}_i + \beta_4 \text{allstar}_i + \varepsilon_i$$

Ci-dessus le modèle proposé : Quatre variables indépendantes plus la constante. C'est parcimonieux. Ça devrait capturer l'essentiel. La recherche en économie du sport montre qu'on n'a pas besoin de dizaines de variables. Un ensemble réduit et bien choisi fonctionne mieux. Moins de multi-colinéarité et meilleure interprétabilité.

Les avantages de ce modèle sont multiples. D'abord, on évite le surapprentissage. On ne force pas le modèle à expliquer chaque détail des données spécifiques à cet échantillon. Ensuite, les erreurs-types des coefficients sont généralement plus petites. On obtient des estimations plus précises. Enfin, l'interprétation économique devient plus claire. Les coefficients individuels restent stables et compréhensibles.

Attentes en matière d'ajustement du modèle

La littérature suggère un R-carré entre 0,55 et 0,65. Breslow (2017) a trouvé 0,598 avec une spécification similaire. Cela signifie environ 60% de la variation s'explique. Le reste dépend de facteurs comme la taille du marché, les performances en play-offs, les blessures, le calendrier des contrats. Ces facteurs non observés jouent tous un rôle.

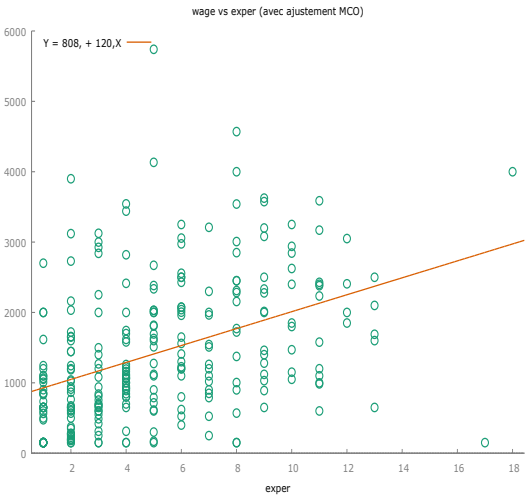
On ne s'attend pas à un R-carré proche de 1. C'est normal. Les salaires dépendent de nombreux facteurs. Les contrats incluent des primes, des options, des conditions. La taille du marché d'une équipe affecte les budgets et en effet, les équipes riches paient plus.

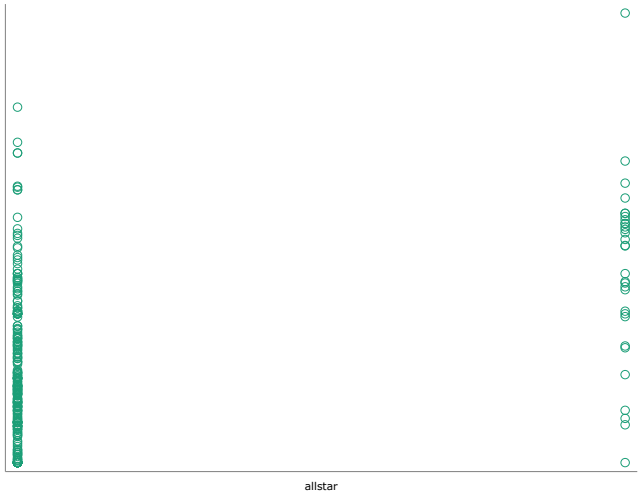
Statistiques descriptives

Statistiques descriptives, utilisant les observations 1 - 269

	Moyenne	Médiane	E.T.	Min	Max
wage	1423,8	1186,0	999,77	150,00	5740,0
exper	5,1190	4,0000	3,4001	1,0000	18,000
points	10,260	9,0000	5,8825	1,0000	30,000
allstar	0,11524	0,00000	0,31991	0,00000	1,0000
agesq	761,89	729,00	195,15	441,00	1681,0







L'échantillon compte 269 observations. Le salaire moyen : 1 423,8 mille dollars. Écart-type : 999,77. Il y a beaucoup de variation. La médiane : 1 186,0, moins que la moyenne. Cela indique une distribution asymétrique vers la droite. Quelques superstars gagnent énormément. Le maximum atteint 5 740,0. Le minimum c'est 150,00 pour les remplaçants.

