



**UNIVERSITÉ DE  
MONTPELLIER**

## **Faculté d'économie**

**Les facteurs d'influence du prix du vin**

**CARENA Gianni - RAË Jolhan**

**M1 - MBFA - IGF - Actuariat - (2024-2025)**

## Résumé

L'étude a permis de mettre en évidence l'importance de l'impact qu'à la notation d'un vin par les consommateurs, la provenance géographique, l'année de production et la couleur sur son prix. L'ajout des variables indicatrices permettant de quantifier l'effet "Pays" a été pertinent. Cependant, cette étude aurait pu être plus poussée par l'utilisation d'une variable "Région" qui aurait rendu les modèles plus précis mais bien plus complexes. La variable "Notes" présente un biais lié au comportement humain, qui peut être influencé par le prix et la réputation de la bouteille avant même que celle-ci soit évaluée. La variable "Année" est également biaisée, car la définition d'un "bon" millésime varie selon les régions et les pays, un aspect qui n'est pas pris en compte dans cette étude.

## 1 Introduction

Analyser le prix du vin est essentiel pour comprendre les dynamiques économiques et culturelles qui influencent ce produit emblématique. Le vin, au-delà de sa simple valeur marchande, est un reflet de traditions, de terroirs et de savoir-faire qui varient d'un pays ou d'une région à l'autre. Comprendre les facteurs qui déterminent son prix permettent non seulement d'apprécier la complexité de ce marché, mais aussi d'aider les producteurs à mieux positionner leurs produits et les consommateurs à faire des choix éclairés.

- **Objectif de l'étude :** L'objectif principal de cette étude est d'analyser les facteurs qui influencent le prix du vin. En utilisant des données provenant de différentes sources, nous chercherons à déterminer l'importance relative de chaque variable explicative. Cette analyse permettra de fournir des informations précieuses pour les producteurs de vin, les distributeurs, et les consommateurs, en leur offrant une meilleure compréhension des éléments qui déterminent le prix du vin.

## 1.1 Définitions des données

L'ensemble des données qui vont être exploitées proviennent principalement de Kaggle, une plateforme qui fournit des jeux de données en libre accès créé par la communauté, qui vient elle-même de données fournies par Vivino, un des sites référence pour les amateurs de vins. Un processus de nettoyage et de normalisation a été effectué sur ces données afin de réduire les erreurs inhérentes aux données d'origine.

### 1.1.1 Variable Expliquée

La variable dépendante  $Y$  sera :

- **l.Price** : Il s'agit du prix Toutes Taxes Comprises, en Euro, du vin étudié exprimé en Logarithme.

### 1.1.2 Variables Explicatives

L'ensemble des variables explicatives  $X$  sont les suivantes :

- **Rating20** : Note sur 20, attribuée par les consommateurs sur la plateforme Vivino.
- **Year** : Année de production (Vins produit de 1988 à 2020).
- **Age** : Âge du vin (Multicolinéarité parfaite avec la variable Year, utilisée afin de simplifier l'interprétation des coefficients).
- **Country** : Pays d'origine (7 pays ont été retenus).
- **Color** : Couleur du vin (Rouge, Blanc, Rosé ou pétillant).

### 1.1.3 Variables Indicatrices :

- **Country** : Pays d'origine (7 pays ont été retenus) Les pays d'origine du vin influencent significativement ses caractéristiques et sa qualité. Chaque région viticole, comme la France, l'Italie ou l'Espagne, apportent des conditions climatiques et des traditions de vinification uniques. Ces facteurs se reflètent dans le goût, l'arôme et la structure des vins produits. Par exemple, les vins de Bordeaux sont réputés pour leur complexité, tandis que les vins de Toscane sont célèbres pour leur richesse.
- **Color** : Couleur du vin (Rouge, Blanc, Rosé ou pétillant) La couleur du vin est également un indicateur important de son potentiel de prix. Les vins rouges, élaborés à partir de raisins noirs, obtiennent leur couleur des pigments présents dans la peau des raisins. Les vins blancs, produits à partir de raisins blancs ou noirs sans la peau, varient du jaune pâle au doré. La couleur peut influencer la perception de la qualité et, par conséquent, le prix. Par exemple, un vin rouge intense et profond peut être perçu comme ayant un potentiel de vieillissement et de qualité, supérieur, justifiant ainsi un prix plus élevé.

## 1.2 Statistiques de Variables

### 1.2.1 Statistiques descriptives

Variable	Moyenne	Médiane	E.T.	Min	Max
Rating20	15.509	15.600	1.1873	10.000	19.600
Price	33.938	16.470	60.325	3.5500	995.00
Year	2015.4	2016.0	3.2282	1988.0	2020.0
Age	7.5507	7.0000	3.0000	3.0000	35.000

TABLE 1 – Statistiques descriptives, Observations 1-11212

Une asymétrie de distribution de la variable Price est à noter. Elle est due à de nombreux vins dont les prix se situent au-delà des 500 euros (ce qui tire la distribution vers le haut).

### 1.2.2 Matrice de corrélation :

Variables	Price	NbOfRatings	Year	Rating20
Price	1.000			
NbOfRatings	0.0372	1.000		
Year	-0.4405	-0.0824	1.000	
Rating20	0.5340	0.0866	-0.3658	1.000

TABLE 2 – Matrice des coefficients de corrélation

Le coefficient entre les variables Year et Price est négatif, ce qui s'explique par le fait que plus un vin est vieux, plus son prix va augmenter. (Plus l'année diminue, plus le prix augmente).

## 2 Modèle Économétrique

### 2.1 Modèle 1

La variable Rating20 est incluse en premier dans la régression pour expliquer la variation du prix du vin. Ce premier modèle n'inclut aucune autre variable afin de permettre l'appréciation de leur importance dans les prochains modèles.

$$\ln(\text{Price}) = \text{const} + \beta_1 \text{Rating20} + \epsilon \quad (1)$$

#### 2.1.1 Test d'hétéroscédasticité :

Le test d'hétéroscédasticité de White est réalisé avant l'analyse de la significativité des coefficients. Une statistique de test  $\text{TR}^2 = 921,048657$  et une probabilité critique p. critique =  $P(\text{Chi-deux}(2) > 921,048657) = 0,000000 < 5\%$  conduit au rejet de l'hypothèse nulle  $H_0$  : homoscélasticité, indiquant ainsi que le modèle est hétéroscédastique. L'hétéroscédasticité masque souvent d'autres problèmes de spécification, tels que des variables omises dans la régression ou la présence de points aberrants.

