

Pondération *ex post* des données de prix pour l'estimation des taux de dépréciation

Marc Tanguay et Pierre Lavallée¹

Résumé

Pour modéliser la dépréciation économique, on utilise une base de données qui contient des informations sur les actifs dont des entreprises se départissent. On connaît les prix d'acquisition et de revente ainsi que les durées d'utilisation de ces actifs. Cependant, les actifs dont on observe les prix sont uniquement ceux qui ont fait l'objet d'une transaction. Bien que la dépréciation d'un actif soit présente de façon continue au cours de sa vie, on ne connaît donc cette valeur que lorsqu'il y a eu transaction. La présente note propose une pondération *ex post* afin d'atténuer, au moins en partie, cet effet dans la détermination des modèles économétriques.

Mots clés : Ratio de prix; données de survie; distribution uniforme; dépréciation des voitures.

1. Contexte

Différents modèles économétriques sont utilisés pour estimer la dépréciation économique. On utilise, à cette fin, une base de données qui contient des informations sur les actifs dont des entreprises se départissent. On connaît les prix d'acquisition et de revente ainsi que les durées d'utilisation de ces actifs. On voudrait en inférer les résultats à la population totale des actifs utilisés par les entreprises. Sur l'utilisation de prix d'actifs usagés pour estimer la dépréciation économique, on peut notamment consulter Gellatly, Tanguay et Yan (2002), ainsi que Hulten et Wykoff (1981).

On s'interroge cependant sur la représentativité de la base de données utilisée. En effet, les actifs dont on observe les prix sont uniquement ceux qui ont fait l'objet d'une transaction. On ignore dans quelle mesure les pertes de valeur observées sur eux sont représentatives de la perte de valeur pour tous les actifs en production, qu'ils aient ou non fait l'objet d'une transaction. Ceci peut constituer une source d'erreur dans l'établissement des modèles économétriques parce que ces derniers cherchent à mesurer la dépréciation des actifs au cours de leur vie, qu'il y ait eu transaction ou non.

Nous nous proposons d'atténuer cette source d'erreur, en introduisant une pondération *ex post* dans la détermination des modèles économétriques. La section 2 de la présente note décrira plus en détail la problématique. À la section 3, nous exposerons la démarche suivie pour la détermination de la pondération. Finalement, à la section 4, nous présenterons quelques résultats numériques.

2. Problématique

On cherche à décrire la relation entre les prix et l'âge des actifs. On dispose d'un échantillon de n actifs où on connaît, pour chaque actif i , le ratio de prix r_i et le temps t_i où ce

ratio a été mesuré. Une fois que les prix sont exprimés en dollars réels, ce ratio est donné par $r_i = P_i^t / P_i^0$ où P_i^0 est la valeur initiale de l'investissement de l'actif i et P_i^t est son prix de revente au temps t . Ce ratio est strictement décroissant par rapport à l'axe du temps t . Au point de départ, on ignore le processus qui génère la perte de valeur et on n'a aucune spécification concernant la fonction qui décrit cette perte sinon qu'elle est strictement décroissante. Il est cependant possible d'examiner la distribution des ratios des prix entre 0 et 1. Voici un exemple construit à partir des données sur les usines de fabrication (on doit noter que les 2/3 de l'échantillon ont été exclus car il correspondent à des mises au rancard (le prix est nul) et les procédures d'estimation prennent en compte, chacune à sa façon, cette composante).

Étant donné que l'on veut utiliser les données pour inférer des statistiques sur la population des actifs en production, on souhaiterait que nos données aient des propriétés analogues à celles d'un échantillon aléatoire qui serait tiré sur cette population. Ceci n'est pas le cas, rappelons-le, parce qu'on ne dispose que des prix des actifs i qui ont fait l'objet d'une transaction au temps t_i , $i = 1, \dots, n$. En effet, bien que l'on voudrait disposer de ratios des prix pour différentes périodes de l'existence d'un actif i donné, ce ratio n'est disponible que lorsqu'il y a eu transaction, ce qui survient de façon non uniforme durant la durée de vie d'un actif.

