

MASTER ÉCONOMISTE D'ENTREPRISE



RECHERCHE, RÉALISATION, RESTITUTION

Application des modèles SFA à l'étude des prix


Corentin DUCLOUX et Aybuké BICAT

14 décembre 2023

| Table des matières

Remerciements	2
Introduction	3
Un prix	3
Partie 2 de l'intro	3
Revue de la littérature	4
Une nouvelle approche de la théorie du consommateur	4
Pricing Hédonique	5
Aspects théoriques	5
Application	6
Fonction de production	8
Le modèle SFA	8
Aspects théoriques	8
Utilisation empirique	11
SFA & Pricing Hédonique	11
Choix et cadrage de la problématique	14
Le marché de la téléphonie mobile, en constante évolution	14
Smartphones et Pricing Hédonique	16
Effet de réputation	16
Acquisition des données	18
Scraping	18
Méthodologie	18
Statistiques descriptives	20
Analyse approfondie des prix	20
Mesures de tendance centrale	20
Mesures de dispersion	21
Prix moyen en fonction des caractéristiques	21
Modélisation économétrique	23
Définitions	24
Références	25

| Remerciements

Nous tenions à remercier chaleureusement Monsieur *Alain BOUSQUET* pour son accompagnement tout au long de ce projet **3R**, qui a toujours été ouvert à l'exploration de nouveaux sujets, à l'expérimentation, et nous a encouragé à creuser diverses pistes de réflexion. Ce sujet a été et sera pour nous l'occasion de mettre en pratique l'ensemble des connaissances acquises dans notre cursus universitaire (microéconomie, économétrie, statistiques, analyse de la concurrence, pricing, développement logiciel sous **R** et **python** ) sur une problématique éminemment appliquée.

* * *

Note : Ce **PDF** a été entièrement rédigé en utilisant **Quarto** ⁽¹⁾, combinant la puissance et la versatilité de R, Python, et \LaTeX . Une présentation interactive **reveal.js** du sujet est aussi disponible. ⁽²⁾

⁽¹⁾ **Quarto** : Système de publication technique et scientifique *open-source* \Rightarrow <https://quarto.org/>.

⁽²⁾ Retrouvez la présentation sur <https://corentinducloux.fr>.

| Introduction

Un prix

En tant que consommateur, nous nous retrouvons souvent face à une question infiniment plus complexe qu'elle n'en a l'air. En des termes simples, elle se traduit par : Pourquoi ce prix ? Pour quelle raison ce stylo, cette nouvelle télévision, ou ce smartphone coûte tant ? Est-ce une là une simple question de coût de production

Comment modéliser l'utilité ? par le service rendu ? par les caractéristiques ?

Partie 2 de l'intro

- un petit `test`

du latex $\frac{2}{x} = 5$

Un test comme ça qui est très long et qui fait 2,3 lignes assez longues on va tester le débordement tetsjeelfnezfejfhezozhhhfezuoueeeeeeeeeeeeee

`2 + 2`

[1] 4

| Revue de la littérature

Une nouvelle approche de la théorie du consommateur

En microéconomie, dans la théorie du consommateur *classique*, le choix du meilleur ensemble de consommation dépend des préférences d'un individu. Les préférences de cet individu sont classiquement représentées par la fonction d'utilité :

$$U(x) = U(x_1, x_2, \dots, x_n) \quad (1)$$

Avec x_1, x_2, \dots, x_n un vecteur de n biens. L'Équation 1 exprime donc la relation entre la quantité de biens consommés et le niveau d'utilité que ces biens procurent à un agent. Dès lors, dans ce cadre, la consommation de biens procure **directement** de l'utilité à l'agent. En pratique pourtant, il est difficile de concevoir comment l'achat d'un bien comme une lampe ou un stylo peut nous apporter de l'utilité en tant que consommateur.

Pour répondre à cette difficulté, Lancaster (1966), propose un nouveau cadre conceptuel théorique décrit par les hypothèses suivantes.

Hypothèses

1. Le bien en lui même ne procure pas d'utilité au consommateur \Rightarrow il possède des **caractéristiques** qui procurent de l'utilité.
2. Un bien est un ensemble (*bundle*) de caractéristiques – il possède le plus souvent de nombreuses caractéristiques.
3. Une combinaison de biens peut posséder des caractéristiques différentes comparé à des biens consommés séparément.

Illustrons ces points avec quelques exemples :

- Un ordinateur n'est pas acheté pour le simple plaisir de posséder un ordinateur. Il est acheté car il permet de naviguer sur Internet, écrire des cours, programmer, regarder une série, etc. C'est donc pour les **services qu'il nous rend**, ce qui est modélisé ici par les caractéristiques possédées du bien.
- Les biens possèdent généralement un grand nombre de caractéristiques. Prenons l'exemple d'une gourde : la couleur, la forme, les dimensions et la capacité isothermique sont autant de caractéristiques qui peuvent influencer sur la décision d'achat.
- En consommant du lait et du café séparément, les caractéristiques retirées du lait sont de la vitamine D et du calcium, tandis que pour le café les caractéristiques retirées sont de la caféine, une boisson chaude, un "*boost*" le matin. En revanche, consommer un café latte permettra d'obtenir une boisson plus douce, moins cafféiné, un goût différent. En bref, les caractéristiques retirées du mélange sont différentes.

Dans le modèle de Lancaster il existe une relation **linéaire** entre les prix des biens et leurs caractéristiques. Le prix total p d'un bien peut donc être considéré comme la somme des prix individuels associé à chaque caractéristique. Cela découle du fait que les attributs des biens étudiés peuvent être considérés comme des composantes distinctes et séparables.

Pricing Hédonique

Aspects théoriques

Rosen (1974) étend ce qui a été apporté par le cadre théorique de Lancaster (1966). La principale différence est qu'il s'intéresse à **l'équilibre de marché de biens différenciés**, (là où Lancaster s'intéresse uniquement à la demande) avec :

- un continuum de biens du côté de l'offre,
- un continuum de consommateurs hétérogènes du côté de la demande.

Dans ce modèle, la relation entre les prix des biens et leurs attributs peut-être **non-linéaire** et permet aussi de capter des effets d'interaction entre plusieurs variables. Au coût d'une modélisation plus complexe que dans le modèle de Lancaster (1966), les résultats gagnent en robustesse.

L'objet de la contribution de Rosen est d'étudier un bien différencié z décrit par le vecteur de ses n caractéristiques mesurables tel que :

$$z = (z_1, z_2, \dots, z_n) \quad (2)$$

Afin de comprendre pourquoi il est important d'étudier des biens différenciés dans ce cadre, regardons en détail le graphique suivant.

