

Révision des équations hédoniques relatives aux prix des ordinateurs

Fred Barzyk

*Division des prix
Statistique Canada*

Immeuble Jean-Talon, 13-C2, Ottawa K1A 0T6

Télécopieur : (613) 951-2848

Téléphone : (613) 951-2493 et 951-3834

Courrier électronique : fred.barzyk@statcan.ca

novembre 1999

Révision des équations hédoniques relatives aux prix des ordinateurs

1. Introduction¹

Notre propos sera de décrire sommairement les travaux récents à la Division des prix de Statistique Canada sur l'estimation d'un nouvel ensemble d'équations hédoniques devant servir au traitement des variations de qualité des ordinateurs². En général, ces équations visent à expliquer les mouvements de la variable d'un bien ou d'un service en les rapportant aux variations des caractéristiques de ce bien ou de ce service. Dans ce cas, on se trouve à mesurer le rapport entre le prix et les diverses caractéristiques d'un ordinateur, qu'il s'agisse de la mémoire vive, de la vitesse du processeur central, de la capacité du disque dur ou de diverses variables qualitatives (fournisseur, présence d'un disque compact ou d'une unité de disques compacts, d'un modem, etc.). On dégage ces estimations pour les ordinateurs de bureau et les ordinateurs portatifs³.

Le principal problème que pose l'estimation d'équations de ce genre réside dans le choix d'une forme fonctionnelle, car celui-ci déterminera la qualité des résultats (pertinence, exactitude, fiabilité, vraisemblance, etc.). Bien que la presque totalité des études consacrées aux équations hédoniques dans le cas des ordinateurs aient limité ce choix aux modèles linéaire, semi-logarithmique et bilogarithmique, Triplett (1987) préconise de ne pas oublier les formes fonctionnelles non linéaires, puisque certaines d'entre elles sont économiquement plus significatives.

Jusqu'à-là, la Division des prix avait procédé à trois estimations des équations hédoniques des ordinateurs (voir le tableau 1). La première de 1989-1990 et la deuxième de 1991 utilisaient toutes deux les données des rapports « Data Pro », base de données américaine. Les résultats de ces études privilégiaient les modèles semi-logarithmique et linéaire pour tous les ordinateurs de bureau (16 et 32 bits et ensemble). Une troisième estimation datant de 1996 a fait intervenir les données du guide des acquisitions de microinformatique de Travaux publics et Services gouvernementaux Canada. Cette étude a fait adopter le modèle semi-logarithmique de correction de qualité pour les ordinateurs de bureau, mais on n'a pas trouvé de modèle convenant aux ordinateurs portatifs.

Cette dernière révision des équations hédoniques relatives aux ordinateurs livre des résultats plus utiles et plus intéressants que les deux autres études. Ils sont plus utiles parce que les données exploitées sont d'une bien meilleure qualité que celles des analyses précédentes; ils sont plus intéressants parce qu'on a expérimenté non seulement les modèles linéaire, semi-logarithmique et bilogarithmique, mais aussi le modèle non linéaire (pour les paramètres). Voici comment se divise le reste de notre exposé : description des données, aperçu des modèles essayés, méthodologie, résultats obtenus et conclusion.

2. Données et variables

La principale source de données aux fins de l'exercice a été les rapports mensuels sur les prix des ordinateurs fournis par IDC Canada (International Data Corporation) à la Division des prix. Par rapport aux

¹ Ce document ne doit pas être cité sans la permission expresse de l'auteur. Celui-ci désire remercier de leurs observations et de leurs suggestions intéressantes Andy Baldwin et Robin Lowe, de la Division des prix, ainsi qu'Erwin Diewert, Jack Triplett et Ralph Turvey. Les vues exprimées sont celles de l'auteur, et non pas nécessairement celles de la Division des prix ni de Statistique Canada.

² Notre propos est de décrire la méthodologie et les résultats de la réestimation des équations, et non pas de nous étendre sur la théorie hédoniste. Le lecteur pourra se faire une idée des études spécialisées dans ce domaine en consultant le chapitre 4 de *The Practice of Econometrics, Classic and Contemporary* de Ernst R. Berndt (1991).