Il est donc nécessaire d'utiliser une estimation d'écart-type robuste à l'hétéroscédasticité pour effectuer les tests de significativité.

Variable dépendante : <b>l_Price</b>				
	Coefficient	éc. type robuste	z	P. critique
const	-5.88414	0.0887926	-66.27	0.000
Rating20	0.571719	0.00580397	98.5	0.000
SCR		4005.840		
R <sup>2</sup>		0.563225	R <sup>2</sup> ajusté	0.563186
F(1,11210)		14455.40	P.critique(F)	0.000

TABLE 3 – Modèle 1 : MCO, estimation robuste, 11212 observations

#### 2.1.2 Test de significativité :

Le modèle présente un  $R^2 = 0.563225$ , ce qui est très satisfaisant avec une seule variable explicative. Le modèle explique 56,32% de la variance de la variable expliquée. Il présente une statistique  $F_{c=14455,40} < F(3,11208)$  ce qui permet de déterminer que l'hypothèse nulle  $H_0 : \forall \beta_i = 0$  est rejeté au risque de première espèce 5% [7]. Cela implique qu'il y a au moins un coefficient de régression significativement différent de 0, le modèle est donc significatif.

## 2.2 Modèle 2

En suivant la littérature empirique viticole, le millésime et l'âge d'un vin sont des indicateurs essentiels de sa qualité et de son caractère. Le millésime reflète les conditions climatiques de l'année de récolte, influençant la maturité des raisins et la concentration des arômes. Une année chaude peut produire des vins plus puissants, tandis qu'une année fraîche peut donner des vins plus acides. Ces facteurs sont cruciaux pour les amateurs et les collectionneurs, car ils déterminent le potentiel de "garde" du vin.

$$\ln(\text{Price}) = \text{const} + \beta_1 \text{Rating20} + \beta_2 \text{Age} + \epsilon \quad (2)$$

### 2.2.1 Test de normalité des résidus :

Le test de normalité des résidus implique un rejet de l'hypothèse nulle pour un Khi-deux(2)=411,794 avec p. critique 0,00000 (inférieur au seuil 5%). Ce test n'étant pas fondamental pour l'obtention d'un bon estimateur avec de nombreuses observations (nécessaire pour les petits échantillons), l'acceptation de l'hypothèse nulle n'est pas nécessaire pour obtenir des coefficients ainsi qu'un modèle significatif.

### 2.2.2 Test d'hétéroscédasticité :

Le test d'hétéroscédasticité de White est effectué avant l'analyse de significativité des coefficients. Une statistique de test  $TR^2 = 1134.347751$  et une probabilité critique p. critique =  $P(\text{Khi-deux}(5) > 1134,347751) = 0,000000 < 5\%$  implique le rejet de l'hypothèse nulle  $H_0$  : homoscedastique, le modèle est donc hétéroscédastique. Le test de Breusch-Pagan [6] vient confirmer ce résultat avec un LM = 682,112381 et p. critique =  $P(\text{Khi-deux}(2) > 682,112381) = 0,000000$  (Rejet évident de  $H_0$ ).

Dans ce modèle, une non-linéarité dans la forme fonctionnelle du modèle peut expliquer la présence d'hétéroscédasticité (figure 2) où l'on observe une répartition non-linéaire des prix par rapport aux années.

Il convient donc de passer par l'estimation d'écart-type robuste à l'hétéro-scédasticité pour pouvoir ensuite effectuer les tests de significativité.

Variable dépendante : <b>l_Price</b>				
	Coefficient	éc. type robuste	z	P. critique
const	-5.10748	0.0806593	-63.32	0.000
Rating20	0.473637	0.00562121	84.26	0.000
Age	0.0985992	0.00196030	50.30	0.000
SCR		3022.011		
R <sup>2</sup>		0.670496	R <sup>2</sup> ajusté	0.670438
F(2,11209)		8209.423	P.critique(F)	0.000

TABLE 4 – Modèle 2 : MCO, estimation robuste, 11212 observations

### 2.2.3 Test de significativité :

Le modèle présente un  $R^2=0.670496$ , le coefficient de détermination a considérablement augmenté avec l'ajout de seulement une variable explicative (Age) par rapport au nombre d'observations. Il présente aussi une statistique  $F_c = 11404.42$ , qui est supérieure à  $F(2,11209)=8209.423$ . Ce qui permet de rejeter l'hypothèse nulle :  $\forall \beta_i = 0$ , au niveau de signification de 5 %. Cela implique que les coefficients sont conjointement différents de zéro, rendant le modèle significatif.

De plus les probabilités critiques des 2 variables explicatives étant  $<0,05$ , cela démontre la significativité des coefficients car les hypothèses  $H_0 : \beta_1 = 0$  et  $H_0 : \beta_2 = 0$  (Coefficient non significatif) sont toutes deux rejetées.

## 2.3 Modèle 3 : Ajout des variables indicatrices :

La base de données de données initiale comprenait un grand nombre de pays et de régions, nécessitant un nettoyage pour simplifier l'utilisation des variables indicatrices. Elle a été réduite de 32 pays et 836 régions à 7 pays et 86 régions. La variable indicatrice "Région" n'est pas utilisée dans ce modèle pour éviter une régression trop complexe, mais la variable "Pays" est incluse.

Les pays inclus sont :

- DCountry\_Fr (France)
- DCountry\_It (Italie)
- DCountry\_Sp (Espagne)
- DCountry\_Ch (Chili)
- DCountry\_Ge (Allemagne)
- DCountry\_SA (Afrique du Sud)
- DCountry\_US (Etats-Unis)

La variable indicatrice pour l'Espagne a été supprimée du modèle afin d'éviter la multicolinéarité des paramètres. La suppression de ce pays en particulier a pour but de simplifier l'interprétation, permet de rendre les autres coefficients positifs. L'Espagne devient donc la variable de référence des pays.