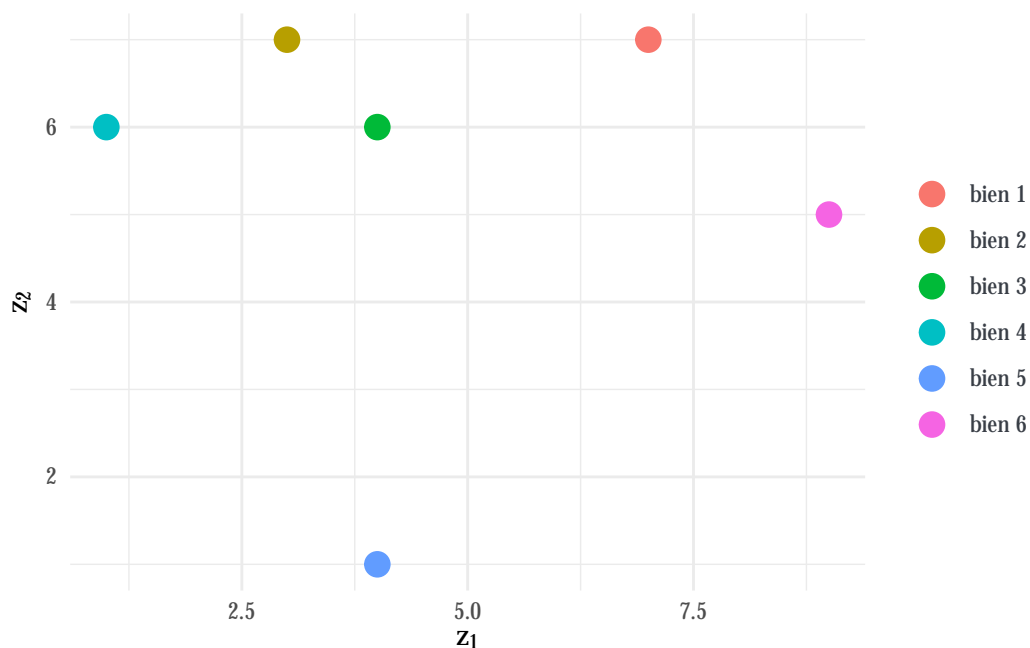


Figure 1 – Plan (z_1, z_2) de différents biens avec 2 caractéristiques.

En général, nous sommes habitués à représenter les préférences des consommateurs en termes de quantités de biens x_1, x_2 . Ici, on assiste à un changement de paradigme : on va représenter les préférences des consommateurs en termes de caractéristiques de biens, c'est à dire dans l'espace z_1, z_2 (on choisit de prendre seulement 2 caractéristiques et 6 biens pour simplifier).

On peut en déduire que les consommateurs achetant le *bien 5* valorisent plus les caractéristiques z_1 que z_2 , et inversement pour le *bien 4*.

En fait, la différenciation horizontale et verticale des produits implique qu'une vaste gamme de paniers est disponible dans cet espace de consommation !

- **Différenciation Horizontale** \Rightarrow A prix donné, il n'y a pas unanimité dans le choix des consommateurs entre 2 biens (jaune et rouge) : ce sont des différences de goûts.
- **Différenciation Verticale** \Rightarrow A prix donné, il y a unanimité dans le choix des consommateurs entre 2 voitures biens : l'un est meilleur que l'autre.

Il faut aussi noter que dans le modèle de Rosen, le consommateur n'achète qu'**une seule** unité de bien qui est une combinaison d'attributs z_1, z_2, \dots, z_n . Historiquement, cela s'explique car Rosen s'intéresse principalement aux biens durables (logements, voitures, smartphones...). Il est en effet beaucoup plus simple d'obtenir des caractéristiques observables sur ces biens durables : que ce soit le nombre de pièces pour un logement, la superficie, ou bien la puissance et la longueur d'une voiture.

De toutes ces informations, on peut formuler 2 questions.

- Pour le **producteur**, quelle combinaison de caractéristiques lui permet de maximiser son profit ?
- Pour le **consommateur**, quelle combinaison de caractéristiques lui rapporte le plus d'utilité sous contrainte budgétaire ?

On aboutit à une relation fonctionnelle entre les caractéristiques des biens et leur prix, appelée fonction de prix hédonique $p(z)$.

$$p(z) = p(z_1, z_2, \dots, z_n) \quad (3)$$

Un prix est donc défini en chaque point du plan et guide les choix de localisation des consommateurs et des producteurs concernant les ensembles de caractéristiques achetés et vendus.

Limites

Il n'en reste pas moins qu'il subsiste un problème indéniable : ce qu'on aimerait réellement mesurer c'est le **service rendu par un produit** et non pas les caractéristiques de ce produit. Mais ce premier est complètement inobservable. Un défi sera donc d'interpréter correctement les résultats des régressions.

Application

Harrison Jr et Rubinfeld (1978) :

Objectif : Examiner comment les données du marché immobilier peuvent être utilisées pour évaluer la *Willingness To Pay* des consommateurs pour une meilleure qualité de l'air.

- Le modèle suppose que les ménages prennent en compte le niveau de pollution de l'air, la quantité et la qualité du logement et d'autres caractéristiques de quartier pour faire leur choix.

- La fonction de la valeur hédonique du logement $p(h)$ traduit les attributs du logement en prix, et suppose que les consommateurs perçoivent avec précision ces attributs et que le marché est en équilibre à court terme.

Définition des variables

- W = WTP *marginale* pour une meilleure qualité de l'air
- NOX = Concentration des oxydes d'azote⁽³⁾
- INC = Revenu du ménage en centaine de dollars

Trois niveaux de revenu par an découpés en variable catégorielles :

- **LOW** si $INC \leq \$ 8500 \Rightarrow Y_0$ (Catégorie de référence)
- **MEDIUM** si $INC \leq \$ 11500 \Rightarrow Y_1$
- **HIGH** si $INC \leq \$ 15000 \Rightarrow Y_2$

$$\log(W) = \beta_0 + \beta_1 \log(NOX) + \beta_2 \log(INC) + \beta_3[Y_1 \cdot \log(NOX)] + \beta_4[Y_2 \cdot \log(NOX)] \quad (4)$$

Coefficients estimés pour la régression log – log (significatifs au seuil $p < 0.01$) :

$$\log(W) = \underbrace{2.2}_{\beta_0} + \underbrace{0.97}_{\beta_1} \log(NOX) + \underbrace{0.8}_{\beta_2} \log(INC) - \underbrace{0.03}_{\beta_3}[Y_1 \cdot \log(NOX)] - \underbrace{0.07}_{\beta_4}[Y_2 \cdot \log(NOX)]$$

Résultats : La WTP marginale pour une meilleure qualité de l'air augmente avec le niveau de pollution de l'air et avec le niveau de revenu des ménages.

Pour finir, l'approche hédonique a été utilisée empiriquement dans de très nombreux domaines comme par exemple :

Berndt et Rappaport (2001) \Rightarrow Secteur informatique.

- L'objectif de cet article est d'examiner l'évolution des prix ajustés en qualité des ordinateurs personnels de bureau et mobiles entre 1976 et 1999.

Chen et Rothschild (2010) \Rightarrow Secteur de l'hôtellerie.

- Analyse l'impact des caractéristiques des hôtels de Taipei sur leurs tarifs en utilisant les données de 73 hôtels collectées auprès d'un agent de voyage en ligne.

Yim, Lee, et Kim (2014) \Rightarrow Secteur de la restauration.

- Explore l'impact des attributs des restaurants à Séoul sur leurs prix moyens de repas en examinant les données de 185 établissements recueillies via diverses sources.