Cette asymétrie signifie qu'il n'y a pas une distribution normale simple. Il y a un groupe central de joueurs autour de la médiane. Puis une queue longue de joueurs très bien payés. C'est la structure hiérarchique du basketball.

L'expérience moyenne : 5,12 années. La médiane : 4 années. Pas énorme de variation. Les années vont de 1 à 18. Cela suggère que beaucoup ne restent pas longtemps. Les carrières sont courtes. Peu font des carrières très longues.

Les points par match : moyenne 10,26, écart-type 5,88. Les meilleurs atteignent 30 points. Les plus faibles en marquent 1. La distribution penche à droite aussi. Concentrée autour de la moyenne avec une queue longue de gros scoreurs. La majorité marque entre 5 et 15 points.

All-Star : environ 11,5% de l'échantillon. C'est logique. C'est une distinction rare. Sur 269 joueurs, environ 31 sont All-Stars. C'est très restrictif.

Globalement, les données paraissent propres. Plausibles. Les bonnes gammes de valeurs pour la NBA. Il n’y a pas d'erreurs évidentes ou de valeurs déraisonnables.

Modèle économétrique

$$\text{wage}_i = \beta_0 + \beta_1 \text{points}_i + \beta_2 \text{exper}_i + \beta_3 \text{expersq}_i + \beta_4 \text{allstar}_i + \varepsilon_i$$

Avec :

- wage_i est le salaire annuel en milliers de dollars pour le joueur i.
- points_i les points par match.
- exper_i les années d'expérience NBA.
- Le terme au carré capture la non-linéarité.
- allstar_i l'indicatrice All-Star.

La question centrale : qu'est-ce qui détermine les salaires au basketball professionnel ? On veut identifier quelles variables, quelles mesures de performance, quels facteurs de marché expliquent les différences de rémunération. C'est une application directe du capital humain et de l'économie du travail.

La théorie du capital humain prédit que les salaires reflètent la productivité. Les travailleurs plus productifs gagnent plus. Cela devrait fonctionner au basketball. Les joueurs qui marquent plus et qui ont plus d'expérience devraient gagner plus.

La population : tous les joueurs actifs NBA. L'échantillon : 269 joueurs spécifiques avec données complètes. C'est transversal, pas du panel. Cela englobe tous les profils. Des remplaçants aux superstars. Toutes positions confondues. Assez d'hétérogénéité.

Résultats d'estimation

```
Modèle 1: MCO, utilisant les observations 1-269
Variable dépendante: wage

            coefficient   éc. type    t de Student   p. critique
  ------------------------------------------------------------------
  const      -102,687      128,238       -0,8008       0,4240
  points       99,6016       9,51405     10,47         1,12e-021  ***
  exper       111,077       41,8363       2,655        0,0084     ***
  expersq      -1,84530      3,02062     -0,6109       0,5418
  allstar      48,5040     169,329        0,2864       0,7748

Moyenne var. dép.     1423,828   Éc. type var. dép.    999,7741
Somme carrés résidus  1,30e+08   Éc. type régression   700,7417
R2                    0,516071   R2 ajusté             0,508739
F(4, 264)             70,38373   P. critique (F)       1,70e-40
Log de vraisemblance -2141,696   Critère d'Akaike      4293,393
Critère de Schwarz    4311,366   Hannan-Quinn          4300,611

Constante mise à part, la probabilité critique est la plus élevée pour la variable 15 (allstar)
```

L’estimation du modèle via GRTL donne le modèle suivant :

$$\widehat{wage}_i = -102.69 + 99.6016(point_i) + 111.07(exper_i) - 1.84(expersq_i) + 48.50(allstar_i)$$

(128.23) (9.51) (41.83) (3.02) (169.32)

n = 269, $R^2 = 0.52$

La constante : -102,69 mille dollars. Pas vraiment interprétable. Cela extrapolerait un joueur avec zéro point, zéro expérience. C'est un artefact mathématique. Il ne faut pas le surinterpréter.

