

Evaluation économétrique de la valeur marchande d'un attaquant de football.

Introduction:

En 2019, pour la première fois de l'histoire, les clubs de football ont dépensé plus de €10 milliards en indemnités de transfert. Ce montant a plus que triplé en une décennie en parallèle à l'augmentation des recettes des clubs. Contrairement à ce qui est souvent avancé par les personnes extérieures au football, les montants des transferts ne sont pas irrationnels. Ils sont fixés par les acteurs du marché sur la base de critères en grande partie objectifs et donc statistiquement modélisables. Le joueur devient l'élément économique central d'un club et les attaquants représente le mieux ce système, car ce sont eux qui ont la plus grande valeur marchande.

Nous allons essayer de voir *quels sont les liens et variables qui jouent sur la valeur marchande d'un attaquant de football.*

Les clubs ont en effet besoin de savoir comment estimer la valeur marchande de leurs joueurs pour construire leur capital. Là où le directeur sportif d'un club cherche d'abord à recruter sur l'aspect technique et sportif de l'attaquant, le club est une entreprise et se voit amené à chercher comment tirer de leurs joueurs le plus d'argent possible en cas de revente.

Pour répondre à cette problématique, nous avons décidé d'effectuer une régression par les moindres carrés ordinaires, où nous effectuerons plusieurs tests pour valider les hypothèses sous-jacentes à cette méthode, et valider ou non la prise en compte de certaines variables.

Le modèle et ses variables:

Nous avons pris de base cette régression :

$$Prix_t = \beta_0 + \beta_1 equipnat10 + \beta_3 taille + \beta_4 ldc + \beta_5 piedfort + \beta_6 champfr + \beta_7 champal + \beta_8 champan + \beta_9 champes + \beta_{10} champit + \beta_{11} ratiobm3ans + \beta_{12} ratiobm20 + \beta_{13} age + \beta_{14} fincontrat + \beta_{15} aboinsta + \beta_{16} cont$$

où $t = 1, \dots, 84$

Les variables explicatives sont les suivantes :

-equipnat10 : dichotomique égal à 1 si le joueur à plus de 10 sélections dans son équipe nationale, 0 sinon.

Valeur attendue : significative et positive, l'équipe nationale étant le regroupement des meilleurs joueurs du pays, si le joueur en fait partie, cela montre qu'il est sportivement très fort, ce qui influe sur sa valeur marchande.

-taille : taille du joueur.

Valeur attendue : non significative, la taille étant un aspect physique plutôt négligeable dans le football notamment à ce poste. Cela étant, il pourrait avoir plus d'importance si la régression portait sur les gardiens de buts et défenseurs)

-ldc : dichotomique égal à 1 si le joueur a participé à la campagne de ligue des champions 2020, 0 sinon.

Valeur attendue : significative positive, la ligue des champions étant la plus prestigieuse des compétitions, nous pouvons nous attendre que les joueurs qui participent à celle-ci comptent parmi les meilleurs attaquants et font évoluer positivement leurs cotes.

-piedfort : dichotomique égal à 1 si le joueur est droitier et 0 s'il est gaucher.

Valeur attendue : non significative, même si les gauchers sont plus rares, nous ne pensons pas que le prix d'un joueur puisse réellement varier entre un gaucher et un droitier.

-champfr, champal, champan, champes, champit, champ12: dichotomique reflétant dans quel championnat joue le joueur, respectivement: la France, l'Allemagne, l'Angleterre, l'Espagne, l'Italie, un autre championnat du top 12 européens (ici la Belgique, le Portugal, la Turquie).

Valeur attendue : nous pensons que les joueurs du championnat d'Angleterre auront une valeur marchande plus importante que les joueurs des autres championnats car leur marché est bien plus élevé que les autres, ou les transferts sont souvent plus chers. Nous estimons aussi que les footballeurs jouant dans le championnat français et hors top 5 seront moins cotés.

-ratiobm3ans : le ratio(moyenne) du nombre de but par match de l'attaquant sur les 3 dernières années.