³ L'ordinateur portable se définit comme un ordinateur de 2 à 8 livres exclusivement (*Le Dictionnaire de l'informatique* Microsoft Press, p. 272-273).

études antérieures, le nombre d'observations est bien plus élevé et les données, plus utiles, étant les données mêmes qui servent à la production des séries de l'indice des prix des ordinateurs⁴.

Une autre source de données a été le site Internet « *The CPU Scorecard* » (voir l'annexe A), qui présente un classement de rendement des processeurs centraux selon des tests comparatifs à l'échelle de l'industrie. Les cotes en question ont servi à élaborer la variable SCORE, dont l'utilité à cet égard réside dans la possibilité d'apporter des corrections de qualité là où, dans les modèles, il y a changement de génération de processeurs centraux⁵.

2.1 - Ordinateurs de bureau

Les données relatives aux ordinateurs de bureau ont consisté en 576 observations. Le tableau 2 présente les variables jugées significatives pour l'explication des différences de prix. Certaines variables sont quantitatives (c'est notamment le cas de la variable HD, qui mesure en mégaoctets la taille des disques durs); d'autres sont dichotomiques et donc traitées comme variables fictives (c'est notamment le cas de la variable MODEM pour la présence ou l'absence d'un tel appareil). Enfin, nous avons élaboré la variable BQUAL en fonction des rapports de qualité de l'industrie qui font un classement de fiabilité et de service de diverses sociétés d'informatique⁶.

2.2 - Ordinateurs portatifs

Pour l'analyse des ordinateurs portatifs, nous disposons de 174 observations pour huit variables (voir le tableau 3). Certaines de ces variables correspondent à celles de l'analyse relative aux ordinateurs de bureau (RAM, HD, SCORE et XCACHE) et d'autres intéressent les seuls ordinateurs portatifs, à savoir WGHT, DSIZE et DQUAL. Avec la variable HIEND, nous distinguons le fournisseur IBM des autres dans l'ensemble de données, conscients de ce que la société IBM destine habituellement les ordinateurs portatifs au secteur haut de gamme à des prix supérieurs à ceux de la concurrence⁷.

3. Méthodologie

Pour pouvoir produire un nouveau jeu d'estimations pour les équations hédoniques, nous devons d'abord résoudre deux questions, celles de la détermination des variables explicatives et du choix d'une forme fonctionnelle. L'une ne va pas sans l'autre et, souvent, on les résout de concert⁸. Dans le second cas, il importe de constater que, s'il existe plusieurs choix de formes fonctionnelles, celles-ci se rangent en gros

⁴ Comme les données disponibles étaient restreintes pour les deux premières études, on a employé dans le premier cas des données américaines et dans le second celles des offres permanentes de vente au gouvernement fédéral du Canada. Dans la présente étude, nous avons cependant utilisé les prix de vente réels dans toutes nos observations. Ce sont des prix représentatifs de ceux qui se pratiquent effectivement sur le marché.

⁵ L'inclusion de cette variable nous permet de faire porter l'estimation d'un coup sur tout l'ensemble de données plutôt que d'avoir à subdiviser les observations relatives aux mêmes processeurs centraux. Par le passé, les changements de modèles d'unités centrales ne se retrouvaient pas dans les corrections de qualité, parce qu'ils n'étaient pas représentés d'une manière significative. Ainsi, l'estimation de 1990 a séparé les processeurs de 16 et de 32 bits et celle de 1996 a visé la seule grande catégorie des ordinateurs Pentium.

⁶ Voir *PC World*, mars 1996, p. 195-197.

⁷ Ce groupe d'entreprises diffère de celui auquel nous nous reportons dans le cas des ordinateurs de bureau où nous subdivisons en fonction de la qualité perçue des gammes d'ordinateurs. Sur le marché des ordinateurs portatifs, il n'y a guère de différence de qualité et de service dans l'ensemble entre les fabricants retenus aux fins de l'analyse (voir *PC World*, novembre 1998, p. 110-116).