On peut donc se demander quelle forme aurait la distribution ci-dessus si elle avait été tirée d'un échantillon où le ratio des prix avait été mesuré, pour un même actif i , à différents temps t . Notre argument est qu'elle devrait converger vers une *distribution uniforme*. Nous allons donc chercher à obtenir une pondération qui nous aidera à recréer une distribution uniforme des ratios de prix. Cette pondération nous aidera à pallier le manque d'uniformité dans la distribution des observations, ce qui peut influencer les analyses statistiques comme, par exemple, la régression linéaire.

1. Marc Tanguay et Pierre Lavallée, Statistique Canada, Ottawa (Ontario), Canada K1A 0T6. Courriel : marc.tanguay@statcan.ca, pierre.lavallee@statcan.ca.

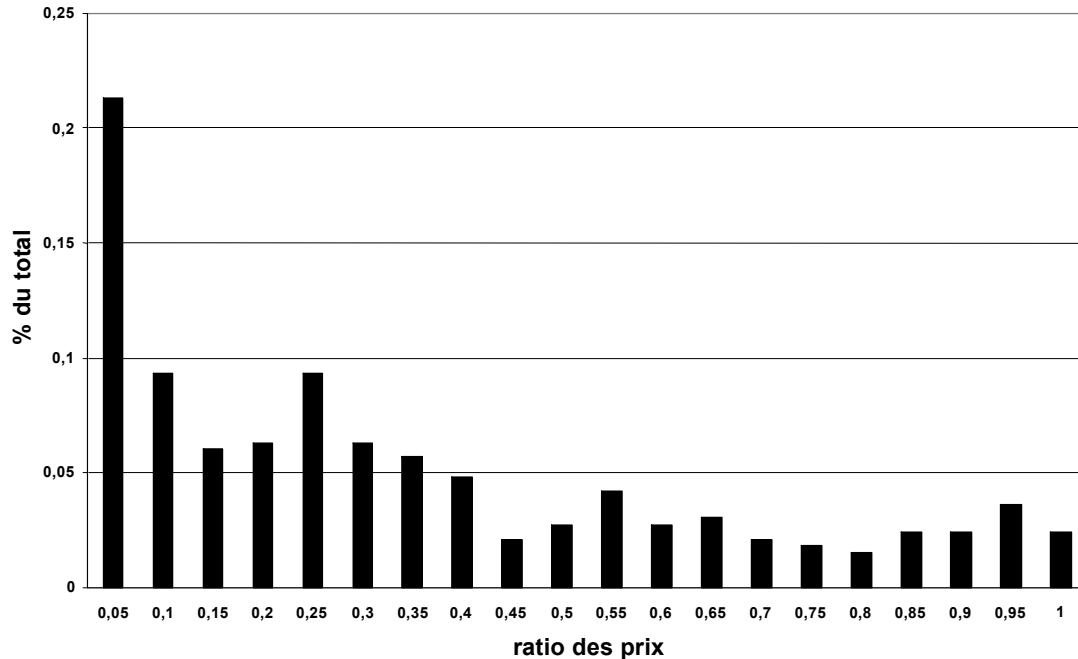


Figure 1 Distribution des observations selon le ratio des prix, usines de fabrication

3. Démarche

Notre point de départ est que les ratios de prix peuvent être considérés comme des réalisations empiriques d'une fonction de survie de forme inconnue. Dans les modèles de durée de vie, la fonction de survie exprime la probabilité qu'une entité dont la vie est limitée survive au-delà d'un certain point sur l'axe du temps. Elle fournit, par conséquent, la même information que la fonction de répartition (ou *Cumulative Distribution Function*). Soit r , une variable aléatoire qui décrit la durée de vie d'une unité de valeur incorporée dans un actif quelconque. La valeur s'épuise au fur et à mesure que le temps passe, et ce, aussi longtemps que l'actif est en service. Le ratio des prix peut donc s'interpréter comme la fraction survivante qui diminue peu à peu. On note cette fraction $S(y)$ et on a

$$S(y) = 1 - F(y)$$

où $F(y) = P(r \leq y)$ est la fonction de répartition, c'est-à-dire la probabilité qu'une unité de valeur soit perdue avant le point y .