Valeur attendue : significative positive : même si on ne juge pas un attaquant uniquement par ses statistiques, elles font partie intégrante de la valeur sportive du joueurs, donc de sa valeur marchande.

-ratiobm20 : le ratio(moyenne) du nombre de but par match en 2020.

Valeur attendue : significative positive, comme dit précédemment, les statistiques d'un joueur sont importantes, que ce soit sur les dernières années ou celle en cours.

-age : l'âge du joueur.

Valeur attendue : significative négative, les jeunes joueurs sont plus chers car la revente est plus facile, et que le joueur peut encore progresser et accroître sa valeur.

-fincontrat : durée en année de la fin de contrat du joueur avec son club.

Valeur attendue : significative positive, un joueur en fin de contrat peut partir gratuitement à la fin de saison, les clubs ont tout intérêt à baisser son prix de vente avant que son contrat se termine pour pouvoir en tirer du profit.

-aboinsta : le nombre d'abonnés Instagram du joueur.

Valeur attendue : non significative, en mettant cette variable, nous cherchions à mettre en avant le côté « populaire » du joueur et si cela a un lien avec sa valeur marchande. Même si le côté extra sportif est très important dans la valeur d'un joueur, nous pouvons douter quant au lien avec seul variable l'estimant, son nombre d'abonnés.

-cont : dichotomique égale à 1 si le joueur est européen, 0 sinon

Valeur attendue : non significative, presque la moitié des attaquants de notre échantillon sont européens, mais valent-ils plus chers ? Nous pensons que ce ne sera pas le cas du fait de la popularité des joueurs étrangers dans leur pays respectifs, et donc l'image que cela génère pour le club.

Les valeurs des prix des joueurs ont été trouvés sur le site de transfermarkt.com, qui estime via son algorithme le prix des joueurs. Ces valeurs se rapprochent de ce que propose le marché mais reste une estimation. Elle se rapproche nettement de l'intervalle de prix que propose le CIES, organisme qui propose parmi les meilleures estimations des prix des joueurs de football. Selon eux, la corrélation entre sommes payées par les clubs et estimées est supérieure à 80%, ce qui signifie que les variables prises en compte expliquent plus de quatre cinquièmes des différences de prix entre les transferts inclus dans leur modèle. Nous n'avons pas la prétention ni les données pour arriver à un modèle aussi pointu, mais nous allons essayer d'arriver à faire un modèle cohérent et représentatif des grands axes.

L'évaluation et tests statistiques:

A l'aide de ce modèle, nous allons procéder à de nombreux tests statistiques pour résoudre les problèmes, s'il y en a, d'hétéroscédasticité, de normalité des erreurs et de multi colinéarité. Nous ne prenons en compte aucune variable de temps donc nous n'aurons pas besoin d'effectuer des tests sur les séries temporelles et l'autocorrélation des erreurs.

Hétéroscédasticité :

Dans un premier temps, nous avons donc procédé à une régression du prix par rapport à toutes ces variables explicatives.

Nous

```
. reg lnprix equipnat10 taille classeq20 piedfort ratio21 champfr champal champan champes champit ratiobm3ans age lnabo  
> insta fincontrat cont
```

Source	SS	df	MS	Number of obs	=	83
				F(15, 67)	=	12.80
Model	47.9114567	15	3.19409711	Prob > F	=	0.0000
Residual	16.7161424	67	.249494663	R-squared	=	0.7413
				Adj R-squared	=	0.6834
Total	64.6275991	82	.788141453	Root MSE	=	.49949