⁸ La contribution apportée à la puissance explicative globale des variables indépendantes a été considérée pour toutes les variables et les formes fonctionnelles soumises à l'analyse.

dans deux catégories, celles des formes linéaires et non linéaires⁹. Au départ, le modèle général se présente sous la forme suivante :

$$P_i = f(X_i, Z_i),$$

où P est le prix de l'ordinateur i , X le jeu de variables explicatives en linéarité ou non et Z le jeu de variables explicatives en linéarité par définition (variables fictives et variables à valeurs non positives). Pour notre propos, nous avons considéré plusieurs fonctions linéaires et loglinéaires (semi-logarithmiques et bilogarithmiques) avec les modèles par transformation de Box-Cox et le modèle non linéaire proposé par Triplett (1987).

3.1 - Modèles linéaire et loglinéaire

Les formes fonctionnelles linéaires, semi-logarithmiques et bilogarithmiques soumises à l'évaluation figurent ci-après :

$$\text{Modèle 1. Linéaire - } P_i = \beta_0 + \sum_{k=1}^n \beta_k X_{ik} + \sum_{j=1}^l \gamma_j Z_{ij} + e_i, \text{ où } e_i \sim \text{NID}(0, \sigma^2) \quad (1)$$

$$\text{Modèle 2. Semi-logarithmique - } \ln P_i = \beta_0 + \sum_{k=1}^n \beta_k X_{ik} + \sum_{j=1}^l \gamma_j Z_{ij} + e_i, \text{ où } e_i \sim \text{NID}(0, \sigma^2) \quad (2)$$

$$\text{Modèle 3. Bilogarithmique - } \ln P_i = \beta_0 + \sum_{k=1}^n \beta_k \ln X_{ik} + \sum_{j=1}^l \gamma_j Z_{ij} + e_i, \text{ où } e_i \sim \text{NID}(0, \sigma^2) \quad (3)$$

Dans la version bilogarithmique, le nombre de versions susceptibles d'être estimées dépend de k^{10} .

3.2 - Modèle de Box-Cox

Le choix portant sur les formes fonctionnelles linéaires et loglinéaires est plutôt limité et d'autres types de transformations sont disponibles. Un type général est la transformation de Box-Cox, qui offre plus de souplesse¹¹. Celle-ci prend la forme suivante :

$$y^{(\lambda)} = \frac{y^\lambda - 1}{\lambda},$$

où y est la variable à considérer. C'est là un modèle fort utile, parce qu'il embrasse une grande diversité de transformations. D'abord, les valeurs de 1 et 0 pour λ donnent des valeurs linéaires et logarithmiques de la

⁹ Qu'on emploie les termes « linéaire » et « non linéaire » à propos de l'estimation d'une forme fonctionnelle peut prêter à confusion. En ce sens, « linéaire » veut dire une linéarité dans les paramètres à estimer. Les transformations logarithmiques (ou loglinéaires), réciproques et exponentielles sont de simples transformations du modèle linéaire général. Ainsi, une fonction bilogarithmique, qui est non linéaire dans sa forme fonctionnelle, le demeure cependant dans ses paramètres et peut donc être estimée par les moindres carrés ordinaires. Des formes fonctionnelles qui sont non linéaires dans leurs paramètres comme dans le modèle d'« identification de t » demandent une autre méthode d'estimation (par les moindres carrés non linéaires, par exemple). Pour un examen détaillé de la question, voir Greene (1990), p. 239-276.

¹⁰ Dans l'analyse, nous présentons des estimations pour deux versions du modèle (3), à savoir le modèle général 3A et le modèle où RAM est la seule variable explicative transformée (3B).