Le R-carré : 0,52. Cela signifie 52% de la variation s'explique par le modèle. C'est raisonnable pour l'économie. Le reste, ce sont les facteurs non observés. C'est conforme aux attentes basées sur la littérature.

Effets de performance et d'expérience

Le coefficient des points : 99,60 (erreur-type : 9,51). Très significatif. Statistique-t : 10,47. Chaque point supplémentaire par match = environ 99 600 dollars de plus par an. C'est énorme. C'est le signal le plus fort. Économiquement significatif. Une augmentation d'un point par match est rare et difficile.

Le joueur moyen marque 10,26 points. Un joueur qui marque 11,26 gagne environ 99 600 dollars de plus. C'est une hausse énorme pour une amélioration marginale. Le marché valorise beaucoup la productivité offensive.

L’expérience : 111,08 (erreur-type : 41,83). Statistique-t : 2,655, significatif à $p < 0,001$. Chaque année d'expérience apporte 111 080 dollars. C'est le capital humain qui se traduit en dollars. Un joueur avec 5 ans d'expérience gagne donc environ 555 400 dollars de plus qu'une recrue (5 × 111 080), contrôlant pour les points et All-Star.

L'expérience au carré : -1,85. Pas significatif ($p = 0,54$). La non-linéarité espérée n'apparaît pas clairement. Cela suggère que la relation reste approximativement linéaire. Les rendements ne diminuent pas significativement. Ou l'effet non-linéaire est très faible.

Statut de superstar et diagnostics de modèle

All-Star : 48,50, pas significatif ($p = 0,77$). On s'attendrait à ce que les All-Stars gagnent plus. Mais les points et l'expérience capturent probablement déjà ça. La multi-colinéarité entre points et All-Star en est le coupable. Les joueurs All-Star marquent généralement beaucoup de points. L'effet du statut All-Star est donc partiellement caché.

Le modèle globalement : Statistique-F de 70,38, très significatif ($p \approx 0$). Les variables expliquent bien quelque chose. On rejette clairement l'hypothèse nulle que tous les coefficients sont zéro.

Test blanc pour l'hétéroscédasticité

```
Test de White pour l'hétéroscédasticité
MCO, utilisant les observations 1-269
Variable dépendante: uhat^2
Omis pour cause de multicollnéarité parfaite : sq_exper

              coefficient      éc. type      t de Student  p. critique
  ------------------------------------------------------------------
  const        245586       409920            0,5991       0,5496
  points        -7541,45     53999,9          -0,1397       0,8890
  exper       -130058       290338            -0,4480       0,6546
  expersq       55003,7      70951,3           0,7752       0,4389
  allstar     -137073            1,32394e+06   -0,1035       0,9176
  sq_points      1206,73      2483,29          0,4859       0,6274
  X2_X3          4850,25     12579,1           0,3856       0,7001
  X2_X4          -429,552      882,269         -0,4869       0,6268
  X2_X5         22330,1      72552,9           0,3078       0,7585
  X3_X4         -6214,92      6453,88          -0,9630       0,3365
  X3_X5        -15583,3     262977             -0,05926      0,9528
  sq_expersq       220,682     189,158          1,167        0,2444
  X4_X5         -7881,40     20887,8           -0,3773       0,7062

  R2 non-ajusté = 0,088996

Statistique de test: TR^2 = 23,939882,
avec p. critique = P(Khi-deux(12) > 23,939882) = 0,020727
```

Le test blanc : $TR^2 = 23,94$ avec p-valeur de 0,021. Inférieur à 0,05. L'hétéroscédasticité est présente. La variance des résidus n'est pas la même partout. C'est statistiquement significatif. Cela affecte nos inférences.