Pour la couleur du vin, la base de données comprend une majorité de vins rouges, représentant 66 % des observations. Les vins blancs sont représentés à 22 %, les vins rosés à 3 % et les vins pétillants à seulement 2 %. La variable indicatrice sera donc présentée sous la forme :

- D\_Red
- D\_White
- D\_Sparkling
- D\_Rose

Même chose que pour les indicatrices de pays, c'est cette fois l'indicatrice D\_Rose qui est supprimée du modèle afin de prévenir la multicolinéarité des paramètres (choix arbitraire).

$$\begin{aligned}
\ln(\text{Price}) = & \text{const} + \beta_1 \text{Rating20} + \beta_2 \text{Age} + \beta_3 \text{D\_Red} \\
& + \beta_4 \text{D\_White} + \beta_5 \text{D\_Sparkling} \\
& + \beta_6 \text{DCountry\_Fr} + \beta_7 \text{DCountry\_It} + \beta_8 \text{DCountry\_Ch} \\
& + \beta_9 \text{DCountry\_Ge} + \beta_{10} \text{DCountry\_SA} + \beta_{11} \text{DCountry\_US} + \epsilon
\end{aligned} \quad (3)$$

### 2.3.1 Test de normalité des résidus :

Les résidus ne suivent pas la loi normale dans ce modèle puisque l'hypothèse nulle est rejetée (p. critique = 0.00000). Ce test nécessairement positif en petits échantillons pour l'obtention d'un bon modèle n'est pas le cas ici (environ 11000 observations).

### 2.3.2 Test d'hétéroscédasticité :

Suite au test de White, une perte de significativité pour la totalité des indicatrices sur les pays est à noter. Le  $R^2$  non ajusté diminue de 0.699850 à 0.127320. Sa statistique étant  $TR^2 = 1427,514053$ , avec une probabilité critique  $P(\text{Khi-deux}(49) > 1427,514053) = 0,000000$ , l'hypothèse nulle de la présence d'homoscédasticité est donc rejetée. Le test de White révèle l'existence d'hétéroscédasticité dans les résidus, résultat confirmé par le test de Breusch-Pagan : Statistique du test LM = 773,660452 pour une probabilité critique  $P(\text{Khi-deux}(11) > 773,660452) = 0,000$ , avec la même conclusion que le test de White. La présence d'hétéroscédasticité nécessite donc l'estimation des écart-types robustes qui permettent de confirmer la significativité des modèles et donc l'interprétation des coefficients.

Variable dépendante : <b>l_Price</b>				
	Coefficient	éc. type robuste	z	P. critique
const	-5.29613	0.0816201	-64.29	0.000
Rating20	0.465848	0.00547761	85.05	0.000
Age	0.0884162	0.00213902	41.33	0.000
D_Red	0.196297	0.0259190	7.573	0.000
D_White	0.160486	0.0258961	6.197	0.000
D_Sparkling	0.173555	0.0364236	4.765	0.000
DCountry_Fr	0.413965	0.0160834	25.74	0.000
DCountry_It	0.108629	0.0146342	7.423	0.000
DCountry_Ch	0.174955	0.0229420	7.626	0.000
DCountry_Ge	0.174257	0.0187363	9.301	0.000
DCountry_SA	0.131423	0.0190672	6.893	0.000
DCountry_US	0.451533	0.0242107	18.65	0.000
SCR		2752.802		
$R^2$		0.699850	$R^2$ ajusté	0.699555
F(11,11200)		1798.849	P.critique(F)	0.000

TABLE 5 – Modèle 5 : MCO, regression robuste, 11212 observations



### 2.3.3 Test de significativité :

Cette régression robuste permet de voir la précision de certains coefficients s'améliorer (les écarts types étant plus petits) comme les indicatrices sur les couleurs et les pays. La précision des autres variables a été réduite par l'estimation robuste avec des écarts-types qui ont augmenté.

Le  $R^2$  du modèle continu d'augmenter avec l'ajout de ces indicatrices, il passe de  $R^2 = 0.670496$  à  $R^2 = 0.699850$ . La probabilité critique  $p.\text{critique}(F) = 0.000$ , ce qui implique le rejet de l'hypothèse nulle  $H_0 : \forall \beta_i = 0$  au niveau de signification de 5 %. Cela implique que les coefficients sont conjointement différents de zéro, le modèle est donc significatif.

L'entièreté des variables du modèle 4 sont significatives. Il est possible de voir que tous les  $t$ . de Student associée à chaque variable sont supérieurs à 1.96 (ou que la  $p.\text{critique}$  associée est inférieur au seuil 5%). Ainsi, toutes les variables du modèle 4 sont pertinentes à l'explication de la variable explicative dans le cadre de l'échantillon de données.

## 2.4 Test de multicolinéarité :

Les tests sur les modèles et les régressions ont été réalisés à l'aide du logiciel Gretel. Aucun test spécifique de multicolinéarité n'a été effectué, car ce logiciel gère automatiquement cette problématique. En cas de multicolinéarité parfaite, une variable est omise lors des analyses. Cela a été observé dans le modèle 2, où une multicolinéarité parfaite existait entre les couleurs de vin et entre les pays. Le logiciel a ainsi désigné des variables de référence : D\_Rose pour la couleur et DCountry\_Sp pour les pays

## 3 Conclusion économétrique :

Tout comme le premier modèle, caractérisé uniquement par un  $R^2$  satisfaisant, les conditions nécessaires à l'établissement d'un BLUE ne sont pas remplies par ce nouveau modèle (hypothèse H4.a sur l'homoscédasticité non valide). Les ajustements effectués, tels que l'ajout de variables indicatrices permettant de mesurer l'effet de la couleur du vin ou de son origine offre une meilleure compréhension de son prix. Cela révèle une relation presque proportionnelle d'élasticité entre son âge (Age) et son prix. Une variation de un an supplémentaire de l'âge entraîne une augmentation de 8.84% du prix. Les relations d'élasticités du modèle ne se limitent pas à ce point, allant au-delà de l'élasticité estimée de l'âge par rapport au prix du vin (qui est égal à 0.0884), le modèle éclaire aussi sur l'élasticité estimée du prix en fonction de sa note qui est de 46.58. Ainsi, une variation de 1 unité dans la note du vin entraîne une variation de 46.58% du prix de ce dernier. Par ailleurs, le modèle informe aussi sur l'élasticité entre le prix et la provenance du vin concerné : une bouteille provenant de France aura un prix 51,28% plus élevé que s'il était espagnol ( $(\text{Exp}(0.413965)-1) \times 100$ ), de même pour un vin américain (57,07%), 11,47% pour l'Italie, 19.03% pour l'Allemagne. . . La couleur du vin influence aussi nettement son prix, un vin rouge devrait avoir un prix en moyenne 21.69% plus élevé que si c'était un rosé ( $(\text{Exp}(0.196297)-1) \times 100$ ), un blanc 17.41% et un champagne 18.95% plus élevé.

Un vin avec un prix élevé devrait donc être un vin rouge, Français ou Américain avec de très bonnes notes sur des sites comme Vivino ou Idealwine, tout en restant assez vieux. Dans la réalité, d'autres facteurs entrent en jeu.