¹¹ Voir Box et Cox (1964).

variable y (modèles 2 et 3) et, si λ se situe entre 0 et 1, les valeurs obtenues sont non linéaires¹². Voici ce que nous obtenons lorsque nous appliquons la transformation de Box-Cox à (1) :

$$\text{Modèle 4. Box-Cox - } P_i^{(\lambda)} = \beta_0 + \sum_{k=1}^n \beta_k X_{ik}^{(\lambda)} + \sum_{j=1}^l \gamma_j Z_{ij} + e_i, \quad \text{où } e_i \sim \text{NID}(0, \sigma^2) \quad (4)$$

C'est le modèle de Box-Cox classique¹³. Il a plusieurs usages : il aide à trouver ou à distinguer des formes fonctionnelles (Darnell, (1994)) et, pour cette raison, nous avons voulu inclure la transformation de Box-Cox dans la présente analyse.

3.3 - Modèle non linéaire

Une seconde catégorie de fonctions est la catégorie non linéaire des fonctions hédoniques, dont la fonction d'« identification de t », proposée par Triplett (1987) et soumise à la présente analyse, constitue un cas d'espèce. Le modèle d'« identification de t » que nous étudions est le suivant :

$$\text{Modèle 5. Non linéaire - } \ln p_i = \beta_0 + \sum_{k=1}^n \beta_k X_{ik}^{\alpha_k} + \sum_{j=1}^l \gamma_j Z_{ij} + e_i, \quad \text{où } e_i \sim \text{NID}(0, \sigma^2) \quad (5)$$

où β_k et α_k sont > 0 pour tous les k et où Z est une fois de plus un jeu de variables linéaires¹⁴.

Théoriquement, cette forme fonctionnelle est des plus attrayantes, car elle permet des courbes hédoniques qui s'excurvent de l'origine, alors que les courbes correspondantes des fonctions linéaire et semi-logarithmique sont linéaires et que celles de la fonction bilogarithmique s'incurvent en réalité.

4. Résultats

4.1 Ordinateurs de bureau

Dans le cas des ordinateurs de bureau, la question du choix d'une forme fonctionnelle s'appliquait aux seules variables PRICE, RAM, HD et SCORE. Les quatre étaient les seules variables non dichotomiques à être toujours positives¹⁵. Les autres figurent dans les équations comme variables linéaires. Nous avons évalué plusieurs formes fonctionnelles en adoptant pour critères d'appréciation les signes et les valeurs des coefficients et les statistiques d'ajustement en comparaison¹⁶.

Pour les modèles linéaire et loglinéaire, les statistiques d'ajustement retenues ont été les fonctions logarithmiques de vraisemblance (avec correction pour les modèles semi-logarithmique et bilogarithmique) et la statistique de régression artificielle dite de double longueur (« Double-Length Artificial Regression

¹² Le modèle classique de Box-Cox présente aussi des inconvénients par une restriction des valeurs possibles de la variable dépendante. Ajoutons que λ influe sur les propriétés des résidus, ainsi que sur la forme fonctionnelle de l'équation (Davidson et Mackinnon (1993), p. 507-510). Voilà pourquoi nous nous sommes uniquement guidés sur les résultats des estimations de Box-Cox pour établir si il y avait des données suffisantes à l'appui de l'application d'un modèle non linéaire.

¹³ Voir Davidson et MacKinnon (1993), p. 480-488.

¹⁴ Le modèle diffère légèrement de celui de Triplett par l'inclusion de variables linéaires.

¹⁵ La variable XCACHE contenait plusieurs valeurs 0 dans le cas des ordinateurs de bureau.

¹⁶ De plus, nous avons évalué les résidus de toutes les estimations relatives aux ordinateurs de bureau et aux ordinateurs portatifs et vérifié la présence d'une multicollinéarité et d'une hétéroscédasticité pour conclure que ni l'une ni l'autre n'était présente à un degré significatif.