Les théorèmes des transformations fondamentales des lois de probabilité permettent de décrire la fonction inverse de $F(y)$ (Greene 1993 et Ross 2002). Soit $z = F(y)$ et supposons que la fonction inverse F^{-1} existe de sorte que $y = F^{-1}(z)$. Il y a donc une concordance directe entre l'espace de y , borné à 0 mais infini à droite, et celui de F qui

est borné entre 0 et 1. La fonction de répartition de z est $F(F^{-1}(z)) = z$. La loi qui génère une telle répartition est une distribution uniforme entre 0 et 1.

Ce résultat est généralement au cœur des processus de génération de données comme les simulations Monte-Carlo puisque lors de la génération d'un échantillon aléatoire, on utilise souvent une distribution uniforme à laquelle on applique ensuite la fonction inverse (Davidson et MacKinnon 1993). Cette approche n'est toutefois pas toujours pratique ou carrément impossible, en particulier si la fonction inverse F^{-1} n'a pas de forme explicite. Ce résultat a aussi été utilisé dans les approches de résidus généralisés, notamment pour la construction de tests de spécification (Lancaster 1985).

Il en résulte que n'importe quel échantillon aléatoire construit à partir de réalisations empiriques de données de proportion de survie doit converger en distribution vers une distribution uniforme.

Dans le cas des données de prix, l'intuition est la suivante : entre l'investissement et la mise au rancard, toute la fourchette des prix relatifs doit forcément être couverte par un actif en production. À la période initiale, la valeur est perdue plus rapidement, il y a donc une plus grande quantité d'observations dont les durées sont courtes. Mais cela est compensé par le fait que la référence correspondante sur l'échelle du temps est également plus courte. Par exemple, il faut moins de temps pour passer de 100 % de la valeur initiale à 90 %, que de 15 % à 5 % de la valeur initiale.

Il est facile de vérifier ces résultats de façon numérique, à l'aide de données simulées et nous ne nous y attarderons pas. Nous allons plutôt examiner comment ce résultat peut être réintroduit dans la base de données pour lui restaurer, au moins en partie, des propriétés semblables à celle d'un échantillon aléatoire. *Pour ce faire, il nous suffit d'imposer ex post, à la distribution empirique des prix, une structure de poids w_i qui soit telle que la distribution empirique des données, dans l'espace des prix, soit uniforme.*

La distribution empirique des ratios de prix r est donnée par

$$\hat{F}_n(y) = \frac{\sum_{i=1}^n I_i(y)}{n} \quad (1)$$

où $I_i(y) = 1$ si la valeur mesurée r_i de l'actif i est inférieure ou égal à y (c'est-à-dire, $r_i \leq y$), et 0 sinon, et n est le nombre total d'observations. Notons que si les n unités de l'échantillon sont indépendantes et identiquement distribuées (i.i.d.), lorsque $n \rightarrow \infty$, $\hat{F}_n(y)$ converge en probabilité vers $F(y)$, c'est-à-dire $\hat{F}_n(y) \xrightarrow{P} F(y)$ (Bickel et Doksum 1977).

Pour obtenir le poids w_i pour chaque actif i , on distribue simplement l'échantillon en un nombre H donné d'intervalles (ou classes) de largeur fixe sur l'échelle des ratios de prix, et on attribue la même probabilité $\pi = 1/H$ à chacun de ces intervalles. Puisque les ratios de prix sont contenus entre 0 et 1, on a alors l'intervalle $h = 1$ donné par $[0, H^{-1}]$, et pour $h = 2, \dots, H$, les intervalles sont donnés par $[(h-1)H^{-1}, hH^{-1}]$. Un poids w_h est ensuite calculé dans chaque intervalle h par le ratio $\pi / \hat{\pi}_h$ où $\hat{\pi}_h$ est la probabilité empirique spécifique à l'intervalle h , c'est-à-dire

$$\hat{\pi}_h = \frac{1}{n} \sum_{i=1}^n \delta_i(h) = \frac{n_h}{n} \quad (2)$$

où $\delta_i(h) = 1$ si $r_i \in h$, 0 sinon. On pose alors

$$\begin{aligned} w_i = w_h &= \frac{\pi}{\hat{\pi}_h} \\ &= \frac{n}{Hn_h} \end{aligned} \quad (3)$$

pour $r_i \in h$. En utilisant ces poids, la *distribution pondérée empirique* des ratios de prix r est donnée par

$$\hat{F}_{n,w}(y) = \frac{\sum_{i=1}^n w_i I_i(y)}{\sum_{i=1}^n w_i}. \quad (4)$$