## 4 Analyse littéraire :

L'évaluation du prix du vin est un processus multidimensionnel, influencée par une interaction complexe de facteurs intrinsèques et extrinsèques. Parmi ces facteurs, la réputation joue un rôle prépondérant. Dans le cas des Bordeaux, des classements historiques établis depuis 1855 restent un critère déterminant pour les prix, même face à des changements qualitatifs ou des méthodes de production modernes [1]. Cette réputation est souvent renforcée par les critiques d'experts et les classements qui façonnent la perception des consommateurs. Par ailleurs, les perceptions cognitives jouent un rôle majeur : des indices tels que le prix ou l'origine géographique influencent la perception sensorielle des consommateurs, qui peuvent associer un prix élevé et une réputation à une qualité supérieure, indépendamment de la réalité gustative (Spence, 2024). La rareté est un autre levier puissant, amplifiant la désirabilité et les prix des vins [3]. Les vins de la Romanée-Conti illustrent parfaitement ce phénomène. Ce vignoble de seulement 1,81 hectare en Bourgogne produit un nombre très limité de bouteilles, renforçant leur exclusivité et leur attrait. En 2018, une bouteille de Romanée-Conti 1945, réputée pour son millésime exceptionnel et liée à une histoire royale (domaine appartenant autrefois au cousin de Louis XV), s'est vendue 558 000 dollars lors d'une vente aux enchères [2]. Enfin, les facteurs climatiques jouent également un rôle dans les variations de prix. Une étude sur les vins de la Napa Valley révèle que les conditions météorologiques influencent davantage les prix que la qualité perçue, soulignant la sensibilité des marchés aux fluctuations externes [4]. En somme, le prix du vin résulte d'une interaction complexe entre réputation, rareté, qualité intrinsèque et perceptions, chacune jouant un rôle unique dans la construction de la valeur sur ce marché diversifié.

## 5 Conclusion

Le marché du vin, en raison de sa nature spéculative, est extrêmement complexe et influencé par de nombreux facteurs. Les prix des vins ne dépendent pas seulement de leur qualité intrinsèque, mais aussi de facteurs externes comme la réputation, la rareté, les conditions climatiques, les évaluations des experts et les perceptions des consommateurs. Pour comprendre les prix des bouteilles de vin, il est crucial de mesurer un large éventail de variables explicatives. Ces variables incluent :

- Caractéristiques sensorielles : Arômes, saveurs, structure et potentiel de vieillissement.
- Conditions de production : Méthodes de vinification, terroir et pratiques viticoles.
- Facteurs économiques : Offre et demande, tendances du marché et comportements des investisseurs.
- Influences climatiques : Variations annuelles du climat et leur impact sur les récoltes.
- Aspects culturels et historiques : Histoire du domaine, prestige et réputation.

Ce sont ces facteurs qui nous montrent que les variables explicatives de notre modèle possèdent des biais. Pour la variable "Année", il est bien connu que certains millésimes font tendre à la hausse le prix d'un vin, cependant, la qualité du millésime dépend lui-même de facteurs comme la météo sur l'année, la température moyenne, facteur maladie ayant atteint les grappes ou non... l'année 2010 peut être très bonne en France et très mauvaise aux Etats-Unis. Quant à la variable de notation (Rating20), elle explique "très bien" le prix du vin, ce qui peut être simplement un biais comportemental des consommateurs qui note un vin par rapport à son prix ou sa réputation. Un problème de subjectivité apparaît donc, mettant en avant un biais de perception. En intégrant ces différentes dimensions dans les modèles économiques, on peut mieux comprendre la complexité du marché du vin et fournir des estimations de prix plus précises et fiables. Cette approche globale est essentielle pour naviguer dans un marché où les perceptions et les attentes des consommateurs peuvent souvent surpasser les qualités intrinsèques du produit.

## 6 Références et Bibliographie

### 6.1 Sources de données

- Base de données sur Kaggle : Wine Rating and Price
- Nath Johnston

### 6.2 Compréhension des variables explicatives

- IdealWine

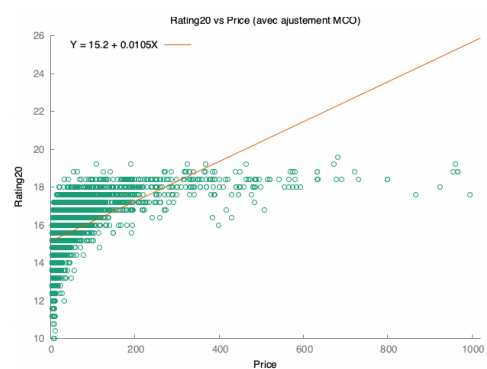
### 6.3 Logiciels statistiques utilisés

- **GRET**L : Logiciel libre pour l'analyse économétrique
- **STATA** : Logiciel avancé pour les analyses statistiques

### 6.4 Références bibliographiques

- 1 Ali, H. H., & Toulouse, G.R.E.M.A.Q. (2003). *Reputation and Quality Effects on Wine Prices : A Comparison Between En Primeur and Bottled Bordeaux Wine*.
- 2 Crum, G. (2018). *Le domaine de la Romanée-Conti*. Lannoo Meulenhoff-Belgium.
- 3 Cardebat, J. M., & Figuet, J. M. (2004). *What explains Bordeaux wine prices ? Applied Economics Letters*, 11(5), 293–296. <https://doi.org/10.1080/1350485042000221544>
- 4 Ramirez, C. D. (2008). *Wine Quality, Wine Prices, and the Weather : Is Napa “Different” ? Journal of Wine Economics*, 3(2), 114–131. <https://doi.org/10.1017/S1931436100001164>
- 5 H. White. (1980) *A Heteroskedasticity-Consistent covariance matrix estimator and a direct test for heteroskedasticity*. *Econometrica*, 48(4), 817. DOI : 10.2307/1912934. <https://doi.org/10.2307/1912934>.
- 6 Breusch, T. S., Pagan, A. R. (1979). *A simple test for heteroscedasticity and random coefficient variation*. *Econometrica*, 47(5), 1287-1294. <https://doi.org/10.2307/1911963>
- 7 Fisher, R. A. (1970). *Statistical methods for research workers* (11th ed. rev.). Oliver Boyd.
- 8 Gosset, W. S. (1908). *The probable error of a mean*. *Biometrika*, 6(1), 1-25.