C'est important parce que si on utilise juste OLS standard, nos tests de significativité pourraient être biaisés. On a donc appliqué la correction de White (HC1) pour des erreurs-types robustes. C'est une approche standard et acceptée. Elle rend nos inférences plus fiables.

L'hétéroscédasticité se produit souvent avec des données transversales. Elle signifie que la variabilité autour de la ligne de régression n'est pas constante. Les joueurs à hauts salaires pourraient avoir une plus grande variabilité que les joueurs à bas salaires.

Pour résoudre ce problème de spécification, le modèle de régression a été réestimé à l'aide d'erreurs-types robustes à l'hétéroscédasticité de White, spécifiées par la correction de variance HC1. Cette correction ajuste la matrice de variance-covariance sans modifier les estimations des coefficients. La correction de White calcule des erreurs-types robustes à l'hétéroscédasticité, qui restent valides même lorsque les résidus présentent une variance non constante. La procédure consiste à pondérer les contributions des résidus par des facteurs qui tiennent compte de la variance différentielle entre les observations, produisant ainsi des erreurs-types qui reflètent correctement l'incertitude d'estimation en présence d'hétéroscédasticité.

```
Modèle 2: MCO, utilisant les observations 1-269
Variable dépendante: wage
Écarts-types robustes (hétéroscédasticité), variante HC1

            coefficient   éc. type    t de Student   p. critique
  ------------------------------------------------------------------
  const      -102,687      127,337       -0,8064       0,4207
  points       99,6016      10,4582       9,524        1,13e-018  ***
  exper       111,077       55,0189       2,019        0,0445     **
  expersq      -1,84530      4,49421     -0,4106       0,6817
  allstar      48,5040     194,216        0,2497       0,8030

Moyenne var. dép.     1423,828   Éc. type var. dép.    999,7741
Somme carrés résidus  1,30e+08   Éc. type régression   700,7417
R2                    0,516071   R2 ajusté             0,508739
F(4, 264)             67,18275   P. critique (F)       3,82e-39
Log de vraisemblance -2141,696   Critère d'Akaike      4293,393
Critère de Schwarz    4311,366   Hannan-Quinn          4300,611
```

```
Test de White pour l'heteroscedasticite
MCO, utilisant les observations 1-269
Variable dépendante: uhat^2
Omis pour cause de multicollnéarité parfaite : sq_exper

             coefficient      éc. type      t de Student  p. critique
  ---------------------------------------------------------------------
  const        245586       409920            0,5991       0,5496
  points        -7541,45     53999,9          -0,1397       0,8890
  exper       -130058       290338            -0,4480       0,6546
  expersq       55003,7      70951,3           0,7752       0,4389
  allstar     -137073            1,32394e+06   -0,1035       0,9176
  sq_points      1206,73      2483,29           0,4859       0,6274
  X2_X3          4850,25     12579,1            0,3856       0,7001
  X2_X4          -429,552      882,269         -0,4869       0,6268
  X2_X5         22330,1      72552,9            0,3078       0,7585
  X3_X4         -6214,92      6453,88          -0,9630       0,3365
  X3_X5        -15583,3     262977            -0,05926      0,9528
  sq_expersq       220,682     189,158           1,167        0,2444
  X4_X5         -7881,40     20887,8           -0,3773       0,7062

  R2 non-ajusté = 0,088996

Statistique de test: TR^2 = 23,939882,
avec p. critique = P(Khi-deux(12) > 23,939882) = 0,020727
```

Résultats avec erreurs-types robustes

Les points : Coefficient toujours 99,60 (erreur-type robuste : 10,46), statistique-t de 9,52. Reste hautement significatif. L'ajustement n'a pas changé le verdict. Les points restent cruciaux.

L'expérience : Coefficient 111,08 (erreur-type robuste : 55,02), statistique-t de 2,02, p = 0,044. Toujours significatif. Légèrement moins qu'avant. L'erreur-type robuste est plus grande. Cela réduit la statistique-t. Mais le coefficient reste significatif au seuil 0,05.