En notant que $\sum_{i=1}^n w_i = \sum_{h=1}^H \sum_{i=1}^{n_h} n / Hn_h = n$, on obtient finalement

$$\hat{F}_{n,w}(y) = \frac{\sum_{i=1}^n w_i I_i(y)}{n}. \quad (5)$$

Puisque $n_h = \sum_{i=1}^n \delta_i(h)$, on a

$$\begin{aligned} \hat{F}_{n,w}(y) &= \frac{\sum_{i=1}^n w_i I_i(y)}{n} \\ &= \frac{1}{H} \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^n \delta_i(h) I_i(y) \\ &= \frac{1}{H} \sum_{h=1}^H \frac{\sum_{i=1}^n \delta_i(h) I_i(y)}{\sum_{i=1}^n \delta_i(h)} \\ &= \frac{1}{H} \sum_{h=1}^H \hat{F}_n(y|h). \end{aligned} \quad (6)$$

Lorsque $n \rightarrow \infty$, on a $(1/n) \sum_{i=1}^n \delta_i(h) I_i(y) \xrightarrow{P} P(r \in h, r \leq y)$ et $(1/n) \sum_{i=1}^n \delta_i(h) \xrightarrow{P} P(r \in h)$. Donc, lorsque $n \rightarrow \infty$,

$$\begin{aligned} \hat{F}_n(y|h) &\xrightarrow{P} \frac{P(r \in h, r \leq y)}{P(r \in h)} \\ &= P(r \leq y | r \in h) = F(y|h) \end{aligned} \quad (7)$$

où $F(y|h)$ est la distribution des ratios de prix r à l'intérieur de l'intervalle h .

Pour n suffisamment grand, on devrait être en mesure de déterminer H de manière à construire les intervalles h pour que $\hat{F}_n(y|h)$ soit approximativement uniformément distribuée, $h = 1, \dots, H$. En d'autres mots, lorsque $n \rightarrow \infty$, pour H suffisamment grand, $F(y|h)$ devrait suivre une distribution uniforme sur l'intervalle h . Notons qu'un tel argument a été utilisé par Dalenius et Hodges (1959) dans un contexte de stratification optimale. Dans ce cas, la distribution $F(y|h)$ est donnée par

$$F(y|h) = \begin{cases} 0 & \text{pour } y \leq (h-1)H^{-1} \\ Hy - h + 1 & \text{pour } (h-1)H^{-1} < y \leq hH^{-1} \\ 1 & \text{pour } y > hH^{-1}. \end{cases} \quad (8)$$

Puisque $F(y) = \sum_{h=1}^H F(y|h) / H$, on a $F(y) = y$, ce qui correspond à la distribution uniforme. On conclue donc que pour n suffisamment grand, l'utilisation de la pondération (3) devrait rendre la distribution empirique pondérée $\hat{F}_{n,w}(y)$ donnée par (5) approximativement uniformément distribuée.

Des simulations Monte-Carlo ont démontré que les estimations résultant d'un échantillon non aléatoire pouvaient être améliorées en utilisant cette approche. Ses principaux avantages résident dans :

- sa simplicité;
- le fait qu'elle peut être introduite *ex ante*, c'est-à-dire avant l'introduction du modèle économétrique comme tel. Par conséquent, elle ne requiert pas d'hypothèses de travail fortes.

Si on reprend l'histogramme exposé plus haut et divisons l'échantillon en $H = 5$ intervalles d'une largeur de 0,2 avec une valeur de $\pi = 1/5 = 0,2$, on obtient alors l'histogramme suivant qui a été pondéré *ex post*.

4. Application

Nous allons illustrer la démarche à partir d'un exemple tiré du Kelly Blue Book, une source d'information largement utilisée pour l'estimation de la dépréciation des voitures. Le tableau 1 présente les prix de deux modèles de voitures pour différents âges entre 1 et 18 ans. Pour chaque voiture, on dispose donc d'un échantillon de $n = 18$ unités. Les prix sont exprimés en valeur relative par rapport à un modèle neuf. Il est en outre nécessaire d'ajuster les ratios pour tenir compte de la probabilité de survie à chacun de ces âges. Pour chaque voiture, le ratio final utilisé r_i de l'année i est donc construit à partir du produit du ratio des prix par la probabilité de survie.