lnprix	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
equipnat10	.4306632	.158824	2.71	0.008	.1136492 .7476772
taille	-.0226846	.0097944	-2.32	0.024	-.0422343 -.003135
classeq20	-.0181652	.022167	-0.82	0.415	-.0624108 .0260804
piedfort	-.2286599	.1339953	-1.71	0.093	-.4961156 .0387959
ratio21	.7433704	.5318261	1.40	0.167	-.318159 1.8049
champfr	.4657691	.2379214	1.96	0.054	-.0091239 .9406622
champal	.7009627	.2252675	3.11	0.003	.2513268 1.150599
champan	.878787	.2414374	3.64	0.001	.3968759 1.360698
champes	.3054134	.2342814	1.30	0.197	-.1622142 .7730409
champit	.75808	.22732	3.33	0.001	.3043473 1.211813
ratiobm3ans	.8571802	.6902249	1.24	0.219	-.5205144 2.234875
age	-.1003367	.0167691	-5.98	0.000	-.133808 -.0668653
lnaboinsta	.1195773	.0373185	3.20	0.002	.0450892 .1940653
fincontrat	.1060958	.0601225	1.76	0.082	-.0139093 .2261009
cont	.0983643	.1351097	0.73	0.469	-.1713158 .3680445
_cons	20.56144	1.82318	11.28	0.000	16.92235 24.20052

proposons un modèle où les variations des prix et des autres variables ont une grande différence due aux chiffres importants (en millions pour les prix par exemple). Nous procédons alors de tester l'homoscédasticité en premier lieu, car il y a de fortes chances que nous trouvions un problème d'hétéroscédasticité. Nous effectuons alors le test de Breusch-Pagan. Logiquement, nous avons un problème, la p-value est inférieure à 5% et nous rejetons donc l'hypothèse nulle d'homoscédasticité.

```
. estat hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
Ho: Constant variance
```

```
Variables: fitted values of prix
```

```
chi2(1) = 19.22
```

```
Prob > chi2 = 0.0000
```

Pour résoudre ce problème, nous pensons qu'un changement de spécification est nécessaire avant d'appliquer des corrections comme les écarts types robuste de White. Les variables prix et aboinsta sont très grandes par rapport aux autres. Nous créons ainsi les variables lnprix et lnaboinsta qui sont les logarithmes des variables prix et aboinsta. Nous posons ainsi une nouvelle régression avec cette nouvelle variable à expliquer et explicative puis effectuons de nouveaux les tests statistiques d'hétéroscédasticité de Breush-Pagan et de White.

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of lnprix

      chi2(1)      =      0.34
      Prob > chi2   =      0.5598

. estat imtest, white

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

      chi2(83)      =      84.00
      Prob > chi2   =      0.4487
```

Comme nous pouvons le voir, la p-value associé au test de Breush-Pagan est bien supérieure à 5%, comme pour le test de White, donc le problème d'hétéroscédasticité est résolu grâce à la nouvelle spécification du modèle.

Normalité des Erreurs :

Tout d'abord, nous allons enregistrer la variable « résidu » contenant les résidus du modèle. Nous pouvons observer le graphe des résidus qui, de notre point de vu nous montre qu'il n'y aura vraisemblablement pas de problème de normalité des erreurs.

Nous validons notre pensée avec le test de Jarque et Berra, où l'on note la p-value inférieur à 5% et donc nous ne remettons pas en cause l'hypothèse nulle de normalité des erreurs.

```
. predict residu, res

. jb residu
Jarque-Bera normality test:  4.584 Chi(2)  .1011
Jarque-Bera test for Ho: normality:

.
```

Multicolinéarité :

C'est ici que le modèle nous à poser le plus de problème. La multicolinéarité partielle est par définition à la forte corrélation des variables explicatives. Cela a pour conséquence de fortes variances des paramètres estimées. Nous allons ainsi étudier en premier lieu la corrélation entre les variables explicatives :