Dans la littérature, une spécification *semi-log* est généralement préférée afin d'améliorer l'ajustement du modèle – voir Bello et Moruf (2010).

⁽³⁾Variable de pollution, NOX est un *proxy* pour la qualité de l'air.

Fonction de production

Avant de passer à l'explication de la seconde partie théorique, c'est à dire les modèles SFA, attardons-nous sur la définition d'une fonction de production, fondement important de la SFA.

💡 Rappel

- Un processus de production représente la transformation d'inputs en outputs.
- Dès lors, une fonction de production $f(\cdot)$ donne la quantité maximum d'output y_i pouvant être produite à partir de vecteurs d'inputs.

$$y_i = f(x_i; \beta) \quad (5)$$

- Avec x_i le vecteur d'inputs.
- Avec β le vecteur de paramètres inconnus à estimer.

$f(x_i; \beta)$ est en fait la frontière de production. Pour l'instant cette frontière ne prend pas en compte l'efficacité technique TE_i et elle n'est pas *stochastique* car elle n'inclut pas de terme aléatoire.

* * *

Farrell (1957) est le premier auteur à définir cette *Frontière de Production*.

"When one talks about the efficiency of a firm one usually means its success in producing as large as possible an output from a given set of inputs."

Cette définition permet donc d'aboutir à la formulation évoquée à l'Équation 5.

Le modèle SFA

Aspects théoriques

Aigner, Lovell, et Schmidt (1977) :

Objectif : Formulation et estimation de fonctions de frontière de production stochastique.

Avant les travaux de Aigner, Lovell, et Schmidt (1977), les économètres utilisaient principalement des fonctions de production moyennes dans la littérature, c'est à dire que la formulation théorique énoncée par Farrell (1957) différait de l'utilisation empirique.

- On repart de la fonction de production (Équation 5), mais en lui ajoutant un terme multiplicatif TE_i .

$$y_i = f(x_i; \beta) \cdot TE_i$$

TE_i représente l'efficacité technique, définie comme le ratio d'output observé sur l'output maximum réalisable, soit $TE_i = \frac{y_i}{y_i^*}$.

- Si $TE_i = 1$ alors la firme i produit l'output maximum réalisable, alors que si $TE_i < 1$, il existe un écart entre l'output maximum et l'output effectivement observé.

Un composant **stochastique** $\exp \{v_i\}$ est en outre ajouté pour représenter les chocs aléatoires affectant la production. La fonction de production devient alors :

$$y_i = f(x_i; \beta) \cdot TE_i \cdot \exp \{v_i\}$$

On peut ré-écrire l'efficacité technique sous la forme $TE_i = \exp \{-u_i\}$. Dès lors :

$$y_i = f(x_i; \beta) \cdot \exp \{-u_i\} \cdot \exp \{v_i\} \quad (6)$$

Note : En réarrangeant l'Équation 6 avec le logarithme népérien, on obtient :

$$\Leftrightarrow \ln(y_i) = f(x_i; \beta) + \underbrace{v_i - u_i}_{\epsilon_i}$$

Le modèle peut alors s'écrire sous la forme suivante :

$$\ln(y_i) = f(x_i; \beta) + \epsilon_i \quad (7)$$

L'avantage de cette écriture est qu'elle facilite la manipulation des termes d'erreur, et il est très simple de retrouver le logarithme de l'output maximum. En effet :

$$\Leftrightarrow \ln(y_i) = \underbrace{f(x_i; \beta) + v_i}_{\ln(y_i^*)} - u_i$$

Et donc le logarithme de l'output observé est simplement $\ln(y_i) = \ln(y_i^*) - u_i$.

Les termes d'erreur ϵ_i ont ainsi une distribution particulière composée :

- v_i est un **erreur aléatoire** \Rightarrow variation inexpliquée par les variables indépendantes du modèle, avec $v_i \sim \mathcal{N}(0, \sigma_v^2)$.
- u_i est un **composant unilatéral** qui peut être choisi parmi plusieurs distributions⁽⁴⁾ et $u_i \geq 0$, puisqu'il est nécessaire d'avoir $TE_i \leq 1$.

Conclusion

La spécification de cette méthode permet donc d'estimer les scores de l'**efficacité technique** de chaque firme.

Enfin, Kumbhakar, Horncastle, et al. (2015) discutent aussi dans la section 3.3 de leur livre des approches dites *distribution-free* sur u_i dans lesquelles aucune hypothèse ne sont faites sur la distribution que suit les u_i . Nous ne nous intéresserons pas à ces méthodes puisqu'elles ont le défaut de ne pas pouvoir correctement distinguer les v_i des u_i , et donc ne sont pas en mesure d'estimer les scores d'efficacité technique.

⁽⁴⁾Dans la littérature, deux distributions sont couramment utilisées : la distribution **semi-normale** et **normale tronquée**.

On l'a vu ci-dessus, la SFA est une méthode **paramétrique** qui requiert une forme fonctionnelle précise. La SFA n'a cependant pas le monopole dans le domaine de l'estimation des frontières de production.

Un autre modèle (non-paramétrique) a aussi été développé : la Data Envelopment Analysis (DEA). Celui-ci a l'avantage de ne pas exiger d'hypothèse particulière sur des termes d'erreur. La structure du modèle n'est pas spécifiée à priori mais est uniquement déterminée à partir des données.

* * *

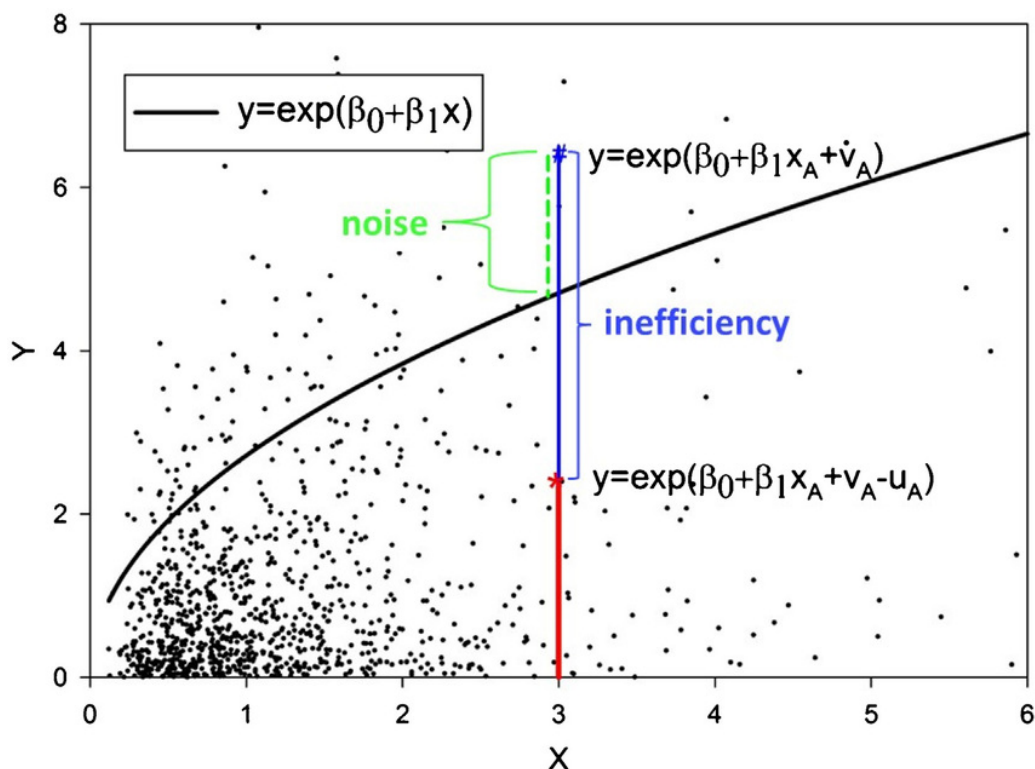


Figure 2 – Représentation graphique d'une SFA.