On s'intéresse au taux de dépréciation moyen $\bar{\tau}$ pour chaque voiture. Ce dernier pourrait être estimé à partir d'une régression des prix (ou d'une fonction de ces derniers) par rapport à l'âge (ou d'une fonction de l'âge). Toutefois, si on présume que le taux est constant et de forme géométrique, on a la relation $r_i = 1 - \bar{\tau}^i$, où r_i est le prix relatif selon l'âge i . Dans ce cas, un taux $\hat{\tau}_i$ peut être estimé à chaque âge i par $\hat{\tau}_i = 1 - r_i^{1/i}$. Une estimation du taux de dépréciation moyen est alors produite à partir de la moyenne pour tous les âges, c'est-à-dire $\hat{\tau} = \sum_{i=1}^{18} \hat{\tau}_i / 18$.

Dans l'exemple ci-dessus, on constate que les taux de dépréciation $\hat{\tau}_i$ varient selon la fourchette d'âge et qu'ils ont tendance à augmenter avec l'âge. Par ailleurs, le fait que

l'on utilise une simple moyenne des âges dans le calcul de $\hat{\tau}$ revient à accorder de façon implicite le même poids à chacun des âges. Mais il est bien évident que ce ne serait pas la distribution que l'on obtiendrait si on tirait un échantillon aléatoire des voitures en services. La figure ci-dessous présente la distribution des cellules de prix entre les ratios de 0 et 1.

La technique de repondération consiste simplement à imposer un poids égal à chacune des fourchettes de prix relatifs. Dans cet exemple, les $n = 18$ âges sont répartis en $H = 7$ classes, ce qui répartit les âges en 18/7 à chacune d'entre elles (en réalité, la structure des cellules a été configurée pour 8 classes mais la dernière est toujours vide). Comme mentionné à la section 3, les poids individuels w_i de chaque âge i sont construits selon (3), c'est-à-dire en divisant 18/7 par le nombre d'observations qui se trouvent dans chaque classe, sauf pour les cellules vides dont le poids demeure nul. Le tableau 2 présente les résultats et l'impact de la repondération sur les statistiques dérivées.

Cet exemple illustre bien les problèmes de biais d'agrégation typique des régressions estimées à partir d'agrégats économiques, sans tenir compte de la distribution réelle des unités au niveau micro. Ainsi, il est assez évident que les unités de 17 et 18 ans ne sauraient avoir le même poids de régression que celles de 1 an puisque le risque de perte à 1 an concerne pratiquement toutes les voitures qui seront mises en circulation alors que très peu d'entre elles seront exposées au risque de perte de valeur à des âges avancés. Il en résulte que l'estimation non pondérée, dans cet exemple, introduit une surestimation du taux de dépréciation de l'ordre de 15 %.

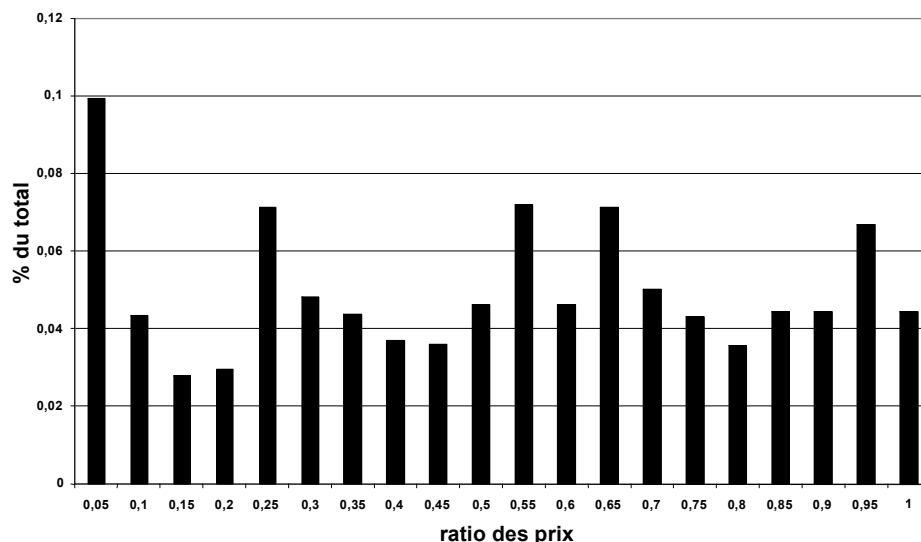


Figure 2 Distribution pondérée des observations selon le ratio des prix, usines de fabrication pondération *ex post*