```
. correlate lnprix equipnat10 taille ratio21 ldc classeq20 piedfort champfr champal champan champes champit ratiobm3ans
> age cont fincontrat lnaboinsta
(obs=83)
```

	lnprix	equip~10	taille	ratio21	ldc	class~20	piedfort	champfr	champal	champan	champes	champit	ratiobm3ans
lnprix	1.0000												
equipnat10	0.3910	1.0000											
taille	-0.1926	-0.1554	1.0000										
ratio21	0.1928	0.0881	0.2940	1.0000									
ldc	0.4486	0.2885	-0.2737	0.0262	1.0000								
classeq20	-0.2206	-0.1492	0.0523	-0.0489	-0.5078	1.0000							
piedfort	-0.1618	-0.0687	0.0223	-0.1118	-0.0761	-0.0083	1.0000						
champfr	0.0480	-0.1448	-0.1899	-0.0804	0.0280	0.1038	0.0983	1.0000					
champal	0.1030	-0.1689	0.1721	0.2272	-0.0050	-0.0184	0.0817	-0.1772	1.0000				
champan	0.3279	0.2281	-0.1941	-0.2584	-0.1233	-0.0507	0.0373	-0.2268	-0.2163	1.0000			
champes	-0.1256	0.1653	-0.0872	-0.0310	0.2269	0.1038	-0.2067	-0.1857	-0.1772	-0.2268	1.0000		
champit	0.0619	0.0878	0.1184	0.1159	0.0943	0.1038	0.0220	-0.1857	-0.1772	-0.2268	-0.1857	1.0000	
ratiobm3ans	0.2497	0.2966	0.1785	0.7645	0.1426	-0.1898	-0.0071	-0.0371	0.1083	-0.0361	-0.1091	0.0099	1.0000
age	-0.3214	0.2238	-0.0337	0.1506	-0.0251	0.0222	0.0430	-0.2665	-0.0707	0.0099	0.1663	0.0099	0.1663
cont	0.0242	-0.1983	0.0100	0.0450	-0.0958	0.2152	-0.0943	-0.0663	0.2777	-0.1488	0.0766	0.0766	0.0766
fincontrat	0.2249	-0.0154	0.2095	-0.1328	-0.0227	-0.0293	-0.1316	0.1098	-0.0603	0.0245	-0.0056	-0.0056	-0.0056
lnaboinsta	0.4986	0.4290	-0.1895	0.1512	0.4392	-0.3619	-0.0190	-0.0966	0.0093	0.3090	0.0644	0.0644	0.0644

	champit	ratiob~s	age	cont	fincon~t	lnaboi~a
champit	1.0000					
ratiobm3ans	0.0974	1.0000				
age	0.1815	0.3513	1.0000			
cont	0.0052	-0.1258	-0.0226	1.0000		
fincontrat	-0.0632	-0.2727	-0.4638	0.1217	1.0000	
lnaboinsta	0.0890	0.3878	0.3350	-0.0365	-0.1830	1.0000

On retrouve une forte corrélation entre les variables ratiobm20 et ratiobm3ans notamment avec des corrélations supérieures à 0. Ces valeurs sont bien significatives, on le remarque avec la deuxième commande *pwcorr* ci-dessous.