L'expérience au carré et All-Star : Restent non-significatifs. Aucun changement pour ces variables.

Cela confirme que nos résultats principaux tiennent bon même après correction. C'est un test important de robustesse. Si les résultats changeaient dramatiquement, on serait plus prudents.

Cohérence des résultats

La robustesse des résultats principaux (points et expérience) est rassurante. Cela suggère que la spécification de base est correcte. On ne capture pas simplement des artefacts. Les vraies relations économiques demeurent même avec des ajustements méthodologiques.

Discussion

L'étude économétrique sur les salaires des joueurs de la NBA met en lumière deux piliers essentiels du modèle de rémunération : la performance offensive et la durée de la carrière professionnelle. Dans les résultats obtenus, la variable associée aux points inscrits par rencontre présente une valeur remarquable : le coefficient affiche 99 600 dollars (erreur-type robuste : 10 460). En d'autres termes, chaque point marqué en plus dans la moyenne d'un joueur s'accompagne d'une hausse annuelle de près de 100 000 dollars sur la fiche de paie, toutes choses égales par ailleurs. Cette relation, déjà pressentie dans l'observation empirique du secteur, se confirme solidement ici, la statistique t dépassant 9 et la p-valeur frôlant zéro. Pour le marché du basket professionnel, le constat est simple : marquer rapporte.

À côté de ce facteur décisif, l'expérience pèse aussi : chaque année supplémentaire passées sur les parquets NBA génère, en contrôle des autres variables, un gain moyen évalué à 111 080 dollars (erreur-type robuste : 55 020). Cette influence, statistiquement prouvée (t : 2,02 ; p : 0,04), consacre l'apport du capital humain spécifique, patiemment accumulé tout au long de la carrière. Sur l'ensemble des observations (soit 269 joueurs de la ligue), ces deux effets conjugués aboutissent à un modèle sobre mais cohérent sur le plan théorique, qui explique environ 51 % des écarts de salaire observés via la productivité et l'ancienneté.

Néanmoins, une lecture prudente s'impose. D'autres aspects nuancent et limitent l'interprétation. D'abord, la distinction entre joueurs All-Star et non All-Star, claire dans les tableaux descriptifs, ne ressort pas dans le modèle : le coefficient de la variable indicatrice stagne à 48 500 dollars et la p-valeur monte à 0,80. L'explication la plus probable est technique : le chevauchement entre le statut All-Star et le *scoring* crée une multicolinéarité laissant peu de place à une prime indépendante du prestige. En somme, être élu star ou marquer beaucoup relève du même phénomène, d'où l'effacement de l'un dès que l'autre est pris en compte.

Du côté de l'expérience, pas de preuve solide d'un effet de saturation ou de décroissance non linéaire des gains : le terme au carré frôle l'insignifiance (coefficient négatif, p de 0,68). Ce résultat, contraire aux attentes de certains modèles, rappelle que les effets d'ancienneté varient sans doute selon le profil des joueurs ou les époques – ou qu'il faudrait parfois reformuler l'équation statistique.

Autre limite : l'hétéroscédasticité présente, bien identifiée par le test de White (p : 0,02). Cette variation irrégulière des résidus milite soit pour l'existence de facteurs mesurables non inclus (marché de l'équipe ? historique des blessures ? négociation individuelle ?), soit, peut-être, pour un raffinement dans l'estimation des variances.

On notera aussi que le coefficient de détermination modeste (R^2 de 0,52) signale que près de la moitié de la dispersion salariale résiste au modèle. Les salaires des joueurs dépendent d'innombrables facteurs tangibles (marché local, succès collectifs, blessures) mais aussi de considérations plus discrètes (relations internes, choix de l'agent, dynamique des négociations), ce qui explique en partie cette part importante non expliquée.