## 7 Annexe :



(a) Rating20/Price



(b) Year/Price

FIGURE 1 – Graphiques de dispersion

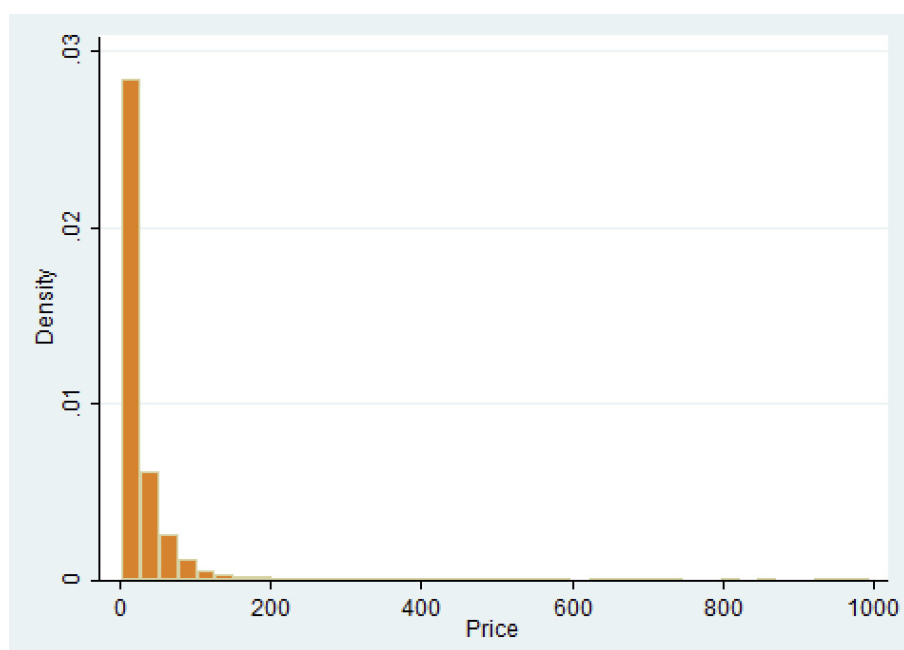


FIGURE 2 – Distribution de PRICE

## 7.0.1 Modèle 1 :

Variable dépendante : <b>l_Price</b>				
	Coefficient	éc. type	t de Student	P. critique
const	-5.88414	0.0739647	-79.55	0.000
Rating20	0.571719	0.00475518	120.2	0.000
SCR		4005.840		
R <sup>2</sup>		0.563225	R <sup>2</sup> ajusté	0.563186
F(1,11210)		14455.40	P.critique(F)	0.000

TABLE 6 – Modèle 1 : MCO, 11212 observations

Test de White pour l'hétéroscédasticité  
MCO utilisant les observations 1–11212

	Coefficient	éc. type	t de Student	P. critique
const	13.0616	0.600308	21.76	0.000
Rating20	-1.73945	0.0775021	-22.44	0.000
sq_Rating20	0.0589937	0.00249466	23.65	0.000
R <sup>2</sup> non ajusté		0.082148		

TABLE 7 – Modèle 1 : Test de White, 11212 observations

Statistique du test :  $TR^2 = 921.048657$ ,  
avec  $p.critique = P(Khi\text{-deux}(2) > 921.048657) = 0.000000$   
Hypothèse  $H_0$  : présence d'homoscédasticité rejetée. Le modèle présente de l'hétéroscédasticité.

Test de normalité des résidus :

Distribution des fréquences pour residual, obs = 1-11212

nombre de classes = 29, moyenne = 4.83334e-14, éc type = 0.597784

Intervalle	Centre	Fréquence	Rel.	Cum.
<-1,9548	-2,0399	3	0,03%	0,03%
-1,9548 – -1,7846	-1,8697	6	0,05%	0,08%
-1,7846 – -1,6145	-1,6996	12	0,11%	0,19%
-1,6145 – -1,4443	-1,5294	38	0,34%	0,53%
-1,4443 – -1,2742	-1,3592	85	0,76%	1,28%
-1,2742 – -1,1040	-1,1891	145	1,29%	2,58%
-1,1040 – -0,93383	-1,0189	273	2,43%	5,01%
-0,93383 – -0,76367	-0,84875	443	3,95%	8,96%
-0,76367 – -0,59351	-0,67859	680	6,06%	15,03%
-0,59351 – -0,42335	-0,50843	918	8,19%	23,22%
-0,42335 – -0,25319	-0,33827	1216	10,85%	34,06%
-0,25319 – -0,083028	-0,16811	1374	12,25%	46,32%
-0,083028 – 0,087133	0,0020528	1335	11,91%	58,22%
0,087133 – 0,25729	0,17221	1227	10,94%	69,17%
0,25729 – 0,42746	0,34238	1001	8,93%	78,09%
0,42746 – 0,59762	0,51254	739	6,59%	84,69%
0,59762 – 0,76778	0,68270	559	4,99%	89,67%
0,76778 – 0,93794	0,85286	441	3,93%	93,61%
0,93794 – 1,1081	1,0230	280	2,50%	96,10%
1,1081 – 1,2783	1,1932	170	1,52%	97,62%
1,2783 – 1,4484	1,3633	98	0,87%	98,49%
1,4484 – 1,6186	1,5335	69	0,62%	99,11%
1,6186 – 1,7887	1,7037	46	0,41%	99,52%
1,7887 – 1,9589	1,8738	21	0,19%	99,71%
1,9589 – 2,1291	2,0440	16	0,14%	99,85%
2,1291 – 2,2992	2,2141	7	0,06%	99,91%
2,2992 – 2,4694	2,3843	4	0,04%	99,95%
2,4694 – 2,6396	2,5545	4	0,04%	99,98%
≥ 2,6396	2,7246	2	0,02%	100,00%

TABLE 8 – Modèle 1 : Normalité des résidus, 11212 observations

Test de l'hypothèse nulle de normalité de la distribution :