* Droits d'auteur : **Lutz Bornmann**

A partir de cette représentation on peut clairement distinguer les effets de v_i (**noise**) et ceux de u_i (**inefficiency**) dans un espace à deux dimensions avec X la quantité d'inputs et Y la quantité d'outputs. La frontière optimale de production est ici représentée en noir par $y = \beta_0 + \beta_1 x$.

- 2 entreprises utilisant la même quantité d'inputs ($X = 3$) sont mises en évidence dans le graph. La première se situe en dessous de la frontière de production avec $Y \simeq 2$ et la seconde est au-dessus de celle-ci avec $Y > 6$.
- Les 2 firmes utilisent donc la même quantité d'inputs pour une quantité d'output différente, à savoir : la première firme est moins efficace dans l'utilisation optimale de ses inputs, donc son efficacité technique est inférieure à la seconde.

Utilisation empirique

Quelques exemples d'application de la SFA dans le cadre de la mesure d'efficacité :

Reinhard, Lovell, et Thijssen (2000) \Rightarrow Secteur Environnemental.

- L'objectif de cet article est d'estimer l'efficacité environnementale pour les fermes laitières aux Pays-Bas.

Rosko et Mutter (2008) \Rightarrow Secteur Hospitalier.

- Cet article est quant à lui une méta-analyse de l'ensemble des articles de SFA et de DEA existants sur l'efficacité hospitalière aux Etats-Unis.

Mohamad, Hassan, et Bader (2008) \Rightarrow Secteur Bancaire.

- Compare l'efficacité des coûts et des profits de 80 banques dans 21 pays comprenant 37 banques conventionnelles et 43 banques islamiques.

En bref, il existe de nombreux domaines d'application !

Un domaine en particulier n'a pourtant pas été évoqué jusqu'ici : pourquoi ne pas utiliser la SFA pour mesurer l'efficacité d'un prix (**best-buy frontier**) ?

C'est précisément le cadre du prochain article de notre revue de la littérature.

SFA & Pricing Hédonique

Arrondo, Garcia, et Gonzalez (2018) :

Objectif : déterminer les attributs principaux des prix des sneakers en Espagne et leur efficacité.

Six caractéristiques⁽⁵⁾ sont étudiées sur $n = 171$ sneakers.

- **Lightweight** : poids des sneakers.
- **Cushioning** : capacité de la chaussure à absorber les chocs au cours d'une course et tout au long du cycle de vie du produit.
- **Flexibility** : les baskets flexibles s'adaptent mieux à la forme naturelle du pied.
- **Response** : capacité du matériau à retrouver sa forme après les déformations provoquées par l'impact sur le sol.
- **Grip** : l'adhérence donne aux coureurs une certaine assise sur le sol.
- **Stability** : mesure la stabilité du pied à l'intérieur de la chaussure.

En plus de ces 6 caractéristiques techniques, la marque est ajoutée en tant que variable qualitative pour mesurer la *Brand Equity* (la valeur d'une marque pour le consommateur).

⁽⁵⁾Variables quantitatives discrètes $\in [1, 10[$.

Le modèle s'écrit alors :

$$\ln(p_{ik}) = \alpha_k + \beta X_{ik} + v_{ik} + u_{ik} \quad (8)$$

- p_{ik} est le prix du i -ème modèle de marque k
- α_k est l'effet marque sur le prix de la marque k
- X_{ik} est le vecteur des attributs mesurables du i -ème modèle de marque k
- β est un vecteur de coefficients pour ces attributs
- v_{ik} est une erreur aléatoire
- u_{ik} représente l'inefficacité

Note : On retrouve bien la forme spécifique d'une SFA, caractérisée par la présence des termes v_{ik} et u_{ik} . La seule différence est que le terme d'erreur composée est $\epsilon_{ik} = v_{ik} + u_{ik}$ car nous sommes dans le cadre d'une **frontière de coût** et non de production.

Résultats :

Table 1 – Résultats de la régression hédonique

Variables	Coefficient	SE
<i>Lightness</i>	0.007	0.028
<i>Cushioning</i>	0.064 **	0.025
<i>Flexibility</i>	0.058 **	0.026
<i>Response</i>	0.050 *	0.30
<i>Stability</i>	0.070 ***	0.025
<i>Grip</i>	-0.045	0.028
Adidas	2.697 ***	0.401
Asics	2.679 ***	0.389
Saucony	2.779 ***	0.403
Nike	2.714 ***	0.422
Brooks	2.834 ***	0.404
Mizuno	2.524 ***	0.397
New Balance	2.544 ***	0.410
Reebok	2.522 ***	0.403

Les variables *Cushioning*, *Flexibility* et *Stability* sont statistiquement significatives à $p < 0.05$.

De plus, nous sommes ici dans le cadre d'une régression log-linéaire donc les coefficients peuvent être interprétés comme des **semi-élasticités**, c'est à dire :

⇒ Pour une augmentation d'une unité de *Stability*, p_{ik} va augmenter de 7%, *cet. par.* ⁽⁶⁾

⇒ Pour une augmentation d'une unité de *Cushioning*, p_{ik} va augmenter de 6.4%, *cet. par.*

⇒ Pour une augmentation d'une unité de *Flexibility*, p_{ik} va augmenter de 5.8%, *cet. par.*

Par conséquent, la caractéristique *Stability* va avoir le plus grand impact sur le prix d'une sneakers, suivi de *Cushioning* et *Flexibility*.

⁽⁶⁾Toutes choses égales par ailleurs.

Table 2 – Indice d'efficacité moyen par marque

Marque	$\hat{\theta}_k$
Adidas ($n = 28$)	0.832
Asics ($n = 35$)	0.864
Saucony ($n = 15$)	0.875
Nike ($n = 25$)	0.824
Brooks ($n = 16$)	0.860
Mizuno ($n = 29$)	0.858
New Balance ($n = 18$)	0.848
Reebok ($n = 5$)	0.859

$\hat{\theta}_k$ représente l'indice d'efficacité moyen estimé par marque, compris entre 0 et 1.

On remarque tout d'abord que cet indice est compris entre 0.8 et 0.9 pour l'ensemble des marques, c'est à dire qu'il n'y a pas de marque globalement **très inefficace** (si une marque l'était, elle n'arriverait probablement pas à vendre et serait évincée par ses concurrents).