Statistic » ou DLR) proposée par Davidson et MacKinnon (1993) comme test de forme fonctionnelle¹⁷. Des trois formes fonctionnelles soumises à l'analyse (modèles linéaire, semi-logarithmique et bilogarithmique), la bilogarithmique avec le logarithme de la variable indépendante RAM offre le meilleur ajustement (voir les tableaux 4 à 7). Dans le cas du modèle linéaire, nous avons obtenu une valeur négative pour CD, et la fonction logarithmique de vraisemblance donnait la plus petite valeur pour les trois modèles. Précisons en outre que la statistique DLR rejette fortement l'hypothèse nulle de linéarité par rapport au modèle général de Box-Cox. Pour le modèle semi-logarithmique, elle s'améliore très nettement et la statistique DLR ne rejette pas l'hypothèse nulle de la transformation de la variable indépendante. De même, il y a non-rejet de l'hypothèse nulle de $\lambda=0$ pour le modèle bilogarithmique lorsque seul le logarithme de RAM est considéré. Toutefois, lorsque nous prenons le logarithme de RAM, HD et SCORE, la DLR rejetait l'hypothèse nulle. De ces résultats, il semble ressortir que toutes les variables indépendantes sauf une, RAM, devraient figurer dans l'équation comme variables linéaires; les résultats indiqueraient aussi que la variable dépendante PRICE devrait figurer sous sa forme logarithmique.

Comme les modèles linéaire et loglinéaire ne sont que deux cas d'espèce de la transformation de Box-Cox, nous avons estimé d'autres formes du modèle général du même nom par la méthode des moindres carrés non linéaires pour obtenir la valeur de λ en (4). Les résultats figurent au tableau 8. Nous avons analysé plusieurs modèles par la transformation de Box-Cox d'abord pour le cas restreint de λ égale 0 pour P_i et pour le cas non restreint où il y avait estimation de λ tant pour P_i que pour X_i . Des modèles du cas restreint (où $\lambda = 0$ pour P), le modèle où seule la variable RAM subit la transformation de Box-Cox offre le meilleur ajustement. La valeur estimée de λ est 0,34, indice d'une relation non linéaire possible du log de P et de RAM¹⁸.

Pour la fonction non linéaire d'« identification de t » de Triplett (1987), nous avons vérifié plusieurs versions de 5 pour trouver le modèle au meilleur ajustement (voir les tableaux 9 à 11). À en juger par les signes des coefficients, les tests F, les rapports de vraisemblances et les résultats de convergence, la version avec RAM pour seule variable non linéaire offre le meilleur ajustement¹⁹. La valeur estimée de α est 0,34186 et la condition α (comme d'ailleurs tous les autres coefficients) > 0 est donc remplie.

¹⁷ Pour l'essentiel, la statistique DLR est une statistique chi-carré à un degré de liberté et permet de choisir entre les modèles linéaire, semi-logarithmique et bilogarithmique par rapprochement analytique avec le modèle général de Box-Cox. Il peut se présenter des cas où les modèles linéaire et loglinéaire et le modèle général pourront sembler acceptables à la fois, auquel cas la statistique DLR servira non seulement à comparer les valeurs des fonctions logarithmiques de vraisemblance, mais aussi à faire un choix entre les modèles linéaire, semi-logarithmique et bilogarithmique. S'il s'agit d'un modèle linéaire, nous évaluons l'hypothèse nulle que $\lambda = 1$ et, s'il s'agit des modèles semi-logarithmique et bilogarithmique, l'hypothèse est que $\lambda = 0$. Voir Davidson et MacKinnon (1993), p. 502-504.

¹⁸ Il faut d'autant plus rejeter le modèle linéaire que λ est estimé tant pour P que pour RAM (cas non restreint) et que les résultats sont plus robustes. La fonction logarithmique de vraisemblance s'améliore sensiblement, mais l'interprétation de ces résultats doit demeurer prudente, puisque, dans le modèle classique de Box-Cox, λ influe sur les résidus aussi bien que sur la forme fonctionnelle de l'équation (voir Davidson et MacKinnon (1993), p. 507-510.). Ajoutons que, avec le modèle de Box-Cox, on peut théoriquement estimer différentes valeurs de λ pour *chaque* variable, quoique des problèmes puissent s'ensuivre pour l'application effective de la méthode d'estimation (voir White (1993), p. 152). Lorsque nous avons appliqué la méthode, il n'y avait pas de convergence pour le modèle.