Khi-deux(2) = 190.400 avec p.critique 0.00000

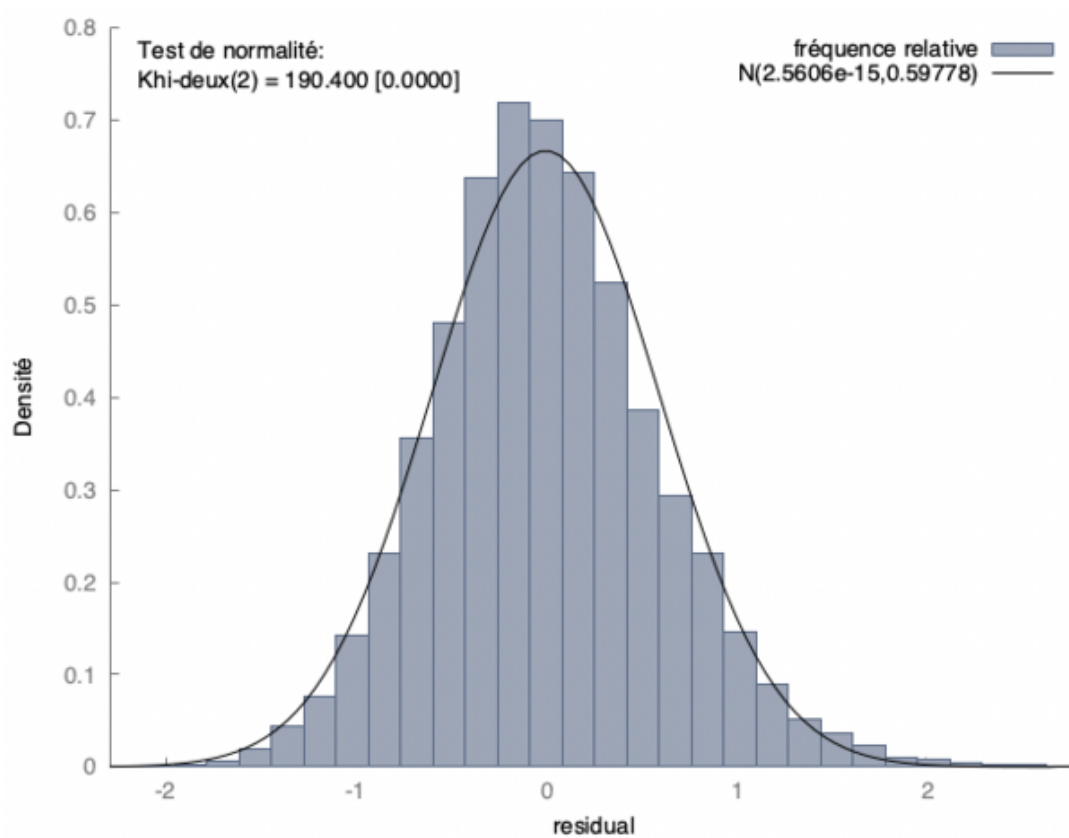


FIGURE 3 – Graphique distribution résidus, modèle 1



## 7.0.2 Modèle 2 :

Variable dépendante : <b>l_Price</b>				
	Coefficient	éc. type	t de Student	P. critique
const	-5.10748	0.0655197	-77.95	0.000
Rating20	0.473637	0.00443803	106.7	0.000
Age	0.0985992	0.00163222	60.41	0.000
SCR		3022.011		
R <sup>2</sup>		0.670496	R <sup>2</sup> ajusté	0.670438
F(2,11209)		11404.42	P.critique(F)	0.000

TABLE 9 – Modèle 2 : MCO, 11212 observations

Test de White pour l'hétéroscédasticité  
MCO utilisant les observations 1–11212

	Coefficient	éc. type	t de Student	P. critique
const	11.4922	0.500783	22.95	0.000
Rating20	-1.66976	0.0667346	-25.02	0.000
Age	0.261789	0.0191599	13.66	0.000
sq_Rating20	0.0613273	0.00226945	27.02	0.000
X2_X3	-0.0191966	0.00130543	-14.71	0.000
sq_Age	0.00199460	0.000215583	9.252	0.000
R <sup>2</sup> non ajusté		0.101173		

TABLE 10 – Modèle 2 : Test de White, 11212 observations

Statistique du test :  $TR^2 = 1134,347751$

avec  $p.critique = P(\text{Khi-deux}(5) > 1134.347751) = 0.000000$

Hypothèse  $H_0$  : présence d'homoscédasticité rejetée. Le modèle présente de l'hétéroscédasticité.

Test de normalité des résidus :

Distribution des fréquences pour residual, obs = 1-11212

nombre de classes = 29, moyenne = 4.46551e-14, éc. type = 0.519236

Intervalle	Centre	Fréquence	Rel.	Cum.
<-1.8216	-1.9072	2	0.02%	0.02%
-1.8216 – -1.6505	-1.7361	1	0.01%	0.03%
-1.6505 – -1.4793	-1.5649	7	0.06%	0.09%
-1.4793 – -1.3081	-1.3937	32	0.29%	0.37%
-1.3081 – -1.1369	-1.2225	60	0.54%	0.91%
-1.1369 – -0.9658	-1.0514	159	1.42%	2.33%
-0.9658 – -0.7946	-0.88018	322	2.87%	5.20%
-0.7946 – -0.6234	-0.70900	503	4.49%	9.69%
-0.6234 – -0.4522	-0.53783	928	8.28%	17.96%
-0.4522 – -0.2810	-0.36666	1245	11.10%	29.07%
-0.2810 – -0.1098	-0.19548	1542	13.75%	42.82%
-0.1098 – 0.061282	-0.02436	1672	14.91%	57.73%
0.061282 – 0.23246	0.14687	1440	12.84%	70.58%
0.23246 – 0.40363	0.31804	1161	10.35%	80.93%
0.40363 – 0.57481	0.48922	771	6.88%	87.81%
0.57481 – 0.74598	0.66059	502	4.48%	92.29%
0.74598 – 0.91716	0.83157	318	2.84%	95.12%
0.91716 – 1.0883	1.0027	205	1.83%	96.95%
1.0883 – 1.2595	1.1739	133	1.19%	98.14%
1.2595 – 1.4307	1.3451	50	0.45%	98.59%
1.4307 – 1.6019	1.5163	37	0.33%	98.92%
1.6019 – 1.7730	1.6874	20	0.18%	99.10%
1.7730 – 1.9442	1.8586	21	0.19%	99.29%
1.9442 – 2.1154	2.0290	9	0.08%	99.38%
2.1154 – 2.2866	2.2008	3	0.03%	99.41%
2.2866 – 2.4577	2.3721	2	0.02%	99.43%
2.4577 – 2.6289	2.5433	0	0.00%	99.43%
2.6289 – 2.8001	2.7145	0	0.00%	99.43%
≥ 2.8001	2.8857	1	0.01%	100.00%

TABLE 11 – Modèle 2 : Normalité des résidus, 11212 observations

Test de l'hypothèse nulle de normalité de la distribution :