- Nike est la marque qui possède la pire relation prix~attributs de la sélection.
- Saucony est la marque qui possède la meilleure relation prix~attributs de la sélection.

Résultats

- En estimant l'efficacité des produits, l'article permet de déterminer le montant des réductions à accorder aux sneakers **overprice** afin de les rendre compétitives.
- Il existe une relation inverse entre l'efficacité du produit et la réduction de prix : la réduction de prix est d'autant plus grande que la sneakers est **overprice**.

| Choix et cadrage de la problématique

L'objectif fixé par notre sujet est de combiner les modèles SFA à une problématique d'étude des prix hédoniques, de manière similaire à ce qui a été entrepris par Arrondo, Garcia, et Gonzalez (2018).

L'ensemble des articles de la littérature exposés ci-dessus ont permis d'affiner notre compréhension théorique des modèles et nous ont aidé à déterminer un marché à étudier. Pour des raisons de disponibilité des caractéristiques et parce que peu d'articles dans la littérature se sont intéressés au pricing hédonique des smartphones, nous avons fait le choix d'analyser le marché de la téléphonie mobile.

Notre problématique est donc la suivante :

Combinaison d'un modèle SFA et d'une régression hédonique pour évaluer la discordance entre les prix de smartphones et leur valeur (intrinsèque).

Le marché de la téléphonie mobile, en constante évolution

Depuis l'apparition des téléphones mobiles au début des années 1990, de nombreuses innovations technologiques ont ajouté des caractéristiques rendant ces téléphones de plus en plus polyvalents. Cette chronologie présente en X les années et les rectangles des différentes catégories correspondent à des **débuts** et des **fin de commercialisation**. L'axe Y permet quant à lui d'améliorer la lisibilité.

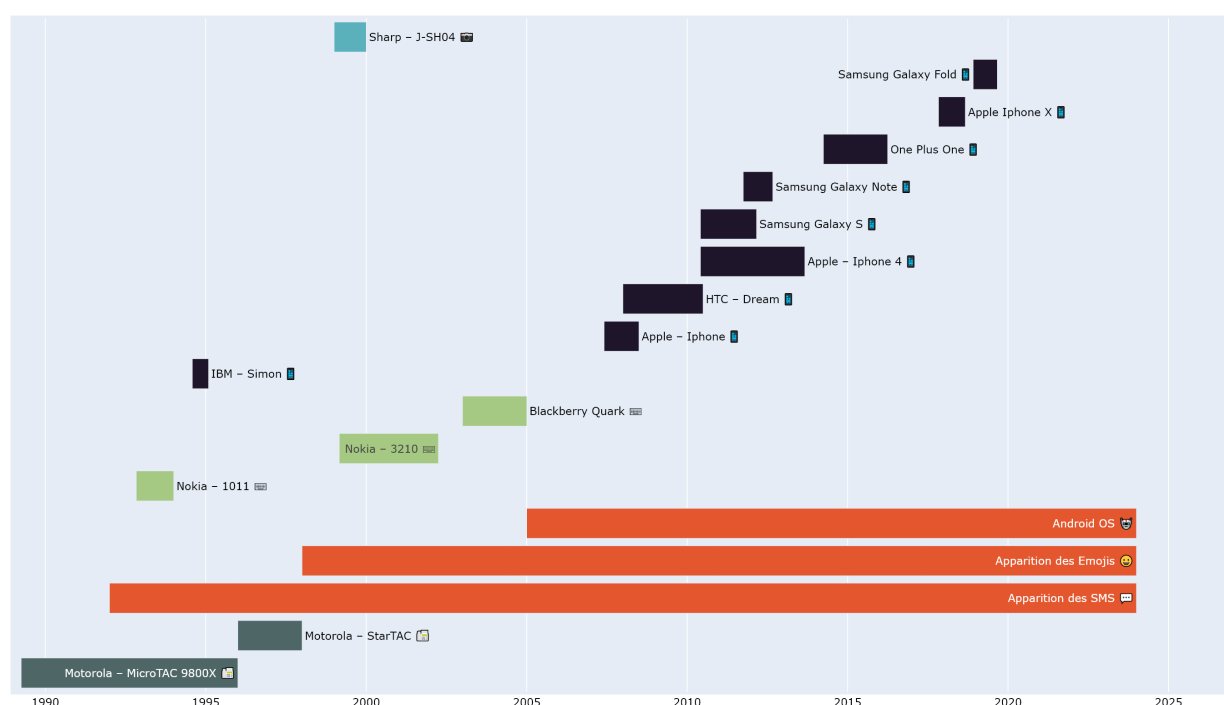




Figure 3 – Smartphone Timeline.

■ Téléphone à clapet ■ Téléphone à clavier ■ Téléphone à appareil photo ■ Smartphone

Examinons quelques modèles de téléphone pour mieux saisir l'impact des innovations majeures sur le marché.

- *Nokia 1011* : Premier écran LCD.
- *IBM Simon* : Premier véritable smartphone avec stylet, commercialisé pendant seulement 6 mois à cause d'un prix élevé de \$899.
- *Nokia 3210* : Premier téléphone à intégrer les SMS et plusieurs jeux. C'est encore aujourd'hui un des téléphones les plus vendus au monde.
- *Sharp J-SH04* : Premier téléphone équipé d'un appareil photo intégré.
- *Blackberry Quark* : Les téléphones  BlackBerry sont les premiers à disposer d'un clavier complet, ce qui, à cette époque, est un avantage majeur. A tel point qu'au début des années 2000 et jusqu'en 2010, BlackBerry devient et reste leader sur le marché de la téléphonie mobile avec 20% de parts de marché à son apogée.
- *Apple Iphone* : En 2007,  Apple annonce l'iPhone. Ce téléphone, qui intègre un écran tactile multitouch, va bouleverser le marché des téléphones mobiles. La vraie révolution, plus que le téléphone en lui-même, est l'*App Store*, qui va permettre d'accélérer le développement de nombreuses applications mobiles.
- *HTC Dream* : Un an après la sortie de l'iPhone, les constructeurs bataillent pour tenter de le concurrencer. HTC est dans ce cadre le premier à intégrer Android OS. Il reste néanmoins un entre-deux (il possède un clavier et un écran tactile).
- *Samsung Galaxy S* : Avec le Galaxy S, Samsung concurrence directement l'Apple iPhone 4 et sort un téléphone meilleur en tout point sur le plan des caractéristiques techniques. L'écran est plus grand, il existe une possibilité d'augmenter le stockage, il possède un meilleur cpu et une meilleure autonomie, tout en étant moins cher.

Conclusion

Toutes ces innovations vont avoir un impact dans les caractéristiques les plus valorisées par les consommateurs. Par exemple, il est difficile d'imaginer qu'un consommateur valorisera aujourd'hui un téléphone sans capteur de caméra frontale et arrière ou qui serait incapable d'envoyer des SMS.

Cela permet d'ailleurs d'évoquer une des limites majeures des modèles de pricing hédonique. Comment va-t-on pouvoir modéliser l'arrivée d'une nouvelle caractéristique ? On ne peut pas trouver dans le passé quelle va être la valorisation de cette nouvelle caractéristique.