```
. pwcorr lnprix equipnat10 taille ratio21 ldc classeq20 piedfort champfr champal champan champes champit ratiobm3ans ag
> e cont fincontrat lnaboinsta, star(0.5)
```

	lnprix	equip~10	taille	ratio21	ldc	class~20	piedfort
lnprix	1.0000						
equipnat10	0.4092*	1.0000					
taille	-0.2050*	-0.1707*	1.0000				
ratio21	0.1929*	0.0896*	0.2908*	1.0000			
ldc	0.4576*	0.3024*	-0.2815*	0.0278	1.0000		
classeq20	-0.2206*	-0.1492*	0.0523	-0.0489	-0.5078*	1.0000	
piedfort	-0.1278*	-0.0313	0.0036	-0.1068*	-0.0542	-0.0083	1.0000
champfr	0.0070	-0.1844*	-0.1590*	-0.0820*	-0.0000	0.1038*	0.0484
champal	0.1088*	-0.1571*	0.1667*	0.2277*	-0.0000	-0.0184	0.0884*
champan	0.3323*	0.2345*	-0.1985*	-0.2570*	-0.1161*	-0.0507	0.0471
champes	-0.1164*	0.1710*	-0.0913*	-0.0302	0.2304*	0.1038*	-0.1943*
champit	0.0685	0.0950*	0.1131*	0.1166*	0.0987*	0.1038*	0.0303
ratiobm3ans	0.2469*	0.2917*	0.1773*	0.7644*	0.1420*	-0.1898*	-0.0065
age	-0.2898*	0.2458*	-0.0481	0.1514*	-0.0081	0.0222	0.0695
cont	0.0489	-0.1619*	-0.0059	0.0470	-0.0765*	0.2152*	-0.0621
fincontrat	0.2088*	-0.0300	0.2156*	-0.1337*	-0.0312	-0.0293	-0.1434*
lnaboinsta	0.5159*	0.4543*	-0.2067*	0.1507*	0.4501*	-0.3619*	0.0257

	champfr	champal	champan	champes	champit	ratiob~s	age
champfr	1.0000						
champal	-0.1826*	1.0000					
champan	-0.2335*	-0.2132*	1.0000				
champes	-0.1914*	-0.1747*	-0.2235*	1.0000			
champit	-0.1914*	-0.1747*	-0.2235*	-0.1831*	1.0000		
ratiobm3ans	-0.0366	0.1083*	-0.0359	-0.1089*	0.0974*	1.0000	
age	-0.2923*	-0.0631	0.0184	0.1713*	0.1863*	0.3477*	1.0000
cont	-0.1026*	0.2810*	-0.1376*	0.0831*	0.0126	-0.1238*	0.0019
fincontrat	0.1254*	-0.0636	0.0199	-0.0092	-0.0667	-0.2721*	-0.4689*
lnaboinsta	-0.1493*	0.0197	0.3133*	0.0736	0.0975*	0.3772*	0.3574*

On applique aussi le test de facteur d'inflation de la variance (VIF) et remarque en effet que ces 2 variables portent à la présomption de multicollinéarité car elles approchent voire dépassent la valeur 4.

Variable	VIF	1/VIF
ratiobm3ans	4.07	0.245560
ratio2l	3.40	0.294551
champan	3.29	0.303630
champfr	2.49	0.402005
lnaboinsta	2.46	0.407184
champes	2.41	0.414593
champt	2.27	0.440375
champal	2.09	0.478964
age	1.78	0.560687
classeq20	1.71	0.586132
fincontrat	1.59	0.629540
equipnatl0	1.53	0.651533
taille	1.46	0.683048
cont	1.31	0.765452
piedfort	1.13	0.885827
Mean VIF	2.20	

C'est pourquoi cela nous amène à devoir re-spécifier notre modèle et enlever une des variables « ratio ». Notre choix se porte sur la variable ratiobm20, nous estimons qu'elle favorise moins l'estimation du prix d'un attaquant de football. Nous prenons en compte le fait qu'un joueur, s'il est blessé toute une saison, ne peut pas marquer de but et nous ne pouvons pas retirer cette variable du nombre de but par match très importante. Nous effectuons ainsi les mêmes tests de multicollinéarité sans cette variable et nous arrivons à des valeurs plus cohérentes. La moyenne des VIF est de 2 ce qui reste correct.

Variable	VIF	1/VIF
champan	3.13	0.319355
champes	2.68	0.372821
lnaboinsta	2.57	0.388868
champfr	2.46	0.405921
champt	2.43	0.410817
ldc	2.24	0.446544
classeq20	2.15	0.465711
champal	2.14	0.467873
age	1.80	0.556866
ratiobm3ans	1.68	0.596741
fincontrat	1.59	0.630430
taille	1.56	0.640715
equipnatl0	1.54	0.649616
cont	1.30	0.769133
piedfort	1.11	0.900828
Mean VIF	2.02	

Les valeurs ne sont pas idéales, notamment les valeurs de la dichotomie des championnats, mais le manque d'observations peut traduire quelques légers écarts.

Interprétations des résultats

Voici la régression finale de notre modèle de départ, après avoir vu que les tests d'homoscédasticité, de multicollinéarité, de normalité des erreurs soit validée.

$$\ln \text{Prix}_t = \beta_0 + \beta_1 \text{equipnat10} + \beta_3 \text{taille} + \beta_4 \text{ldc} + \beta_5 \text{classeq20} + \beta_6 \text{piedfort} + \beta_7 \text{champfr} + \beta_8 \text{champal} + \beta_9 \text{champan} + \beta_{10} \text{champes} + \beta_{11} \text{champit} + \beta_{12} \text{ratiobm3ans} + \beta_{13} \text{age} + \beta_{14} \text{fincontrat} + \beta_{15} \text{lnaboinsta} + \beta_{16} \text{cont}$$

où $t=1, \dots, 84$