Khi-deux(2) = 411.794 avec p. critique 0.00000

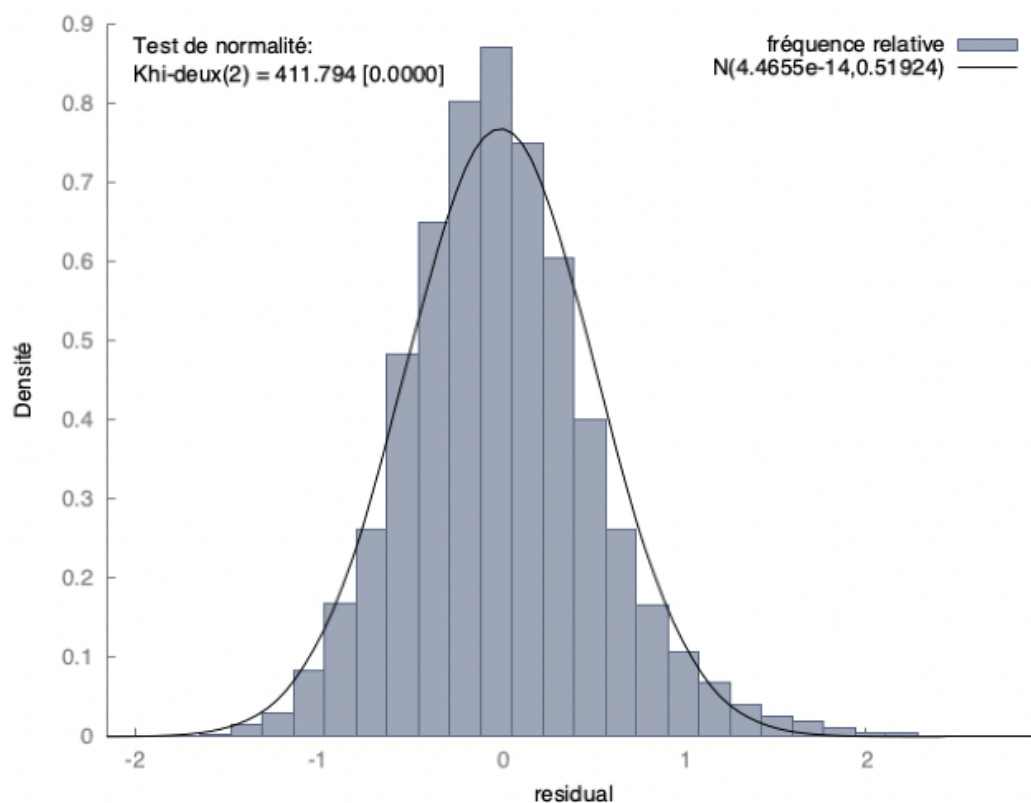


FIGURE 4 – Graphique distribution résidus, modèle 2

### 7.1 Modèle 3 :

Variable dépendante : <b>l_Price</b>				
	Coefficient	éc. type	t de Student	P. critique
const	-5,29613	0,0673394	-78,65	0,000
Rating20	0,465848	0,00428827	108,6	0,000
Age	0,0884162	0,00171512	51,55	0,000
D_Red	0,196297	0,0275292	7,130	0,000
D_White	0,160486	0,0281131	5,709	0,000
D_Sparkling	0,173555	0,0415691	4,175	0,000
DCountry_Fr	0,413965	0,0158123	26,18	0,000
DCountry_It	0,108629	0,0153841	7,061	0,000
DCountry_Ch	0,174955	0,0272687	6,416	0,000
DCountry_Ge	0,174257	0,0205258	8,490	0,000
DCountry_SA	0,131423	0,0216271	6,077	0,000
DCountry_US	0,451533	0,0252481	17,88	0,000
SCR		2752,802		
$R^2$		0,699850	$R^2$ Ajusté	0,699555
F(11,11200)		2374,057	P.critique(F)	0,000

TABLE 12 – Modèle 3 : MCO, 11212 observations

Test de White pour l'hétéroscédasticité  
MCO utilisant les observations 1–11212

	Coefficient	éc. type	t de Student	P. critique
const	11.9972	0.575778	20.84	1.26e-094
Rating20	-1.67378	0.0654109	-25.59	2.17e-140
Age	0.253509	0.0284583	8.908	6.00e-019
D_Red	-0.731855	0.361290	-2.026	0.0428
D_White	-0.363323	0.369808	-0.9825	0.3259
D_Sparkling	1.01214	0.586195	1.727	0.0843
DCountry_Fr	-0.434604	0.176334	-2.465	0.0137
DCountry_It	-0.217270	0.170534	-1.274	0.2027
DCountry_Ch	-0.322764	0.320427	-1.007	0.3138
DCountry_Ge	0.224006	0.260916	0.8585	0.3906
DCountry_SA	0.923493	0.237280	3.892	0.0001
DCountry_US	0.190170	0.267635	0.7106	0.4774
sq_Rating20	0.0587365	0.00211474	27.77	3.11e-164
X2_X3	-0.0171852	0.00133235	-12.90	8.56e-038
X2_X4	0.0550616	0.0227827	2.417	0.0157
X2_X5	0.0285259	0.0233883	1.220	0.2226
X2_X6	-0.0580907	0.0388115	-1.497	0.1345
X2_X7	0.0393979	0.0114381	3.444	0.0006
X2_X8	0.0192769	0.0111687	1.726	0.0844
X2_X9	0.0165104	0.0189621	0.8707	0.3839
X2_X10	-0.0219035	0.0161510	-1.356	0.1751
X2_X11	-0.0748038	0.0158906	-4.707	2.54e-06
X2_X12	-0.0186146	0.0172793	-1.077	0.2814
sq_Age	0.00282599	0.000229713	12.30	1.47e-034
X3_X4	-0.0296282	0.0216978	-1.365	0.1721
X3_X5	-0.0202908	0.0219641	-0.9238	0.3556
X3_X6	-0.0177973	0.0231702	-0.7681	0.4424
X3_X7	-0.0333719	0.00431207	-7.739	1.09e-014
X3_X8	-0.0179932	0.00490005	-3.672	0.0002
X3_X9	-0.00897863	0.0104209	-0.8616	0.3889
X3_X10	-0.00111519	0.00829924	-0.1344	0.8931
X3_X11	0.0272947	0.00922050	2.960	0.0031
X3_X12	-0.0154533	0.0119818	-1.290	0.1972
X4_X7	0.221104	0.0689675	3.206	0.0014
X4_X8	0.100510	0.0685866	1.465	0.1428
X4_X9	0.107035	0.189905	0.5636	0.5730
X4_X10	0.135677	0.121765	1.114	0.2652
X4_X11	0.0194696	0.0965560	0.2016	0.8402
X4_X12	0.233176	0.125773	1.854	0.0638
X5_X7	0.0693667	0.0722783	0.9597	0.3372
X5_X8	0.0167661	0.0714339	0.2347	0.8144
X5_X9	0.0712244	0.193116	0.3688	0.7123
X5_X10	0.112658	0.120049	0.9384	0.3480
X5_X11	0.0271717	0.0975784	0.2785	0.7807
X5_X12	0.137375	0.127290	1.079	0.2805
X6_X7	-0.0370266	0.129500	-0.2859	0.7749
X6_X8	-0.00618540	0.120698	-0.05125	0.9591
X6_X10	-0.0201042	0.170697	-0.1178	0.9062
X6_X11	-0.191643	0.182003	-1.053	0.2924
X6_X12	0.159059	0.293391	0.5421	0.5877
R <sup>2</sup> non ajusté		0.127320		