Illustrons cette remarque avec l'iPhone. Un modèle de régression des prix hédoniques réalisé juste avant la sortie de l'iPhone aurait probablement trouvé (sans surprise) que BlackBerry était la marque la plus valorisée par les consommateurs et qu'il faut augmenter la taille du téléphone pour lui permettre d'avoir un plus grand clavier. Il va sans dire que deux mois plus tard, ces résultats sont inutilisables à cause d'une innovation technologique.

Enfin, il existe relativement peu d'articles sur les prix hédoniques des smartphones, ou alors ils sont assez anciens (2004-2005), et on l'a vu, étant donné la vitesse à laquelle évolue le marché, avoir des données récentes est primordial pour estimer correctement les caractéristiques valorisées par les consommateurs à un instant T .

Smartphones et Pricing Hédonique

Il existe néanmoins quelques articles récents traitant du sujet, dont celui de Ahmad, Ahmed, et Ahmad (2019) :

Objectif : Pricing des attributs des smartphones au Pakistan

Les données des attributs ont été collectées sur des sites webs et les prix pratiqués relevés dans les magasins de 2 villes du Pakistan ($n = 348$ smartphones).

Le prix moyen d'un smartphone dans leur étude est de \$136,35. En outre, l'**écart-type** du prix des smartphones est élevé (181), c'est à dire que la dispersion en prix est assez importante, ce qui confirme l'hypothèse que les smartphones sont des biens différenciés.

Ils proposent alors l'estimation du modèle suivant avec les caractéristiques découpées en variables catégorielles.

$$\ln(PRICE_i) = \beta_0 + \beta_{1i}BRAND_i + \beta_{2i}WEIGHT_i + \beta_{3i}BATTERY_i + \beta_{4i}OS_i + \beta_{5i}RAM_i + \beta_{6i}MEMORY_i + \beta_{7i}DISPLAY_i + \beta_{8i}NETWORK_i + \beta_{9i}BCAM_i + \beta_{10i}FCAM_i + \epsilon_i$$

Résultats :

- La marque, la batterie, le poids, l'OS, la RAM, la mémoire et la taille de l'écran ont un effet positif statistiquement significatif sur les prix des smartphones.

Plus précisément, les résultats indiquent que les fabricants doivent se concentrer sur un téléphone :

- avec une RAM de plus d'1 Go.
- avec une Mémoire de plus de 8 Go.
- avec un Ecran de plus de 5 pouces.
- compatible avec la 4G.
- avec une caméra arrière de plus de 15 mégapixels.

Le Pakistan étant un pays en voie de développement, et l'étude datant de 2019, on peut s'attendre à trouver des résultats différents dans nos données.

De plus, sur les 348 smartphones, 127, sont de la marque **QMOBILE**, une société Pakistanaise qui vend des smartphones à bas prix, ce qui peut aussi expliquer le prix moyen assez bas.

Effet de réputation

Dans la section précédente, les résultats de l'étude de Ahmad, Ahmed, et Ahmad (2019) ont permis de discerner que la marque a un effet statistiquement significatif sur le prix des smartphones, c'est pourquoi nous voulions explorer rapidement des questions d'analyse de la concurrence que l'on peut relier à notre sujet.

Boistel (2008) parle spécifiquement de cet effet de réputation.

Objectif : Analyser l'impact de la réputation sur les fonctions clés de l'entreprise et son intégration dans le management stratégique actuel.

Réputation et marketing

- **Considéré comme une priorité majeure :** La réputation est une ressource essentielle, reconnue comme une priorité de recherche par le *Marketing Science Institute*.
- **Influence le comportement des consommateurs:** La réputation impacte l'intention d'achat, la confiance envers les nouveaux produits et est liée à la satisfaction client. Elle agit comme une "garantie".
- **Avantage compétitif :** Une solide réputation permet de gagner un avantage compétitif sur le marché, voire un avantage concurrentiel, en attirant les clients et en se différenciant des concurrents.
- **Limitation de la concurrence :** les produits ou services d'une entreprise réputée sont moins facilement remplaçables ou imitables en raison de la perception limitée des consommateurs. Au lieu d'examiner les caractéristiques en détail, ils vont se fier à la marque et à la réputation.
- **Corrélation positive avec le prix :** Une meilleure réputation va permettre de fixer des prix plus élevés et d'obtenir un avantage sur les ventes par rapport à la concurrence.
Exemple : Toyota et General Motors forment la joint-venture New United Motor Manufacturing Inc. La société a produit 2 voitures identiques :
 - la Toyota Corrola.
 - la GM's Geo Prizm.

⇒ La meilleure réputation de Toyota lui a permis de vendre 200000 voitures à \$ 11100, contre seulement 80000 véhicules à \$ 10700 vendu pour General Motors.

| Acquisition des données

Scraping

Il n'existe **évidemment** pas de données directement disponibles recoupant le prix et l'ensemble des caractéristiques des smartphones. Ce qui pourrait le plus s'en rapprocher sont les fiches techniques de téléphones disponibles sur <https://www.01net.com/>. Scraper ce site pourrait être une idée intéressante, mais *01net* n'est pas un revendeur de smartphones.

L'idée est donc de récupérer ces données sur le site d'un revendeur (*Fnac, Darty, Boulanger*). En effet, l'avantage de la récupération des données sur un site de revente direct est que nous avons une "*photographie*" du marché au moment où le *crawler* récupère et alimente notre base de données. A ce titre, nous avons choisi de récupérer des données sur **Boulanger**.

⚠ Encadrement du web scraping

Le web scraping est encadré en droit français par l'article **L. 342-3⁽⁷⁾** du Code de la propriété intellectuelle, qui autorise la pratique suivante :

- L'extraction et la réutilisation d'une partie substantielle, appréciée de façon qualitative ou quantitative, à des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche et pour un public composé d'élèves, d'étudiants, d'enseignants ou de chercheurs directement concernés. Ainsi, **ce cas de figure étant limité à des fins pédagogiques, il est totalement exclu de faire usage des données extraites à titre commercial.**

Nous précisons donc que nous ne ferons en aucun usage de ces données dans un cadre commercial.

Méthodologie

L'objectif final est de disposer d'une application permettant aux consommateurs ou aux producteurs de comparer l'efficacité des smartphones en fonction de leurs caractéristiques et de leur indiquer quel est le meilleur choix.