```
. reg lnprix equipnat10 taille ldc classeq20 piedfort champfr champal champan champes champit ratiobm3ans age lnaboinsta
> a fincontrat cont
```

Source	SS	df	MS	Number of obs	=	83
Model	49.2760386	15	3.28506924	F(15, 67)	=	14.34
Residual	15.3515606	67	.229127769	Prob > F	=	0.0000
				R-squared	=	0.7625
				Adj R-squared	=	0.7093
Total	64.6275991	82	.788141453	Root MSE	=	.47867

lnprix	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
equipnat10	.3587084	.1524279	2.35	0.022	.0544611 .6629556
taille	-.0145473	.0096912	-1.50	0.138	-.033891 .0047965
ldc	.4471106	.157264	2.84	0.006	.1332104 .7610108
classeq20	.0182464	.0238317	0.77	0.447	-.0293219 .0658147
piedfort	-.2436643	.1273361	-1.91	0.060	-.4978282 .0104996
champfr	.320395	.226901	1.41	0.163	-.1325012 .7732912
champal	.6131696	.2184209	2.81	0.007	.1771996 1.04914
champan	.8423786	.2256046	3.73	0.000	.39207 1.292687
champes	.0876499	.2367594	0.37	0.712	-.3849239 .5602236
champit	.5813362	.2255449	2.58	0.012	.1311467 1.031526
ratiobm3ans	1.549064	.4243113	3.65	0.001	.7021349 2.395993
age	-.0950981	.0161252	-5.90	0.000	-.1272841 -.0629122
lnaboinsta	.0983059	.0365954	2.69	0.009	.0252611 .1713506
fincontrat	.1118432	.0575757	1.94	0.056	-.0030783 .2267647
cont	.1334968	.1291675	1.03	0.305	-.1243226 .3913161
_cons	19.0271	1.825024	10.43	0.000	15.38434 22.66986

Tout d'abord nous pouvons remarquer que la régression est significativement globale (Prob>F=0.000 <5%) et que le coefficient de détermination R^2 est égal à 0,7625, ce qui est plutôt satisfaisant. Le modèle explique donc environ 76% de la variance. Les résultats suivants seront interprétés « *toutes choses étant égales par ailleurs* » :

Les variables cont, piedfort, taille et classeq20 sont nettement non significative (p-value<5%), elles n'ont ainsi aucun impact sur la détermination du prix d'un attaquant de football. Cela rejoint les valeurs attendues.

La variable fincontrat est non significative à 5%, mais elle l'est à 10% et positivement. C'est une marge d'erreur notable mais interprétable. À 10% près, une augmentation d'un an du contrat du joueur augmenterait son prix d'en moyenne 11,18%. Cela ne s'explique pas le fait qu'un footballeur en fin de contrat peut partir libre dans un autre club, ce qui entraîne à la baisse le prix des joueurs qui s'en rapproche.

La variable equipnat10 est significative à 5%. Ainsi si le joueur à participer à plus de 10 rencontres internationales, son prix augmente en moyenne de 35,87%.

La variable ldc est aussi significative à 5%, cela veut dire qu'en moyenne un joueur qui participe à la ligue des champions voit sa valeur augmenter de 44,7%. Participer à des matchs internationaux, ou dans la compétition la plus prestigieuse d'Europe permet une visibilité mondiale du joueur et démontre qu'il fait partie des meilleurs attaquant donc augmente son prix.

Ici, nous avons choisis de comparer les joueurs évoluant dans les cinq grands championnats par rapport à la variable de référence champ12: les joueurs évoluant dans les championnats mineurs d'Europe (notamment la Belgique, la Turquie, le Portugal). On remarque que champfr et champes sont non significatives. Cela montre qu'il n'y a pas de différence de prix notable entre les attaquants du championnat de France, d'Espagne et des championnats mineurs européens. En revanche les joueurs évoluant en Angleterre Allemagne et Italie voient en moyenne leur prix augmenter de respectivement 84%, 51% et 58% par rapport aux joueurs des autres championnats européens. Il n'est pas