TABLE 13 – Modèle 3 : Test de White, 11212 observations

Statistique du test :  $TR^2 = 1427,514053$

avec  $p.critique = P(Khi\text{-deux}(49) > 1427.514053) = 0.000000$

Hypothèse  $H_0$  : présence d'homoscédasticité rejetée. Le modèle présente de l'hétéroscédasticité.

Distribution des fréquences pour residual, obs = 1–11212  
 nombre de classes = 29, moyenne = -4.73328e-015, éc. type = 0.495768

Intervalle	Centre	Fréquence	Rel.	Cum.
<-1.7656	-1.8508	1	0.01%	0.01%
-1.7656 – -1.5952	-1.6804	4	0.04%	0.04%
-1.5952 – -1.4247	-1.5099	7	0.06%	0.11%
-1.4247 – -1.2543	-1.3395	32	0.29%	0.39%
-1.2543 – -1.0839	-1.1691	71	0.63%	1.03%
-1.0839 – -0.91343	-0.99864	179	1.60%	2.62%
-0.91343 – -0.74299	-0.82821	354	3.16%	5.78%
-0.74299 – -0.57256	-0.65777	594	5.30%	11.08%
-0.57256 – -0.40212	-0.48734	954	8.51%	19.59%
-0.40212 – -0.23169	-0.31691	1340	11.95%	31.54%
-0.23169 – -0.061255	-0.14647	1650	14.72%	46.25%
-0.061255 – 0.10918	0.023962	1728	15.41%	61.67%
0.10918 – 0.27961	0.19440	1422	12.68%	74.35%
0.27961 – 0.45005	0.36483	1067	9.52%	83.87%
0.45005 – 0.62048	0.53526	731	6.52%	90.39%
0.62048 – 0.79092	0.70570	419	3.74%	94.12%
0.79092 – 0.96135	0.87613	269	2.40%	96.52%
0.96135 – 1.1318	1.0466	144	1.28%	97.81%
1.1318 – 1.3022	1.2170	95	0.85%	98.65%
1.3022 – 1.4727	1.3874	62	0.55%	99.21%
1.4727 – 1.6431	1.5579	49	0.44%	99.64%
1.6431 – 1.8135	1.7283	11	0.10%	99.74%
1.8135 – 1.9840	1.8987	9	0.08%	99.82%
1.9840 – 2.1544	2.0692	11	0.10%	99.92%
2.1544 – 2.3248	2.2396	5	0.04%	99.96%
2.3248 – 2.4953	2.4100	2	0.02%	99.98%
2.4953 – 2.6657	2.5805	1	0.01%	99.99%
2.6657 – 2.8361	2.7509	0	0.00%	99.99%
≥ 2.8361	2.9213	1	0.01%	100.00%

TABLE 14 – Modèle 3 : Normalité des résidus, 11212 observations

Test de l'hypothèse nulle de normalité de la distribution :  
 Khi-deux(2) = 407.270 avec p. critique 0.00000

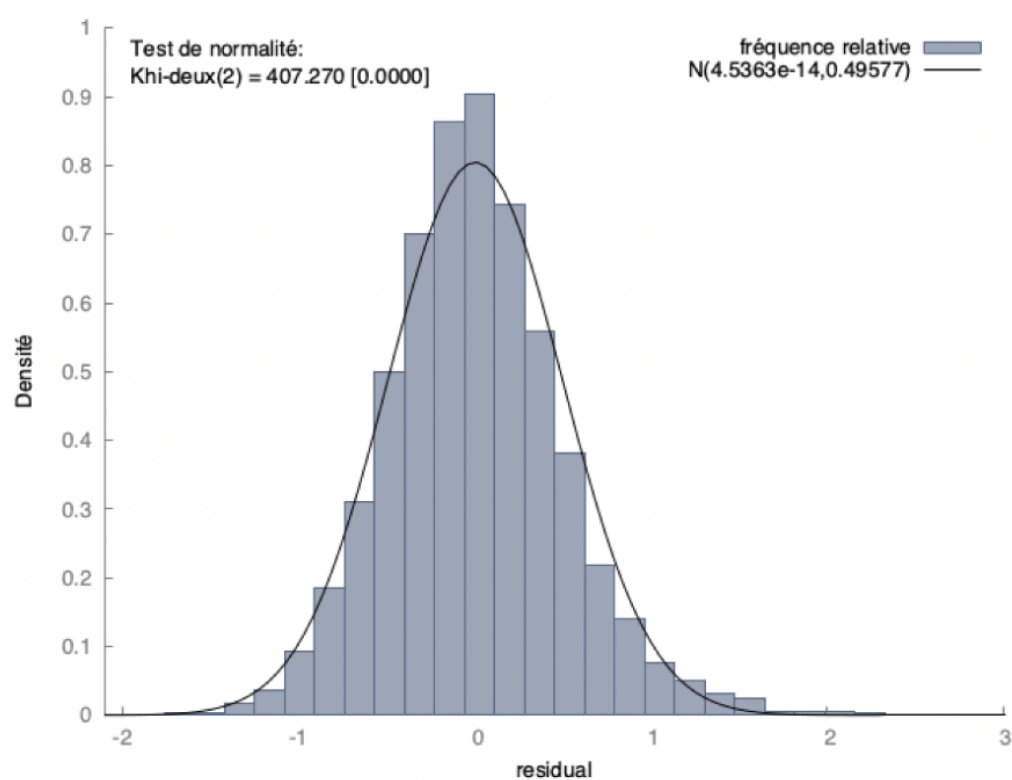


FIGURE 5 – Graphique distribution résidus, modèle 3