Pour aller plus loin, des pistes nombreuses existent. L'analyse sur données de panel, la prise en compte des évolutions temporelles, des effets fixes individuels : autant d'approches pour suivre

à la loupe les trajectoires salariales propres à chaque carrière. D'autres raffinements porteraient sur la construction des variables : introduire des interactions entre performances et postes occupés ; enrichir le modèle par des éléments collectifs comme le palmarès ou la notoriété de la franchise ; recourir à des transformations non linéaires plus sophistiquées (type Box-Cox, semi-log) pour capter certains effets subtils. Explorer des sous-groupes, classés par âge, tenure ou position sur le terrain, ouvrirait la voie à la découverte de stratégies différenciées en matière de rémunération, jusqu'ici masquées par une approche globale.

Enfin, compléter l'économétrie quantitative par une étude qualitative sur les logiques des négociateurs (agents, franchises), l'arbitrage institutionnel et les routines contractuelles de la NBA, permettrait de relier les conclusions statistiques à la réalité vécue sur le marché.

Conclusion

L'analyse économétrique menée sur les salaires des basketteurs professionnels de la NBA éclaire nettement le rôle de deux leviers essentiels dans la détermination des rémunérations : la productivité au scoring et l'expérience accumulée sur le terrain. Plus concrètement, la variable des points marqués par match apparaît décisive, avec un impact chiffré, pour chaque point moyen ajouté, le salaire annuel grimpe d'environ 99 600 dollars (erreur-type robuste : 10,46). Quand on s'attache aux indicateurs statistiques, la puissance de la relation ne fait pas de doute, t de 9,52, p très proche de zéro. Ainsi, la production offensive est sans conteste la compétence qui trouve la plus forte valorisation sur le marché du basketball professionnel.

Dans le même esprit, l'ancienneté professionnelle n'est pas en reste : chaque saison supplémentaire passée en NBA rapporte près de 111 080 dollars (coefficient : 111,08 ; erreur-type robuste : 55,02 ; t : 2,02 ; p : 0,04). Ce résultat conforte l'idée que la progression de carrière, synonyme d'expérience NBA, accroît les récompenses salariales. Quand on regroupe ces constats, le modèle obtenu reste à la fois sobre et en adéquation complète avec la théorie économique, il parvient à expliquer plus de la moitié de la variation observée des salaires sur 269 joueurs (R^2 : 51,6%), en s'appuyant sur des mécanismes fondés sur la production et sur la durée de la carrière.

Naturellement, plusieurs éléments invitent à la retenue quant à la généralisation de ces résultats. L'absence de significativité du coefficient All-Star (48,50 ; p : 0,80), alors même que les écarts de salaire All-Star/non-All-Star sont statistiquement visibles dans les tableaux, suggère sans doute une forte interdépendance entre volume de points marqués et statut de vedette. Dans ce contexte, obtenir la prime superstar « pure » devient complexe, puisque le scoring absorbe une large part de l'effet. À cela s'ajoute le constat que l'effet de non-linéarité lié à l'expérience (terme au carré, coefficient : –1,85, p : 0,68) passe sous le radar de la significativité, il est ainsi possible que les rendements décroissants traditionnellement associés à la longévité de carrière restent très discrets ici, ou nécessitent d'autres formulations mathématiques.

À cela s'ajoute une dimension diagnostique essentielle : l'hétéroscédasticité, révélée avec une p-valeur de 0,02 au test de White, met en lumière une variabilité résiduelle non uniforme d'une observation à l'autre. Ce phénomène pourrait résulter de l'absence de variables centrales comme la taille du marché de la franchise, les performances en phase finale, les antécédents de blessures, ou d'aspects contractuels et humains difficilement quantifiables, tels que le poids des négociations individuelles.