Workflow :

- *Scraping* ⇒ **Python**
- *Nettoyage des données* ⇒ **Python**
- *Modélisation* ⇒ **R**
- *Application* ⇒ **Python**

⁽⁷⁾Plus de détail sur legifrance.gouv.fr

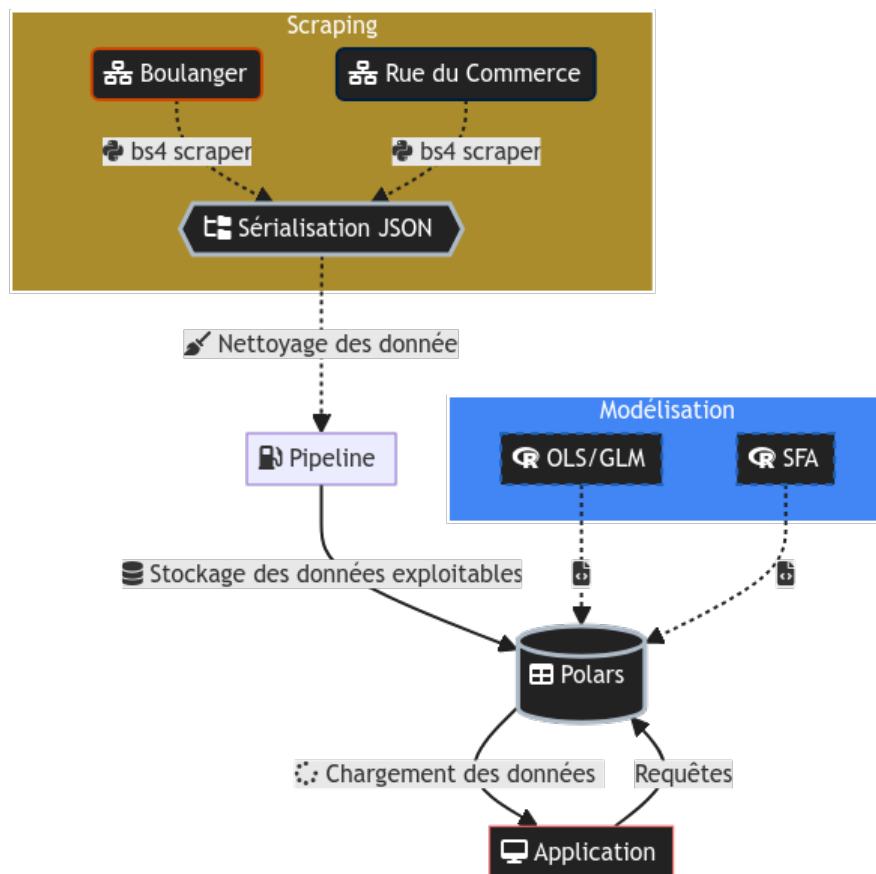


Figure 4 – Diagramme fonctionnel.

Le diagramme fonctionnel permet de comprendre comment interagissent les différents composants logiciels avant leur utilisation dans l'application.

| Statistiques descriptives

Il y a 487 smartphones valides disponibles dans nos données scrapées sur Boulanger avec 35 variables.

- Variables liées à l'écran (*screen_type, screen_size, screen_tech, diagonal_pixels, ppi, resolution_1, resolution_2*).
- Variables liées à la caméra (*mpx_backward_cam, cam_1, cam_2, cam_3, sensor*).
- Variables liées aux caractéristiques physiques du téléphone (*color, thickness, width, height, net_weight*).
- Variables liées aux performances (*network, cpu, ram, storage, upgrade_storage*).
- Variables liées à la batterie (*battery, fast_charging, induction, usb_type_c*).
- Variables liées au DAS (*das_limbs, das_chest, das_head*).
- Autres variables (*repairability_index, model, brand, made_in, stars, reviews*).

Et enfin notre variable à expliquer : **price**.

Analyse approfondie des prix

Mesures de tendance centrale

Le prix moyen d'un smartphone de la sélection est de 711.11 €, soit $\simeq 5$ fois plus élevé que dans l'article de Ahmad, Ahmed, et Ahmad (2019). Cela peut s'expliquer notamment par la différence considérable de **PIB** par habitant.⁽⁸⁾

- En 2022 au Pakistan : \$ 1596.7
- En 2022 en France : \$ 40963.8

La médiane est quant à elle de 597.64 €.

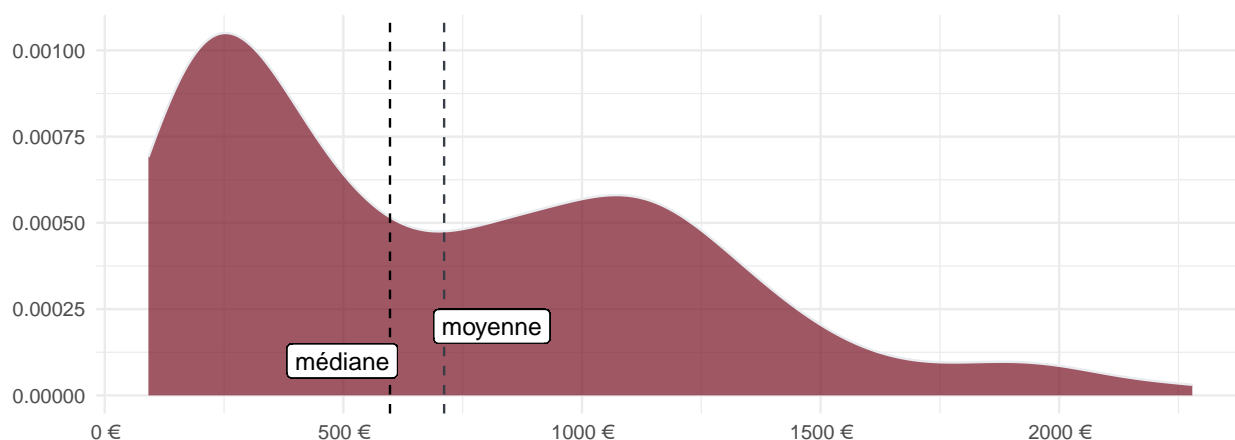


Figure 5 – Distribution des prix des smartphones

⁽⁸⁾Données issues de la *Banque Mondiale* (PIB par habitant en US dollars courants)

On peut aussi tester l'asymétrie de la distribution avec le coefficient d'asymétrie de *Pearson* :

$$SK = \frac{3(\bar{x} - \tilde{x})}{\sigma} \simeq 0.67$$

Ce résultat indique que l'asymétrie de la distribution est positive. Il y a beaucoup plus de valeurs concentrées à gauche de la distribution qu'à droite. Si le coefficient était proche de 0, cela signifierait que la distribution est proche d'une loi normale \mathcal{N} , ce qui n'est pas le cas.

Mesures de dispersion

Le prix minimal d'un smartphone dans notre sélection est de 91.4 € pour le modèle **Xiaomi Redmi A1** et le prix maximal est de 2279 € pour le modèle **Samsung Galaxy Z Fold 5**.

Il existe donc une grande étendue de prix, c'est à dire : 2187.6 €. De plus, l'écart-type du prix est très important (506.36).

Toutes ces mesures nous confirment que la dispersion en prix est très élevée.

Prix moyen en fonction des caractéristiques

Table 3 – Prix moyen en fonction de la RAM.

ram	n	mean_price
2 Go	21	110.11 €
3 Go	16	165.35 €
4 Go	118	413.4 €
6 Go	87	722.22 €
8 Go	163	835.55 €
12 Go	75	1144.02 €
16 Go	7	1106.14 €

Table 4 – Prix moyen en fonction du stockage.

storage	n	mean_price
32 Go	27	116.65 €
64 Go	55	214.76 €
128 Go	207	537.44 €
256 Go	127	930.65 €
512 Go	55	1318.86 €
1000 Go	16	1835.65 €

- On peut voir que plus la RAM augmente, plus le prix moyen du téléphone augmente (sauf à partir de 12 Gigaoctets). On a en effet remarqué plus haut que le téléphone le plus cher était le **Samsung Galaxy Z Fold 5**, qui possède 12 Go de RAM, et non 16. Ce constat, en plus du nombre très limité de téléphones disposant de 16 Go de RAM (7), peut expliquer pourquoi le prix moyen des téléphones ayant 16 Go de RAM est inférieur au prix moyen des téléphones disposant de 12 Go de RAM.