Tableau 1 Prix relatifs de deux modèles de voitures selon le Kelly Blue Book et Taux de dépréciation moyen avant repondération

Année	Pr ($t > S$)*	Prix relatifs				Taux moyens de dépréciation	
		Excluant mises au rancard		Incluant mises au rancard		Incluant mises au rancard	
		Buick	Chrysler	Buick	Chrysler	Buick	Chrysler
1	0,9988	0,8633	0,8257	0,8622	0,8246	0,1367	0,1743
2	0,9901	0,7435	0,6801	0,7361	0,6734	0,1377	0,1753
3	0,9666	0,6410	0,5608	0,6195	0,5420	0,1378	0,1754
4	0,9220	0,5523	0,4621	0,5092	0,4261	0,1379	0,1755
5	0,8526	0,4740	0,3794	0,4042	0,3234	0,1387	0,1762
6	0,7582	0,4034	0,3087	0,3058	0,2341	0,1404	0,1779
7	0,6433	0,3391	0,2482	0,2181	0,1597	0,1432	0,1805
8	0,5164	0,2790	0,1953	0,1441	0,1009	0,1475	0,1846
9	0,3892	0,2227	0,1491	0,0867	0,0580	0,1537	0,1906
10	0,2731	0,1639	0,1050	0,0448	0,0287	0,1654	0,2018
11	0,1770	0,1261	0,0772	0,0223	0,0137	0,1716	0,2077
12	0,1051	0,0892	0,0523	0,0094	0,0055	0,1824	0,2180
13	0,0567	0,0614	0,0344	0,0035	0,0019	0,1932	0,2284
14	0,0276	0,0441	0,0236	0,0012	0,0007	0,1999	0,2347
15	0,0120	0,0320	0,0164	0,0004	0,0002	0,2050	0,2396
16	0,0046	0,0190	0,0093	0,0001	0,0000	0,2194	0,2534
17	0,0016	0,0088	0,0041	0,0000	0,0000	0,2432	0,2761
18	0,0005	0,0051	0,0023	0,0000	0,0000	0,2542	0,2867
						<i>Moyenne</i>	
						0,1727	0,2087

* Probabilité de Survie selon les estimations de la Division des études et de l'analyse micro-économique de Statistique Canada.