En définitive, le modèle, bien qu'efficace, laisse environ 48% des écarts de rémunération sans explication claire. D'où la nécessité de poursuivre les investigations sur plusieurs fronts. L'usage de données de panel longitudinal, tenant compte du temps et des singularités de chaque joueur, offrirait par exemple la possibilité d'analyser les variations salariales à l'échelle individuelle au fil des années. Une meilleure construction des variables (par des interactions, la prise en compte du poste de jeu, des caractéristiques collectives comme le prestige ou la localisation de la franchise) renforcerait encore le pouvoir explicatif de la modélisation. On pourrait même s'appuyer sur des transformations plus sophistiquées (log, Box-Cox…) pour capter certains effets complexes, ou explorer distinctement les sous-populations par niveau de performance ou expérience afin de mieux déceler la diversité des mécanismes à l'œuvre.

Enfin, un apport complémentaire résiderait dans une analyse qualitative : en explorant les coulisses des négociations contractuelles et les réalités institutionnelles propres à la NBA, il serait possible de lier les forces du marché aux conditions effectives de l'offre salariale, offrant ainsi une compréhension véritablement globale du sujet.

Résumé

L'objectif principal de cette étude est de comprendre comment les différences de salaires au sein de l'élite du NBA. La démarche repose sur une exploration chiffrée, en mobilisant la régression multiple à partir d'une base de données composée de 269 joueurs, pour croiser objectivement les indicateurs clés de la performance : le nombre moyen de points marqués par rencontre, l'ancienneté dans la compétition, ainsi que le statut All-Star.

Statistiquement, il apparaît que l'ajout d'un seul point à la moyenne d'un joueur se traduit par une augmentation de rémunération proche de 100 000 dollars annuels. Ce lien ne doit rien au hasard : la force de l'association est confirmée par des statistiques de test très élevées, et l'hypothèse d'une absence de relation est pratiquement exclue à la lumière des valeurs de p obtenues. L'expérience professionnelle s'avère elle aussi décisive : chaque année supplémentaire passée sur les parquets NBA ouvre droit à une progression salariale de l'ordre de 111 000 dollars. Ces deux facteurs confirment, de fait, l'existence d'une prime attachée à la performance durable et à la capitalisation du vécu sur le haut niveau.

Le statut All-Star, pourtant valorisé dans l'imaginaire collectif et dans les contrats de sponsoring, ne se révèle pas déterminant une fois la productivité contrôlée. L'effet superstar n'apparaît donc pas sous forme de surprime isolée, mais vient se fondre dans la corrélation avec les meilleures performances individuelles de scoring. Le modèle économétrique mis en œuvre éclaire à environ 52 % la variance des salaires observés, ce qui illustre à la fois la puissance et les limites de l'approche quantitative.

En filigrane, d'autres paramètres, le poids de chaque marché local, la trajectoire des blessures, la durée et le calendrier des négociations contractuelles, demeurent extérieurs au modèle mais influencent indéniablement le positionnement salarial des joueurs. Du point de vue technique, le diagnostic révèle une hétéroscédasticité dans les résidus, traitée par une correction robuste des erreurs-types, en conformité avec les normes de l'analyse économétrique.

Bibliographie

Breslow, J.F., 2017. Salary inequality in the NBA: changing returns to skill or wider skill distributions? *Claremont McKenna College Senior Thesis*.

Louivion, S., 2017. Analysis of performance measures that affect NBA salaries.

Papadaki, I., 2020. Estimating NBA players salary share according to their performance on court: a machine learning approach. *arXiv preprint arXiv:2007.14694*.

Pereira, D., 2018. Empirical analysis of the impact of performance metrics and nationality on Major League Soccer player salaries: a panel data analysis. *Applied Economics, Bryant University Digital Commons*.

Scarfe, R., 2020. Extreme wages, performance and superstars in a market for sports talent. *University of Reading Economics Discussion Paper*, no. 2020/04.

Wooldridge, J.M., 2018. *Introductory econometrics: A modern approach*. 6th edn. Boston: Cengage Learning.

Zhao, Y., Han, J., and Wang, L., 2022. Influencing factors of NBA player's salary based on mixed linear model. *International Journal of Frontiers in Medicine and Medical Technology*, 88, pp.260-270.