étonnant de voir le championnat anglais devant, car le marché est particulièrement élevé et la concurrence est très rude outre-manche. Le championnat allemand est lui aussi un des meilleurs championnats au monde. En revanche les résultats sont plus étonnants pour l'Italie qui, malgré des clubs très puissants, a connu une baisse de popularité et de niveau ces dernières années. Le championnat français (et espagnol dans une moindre mesure) a un niveau moyen plus faible et donc financièrement moins forte qui fait qu'il se rapproche des autres championnats comme celui du Portugal, de la Belgique ou de la Turquie par exemple.

La variable ratiobut3ans est significative à 5%. Elle démontre que les statistiques individuelles sur les 3 dernières années jouent beaucoup en faveur du prix d'un joueur. En moyenne, l'augmentation du ratio but/match d'une unité augmente la valeur du joueur de 154,9%. Cette analyse pose un problème dans le sens où l'augmentation d'un point du ratio but/match peut varier énormément, nous avons ainsi procédé à une autre régression (les valeurs étaient approximativement les mêmes pour le reste des variables explicatives) pour avoir un aperçu de l'augmentation du prix du joueur pour un but marqué en plus lors des 3 derniers années : environ 1,7%. Nous l'avons supposé au départ, la valeur d'un attaquant se retrouve beaucoup dans ces statistiques, un grand buteur sera très recherché et verra son prix nettement augmenter.

L'âge est aussi significativement différent de 0 à 5%. En moyenne, si l'attaquant a un an en plus, sa valeur sur le marché des transferts baissera de 9,5%. Les jeunes joueurs sont très chers, leur côte a augmenté ces dernières années car il représente une source future de profit très certains par rapport à des buteurs en fin de carrière.

Enfin, lnaboinsta est significative elle aussi. Une augmentation d'1% du nombre d'abonnés sur Instagram augmente le prix du joueur d'en moyenne 11,18%. Cela montre que le prix d'un joueur est aussi dû à sa popularité et ces contrat publicitaire par exemple. Plus un joueur est connu et plus le club auquel il appartient est reconnu.

Il est à noter que lors des régressions avec la variable du ratiobm20 a la place de ratiobm3ans, elle était significative à 10% mais avait un impact bien moins important que le nombre de but par match sur les 3 dernières années.

Conclusion

Le travail effectué dans ce rapport reste très simpliste et l'estimation du prix d'un joueurs de football est beaucoup plus complexe et est liée à de nombreuses autres variables. Nous aurions pu (très difficile à obtenir) obtenir des informations sur la qualité purement sportive du joueur comme sa vitesse, sa force, sa qualité de passe, son nombre d'accélération par match etc... et des informations sur le côté extra sportif du joueur comme ses contrats de sponsoring, son contrat d'agent... Il aurait été aussi idéal de mettre en avant les variables d'inflation du marché et du monde du football en général, ce qui amènerait à des estimations plus sûres à travers le temps. Il ne faut pas oublier tout l'aspect psychologique qu'on peut retrouver dans le transfert d'un joueur, qui peut plus ou moins jouer sur la cote de celui-ci.

Cependant, notre dossier met en évidence les grandes variables du prix d'un attaquant de football, à travers le joueur avec les variables d'âges, de statut, de contrat, de popularité, de statistiques, à travers le club avec le classement de son équipe et son championnat, et enfin à travers le contexte économique.