- Il existe une relation non-linéaire concernant le doublement du stockage. Par exemple, passer de 64 à 128 Go implique une augmentation du prix moyen de 250% alors que passer de 256 à 512 Go de stockage implique seulement 141% d'augmentation du prix moyen

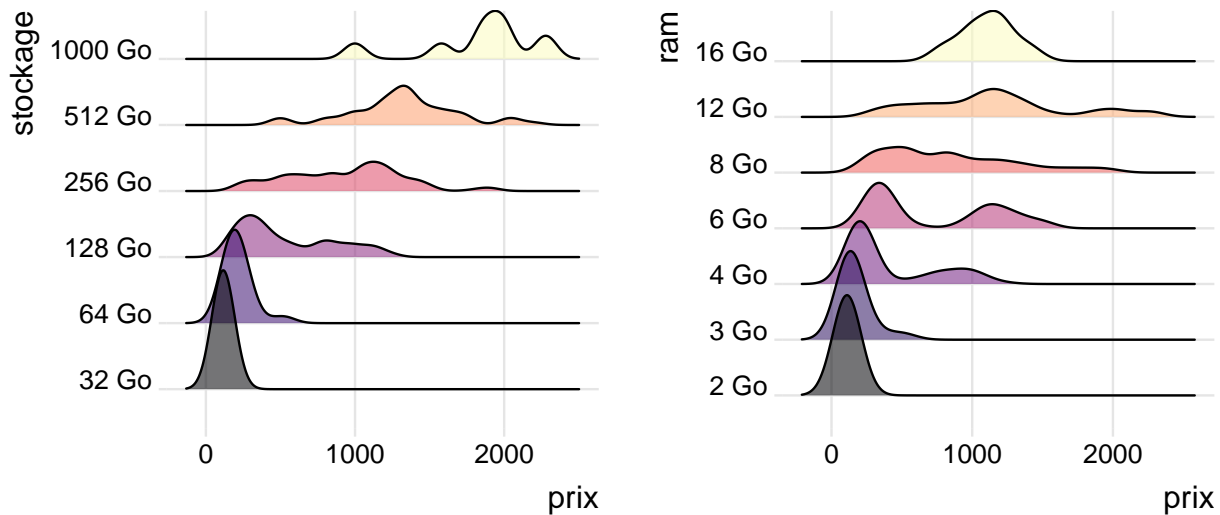


Figure 6 – Ridge plot : RAM et Stockage

Faire un giga tableau avec min/max/mean/q1/q2/q3, etc.

Faire des barcharts pour les variables TRUE/FALSE (induction, usb_type_c, fast_charging)

TODO => ACP ?

tests statistiques ?

| Modélisation économétrique

Verifier les hypothèses BLUE, hétéroscédasticité, découpage en variables dichotomiques, etc.

| Définitions

WTP Willingness To Pay

SFA Stochastic Frontier Analysis

DAS Débit d'Absorption Spécifique

DEA Data Envelopment Analysis

TE Technical Efficiency

| Références

- Ahmad, Waseem, Tanvir Ahmed, et Bashir Ahmad. 2019. « Pricing of mobile phone attributes at the retail level in a developing country: Hedonic analysis ». *Telecommunications Policy* 43 (4): 299-309. <https://doi.org/10.1016/j.telpol.2018.10.002>.
- Aigner, Dennis, C. A.Knox Lovell, et Peter Schmidt. 1977. « Formulation and estimation of stochastic frontier production function models ». *Journal of Econometrics* 6 (1): 21-37. [https://doi.org/10.1016/0304-4076\(77\)90052-5](https://doi.org/10.1016/0304-4076(77)90052-5).
- Arrondo, Ruben, Nuria Garcia, et Eduardo Gonzalez. 2018. « Estimating product efficiency through a hedonic pricing best practice frontier ». *BRQ Business Research Quarterly* 21 (4): 215-24. <https://doi.org/10.1016/j.brq.2018.08.005>.
- Bello, Ajide K, et Alabi Moruf. 2010. « Does the functional form matter in the estimation of hedonic price model for housing market ». *The Social Sciences* 5 (6): 559-64.
- Berndt, Ernst R, et Neal J Rappaport. 2001. « Price and quality of desktop and mobile personal computers: A quarter-century historical overview ». *American Economic Review* 91 (2): 268-73. <https://www.aeaweb.org/articles?id=10.1257/aer.91.2.268>.
- Boistel, Philippe. 2008. « La réputation d'entreprise: un impact majeur sur les ressources de l'entreprise ». *Revue management et avenir*, n° 3: 9-25.
- Chen, Ching-Fu, et Rochelle Rothschild. 2010. « An application of hedonic pricing analysis to the case of hotel rooms in Taipei ». *Tourism Economics* 16 (3): 685-94. <https://journals.sagepub.com/doi/abs/10.5367/000000010792278310>.
- Farrell, Michael James. 1957. « The measurement of productive efficiency ». *Journal of the Royal Statistical Society Series A: Statistics in Society* 120 (3): 253-81. <https://doi.org/10.2307/2343100>.
- Harrison Jr, David, et Daniel L Rubinfeld. 1978. « Hedonic housing prices and the demand for clean air ». *Journal of environmental economics and management* 5 (1): 81-102. [https://doi.org/10.1016/0095-0696\(78\)90006-2](https://doi.org/10.1016/0095-0696(78)90006-2).
- Kumbhakar, Subal, Alan Horncastle, et al. 2015. *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.
- Lancaster, Kelvin J. 1966. « A New Approach to Consumer Theory ». *Journal of Political Economy* 74 (2): 132-57. <https://doi.org/10.1086/259131>.
- Mohamad, Shamsheer, Taufiq Hassan, et Mohamed Khaled I Bader. 2008. « Efficiency of conventional versus Islamic Banks: international evidence using the Stochastic Frontier Approach (SFA) ». *Journal of Islamic economics, banking and finance* 4 (2): 107-30.
- Reinhard, Stijn, CA Knox Lovell, et Geert J Thijssen. 2000. « Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA ». *European Journal of Operational Research* 121 (2): 287-303.
- Rosen, Sherwin. 1974. « Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition ». *Journal of Political Economy* 82 (1): 34-55. <http://www.jstor.org/stable/1830899>.
- Rosko, Michael D, et Ryan L Mutter. 2008. « Stochastic frontier analysis of hospital inefficiency: a review of empirical issues and an assessment of robustness ». *Medical care research and review* 65 (2): 131-66.
- Yim, Eun Soon, Suna Lee, et Woo Gon Kim. 2014. « Determinants of a restaurant average meal price: An application of the hedonic pricing model ». *International Journal of Hospitality Management* 39: 11-20. <https://www.sciencedirect.com/science/article/abs/pii/S027843191400019X>.