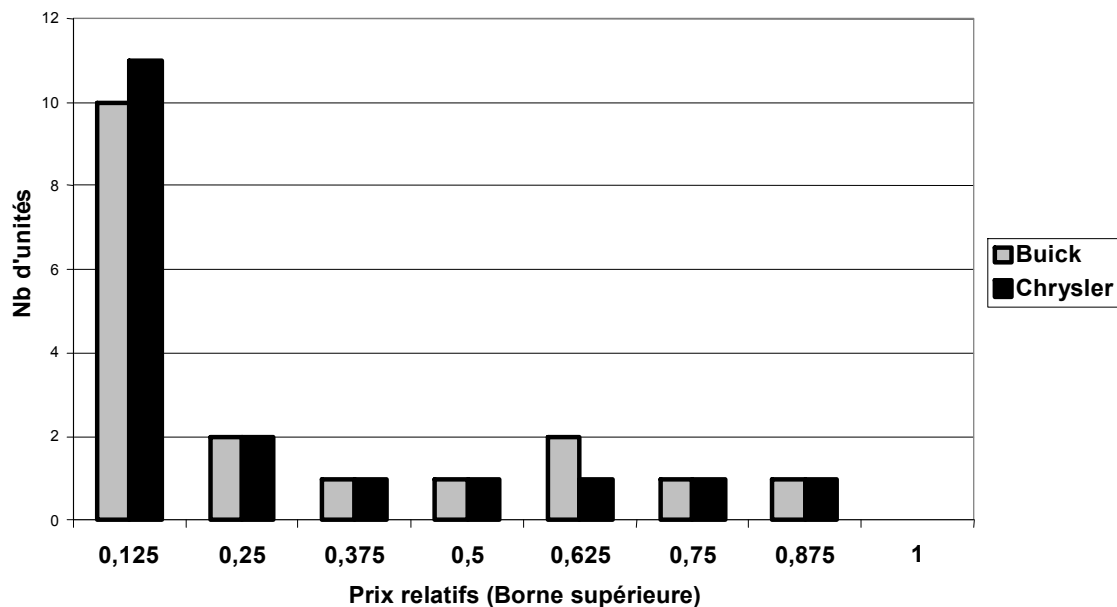


Figure 3 Distribution des cellules utilisées pour l'estimation du taux de dépréciation moyen selon les données du Kelly Blue Book avant repondération (Total = 18)

Tableau 2 Prix relatifs de deux modèles de voitures selon le Kelly Blue Book et taux de dépréciation moyen après repondération

Année	Prix relatifs		Taux moyens de dépréciation		Poids Ex post	
	Incluant mises au rancard		Incluant mises au rancard		Buick	Chrysler
	<i>Buick</i>	<i>Chrysler</i>	<i>Buick</i>	<i>Chrysler</i>		
1	0,8622	0,8246	0,1367	0,1743	2,5714	2,5714
2	0,7361	0,6734	0,1377	0,1753	2,5714	2,5714
3	0,6195	0,5420	0,1378	0,1754	1,2857	2,5714
4	0,5092	0,4261	0,1379	0,1755	1,2857	2,5714
5	0,4042	0,3234	0,1387	0,1762	2,5714	2,5714
6	0,3058	0,2341	0,1404	0,1779	2,5714	1,2857
7	0,2181	0,1597	0,1432	0,1805	1,2857	1,2857
8	0,1441	0,1009	0,1475	0,1846	1,2857	0,2338
9	0,0867	0,0580	0,1537	0,1906	0,2571	0,2338
10	0,0448	0,0287	0,1654	0,2018	0,2571	0,2338
11	0,0223	0,0137	0,1716	0,2077	0,2571	0,2338
12	0,0094	0,0055	0,1824	0,2180	0,2571	0,2338
13	0,0035	0,0019	0,1932	0,2284	0,2571	0,2338
14	0,0012	0,0007	0,1999	0,2347	0,2571	0,2338
15	0,0004	0,0002	0,2050	0,2396	0,2571	0,2338
16	0,0001	0,0000	0,2194	0,2534	0,2571	0,2338
17	0,0000	0,0000	0,2432	0,2761	0,2571	0,2338
18	0,0000	0,0000	0,2542	0,2867	0,2571	0,2338
				<i>Moyenne pondérée</i>	0,1479	0,1836

Remerciements

Les auteurs tiennent à remercier grandement l'arbitre anonyme de *Techniques d'enquête* qui, par ses judicieux commentaires, a contribué à améliorer la qualité de l'article.

Bibliographie

- Bickel, P.J., et Doksum, K.A. (1977). *Mathematical Statistics*, Holden-Day, Oakland, CA.
- Dalenius, T., et Hodges, J.L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54, 88-101.
- Davidson, R., et MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, N.Y.

Gellatly, G., Tanguay, M. et Yan, B. (2002). An alternative methodology for estimating economic depreciation: New results using a survival model. Dans *Productivity Growth in Canada-2002*, Statistique Canada. #15-204-XPE.

Greene, W.H. (1993). *Econometric Analysis*. Deuxième édition, Prentice Hall, Englewood Cliffs, N.J.

Hulten, C.R., et Wykoff, F.C. (1981). The measurement of economic depreciation. Dans *Depreciation, Inflation, and the Taxation of Income from Capital*, (Éd. C.R. Hulten). The Urban Institute Press, Washington, D.C, 81-125.

Lancaster, T. (1985). Generalized residuals and heterogeneous duration model: With applications to the weibull model. *Journal of Econometrics*, 28, 155-69.

Ross, S.M. (2002). *Introduction to Probability Models*, 8^{ième} Édition, Academic Press.