¹⁹ Les modèles où la variable HD figurait sous une forme non linéaire n'étaient pas en convergence pour la plupart, sauf lorsqu'on considérait RAM et HD en combinaison. Vérification faite, il y avait un maximum local plutôt que global, ce qui a fait retirer ce modèle (voir White, Judge et coll. (1988), p. 505).

4.1 - Comparaison des modèles estimés

Les modèles semi-logarithmique (2), bilogarithmique (3B) et non linéaire (5B) paraissent acceptables, bien que, en matière d'ajustement, le dernier semble légèrement l'emporter²⁰. Toutefois, un critère très important de distinction de ces formes fonctionnelles est celui du rendement du modèle dans l'évaluation de changements effectifs de qualité au niveau des valeurs des coefficients estimés. Si les coefficients d'un modèle donné font voir une variation de prix hédoniques en divergence avec ce qu'on peut observer dans l'industrie, le modèle en question doit être délaissé au profit d'un modèle plus réaliste. Si on constate que, en doublant la quantité de mémoire vive, on double le prix d'un ordinateur, le modèle appliqué doit être remplacé par un modèle qui appréhende les variations de prix avec plus de réalisme²¹.

Le tableau 12 compare les prix obtenus par les trois modèles estimés. La comparaison porte sur une évolution type des modèles avec trois changements courants, à savoir un accroissement de 64 à 128 Mo de la mémoire vive (RAM), une augmentation de taille du disque dur (HD) de 8,4 à 12,9 Go et une conversion d'unité centrale (CPU) ou de classe de processeurs (SCORE) de Pentium II 300 à Pentium II 400. Avant ces mesures de mise à niveau, le prix hypothétique de l'ordinateur était de 2 000 dollars canadiens²².

Les résultats montrent clairement que le modèle non linéaire ne livre pas des estimations raisonnables si on le compare aux deux autres modèles. À l'aide du modèle 5B, nous avons obtenu un nouveau prix de 6 060,18 \$ et, par les modèles 2 et 3B, des prix respectifs de 2 945,52 \$ et 2 905,16 \$. Ces deux derniers prix sont nettement plus réalistes. Des deux modèles, le 3B est le meilleur d'après tous les critères mentionnés.

4.3 - Ordinateurs portatifs

Nous avons obtenu des résultats semblables pour les ordinateurs portatifs : le modèle bilogarithmique offrait le meilleur ajustement lorsque seul le logarithme de RAM était inclus (voir les tableaux 13 à 16). D'après les valeurs de la fonction logarithmique de vraisemblance estimée et la statistique DLR, le modèle linéaire peut nettement être rejeté, ce qui laisse à considérer le modèle semi-logarithmique et les deux versions bilogarithmiques 3A et 3B²³. À examiner de plus près les fonctions logarithmiques de vraisemblance respectives, le modèle contenant le logarithme de RAM (3B) l'emportait pour l'ajustement²⁴.

Les résultats des estimations de Box-Cox (voir le tableau 17) privilégient aussi le modèle bilogarithmique 3B. Contrairement au cas des ordinateurs de bureau, il n'y a guère d'indications à l'appui de l'application d'un modèle non linéaire ($0 < \lambda < 1$). Pour le cas restreint $\lambda = 0$ pour P, la valeur estimée de λ est également nulle pour RAM. Nulle coïncidence donc que de piètres résultats aient été obtenus pour le

²⁰ Nous avons respectivement obtenu les valeurs 115,0178 et 122,8366 pour les fonctions logarithmiques de vraisemblance de 2 et 3B; par les moindres carrés non linéaires, nous avons estimé que les modèles en question étaient comparables au modèle non linéaire 5B.

²¹ À en croire les tendances actuelles dans l'industrie, le prix de la mémoire vive en dollars canadiens avoisine les 3 \$ le mégaoctet sans les frais d'installation ni les taxes.

²² Nous avons calculé le prix après mise à niveau en multipliant les variations propres à chaque mesure de mise à niveau par les coefficients correspondants, en sommant les valeurs et en prenant l'antilogarithme pour dégager un facteur de variation de prix. Nous avons alors multiplié le prix d'origine par ce facteur pour obtenir le nouveau prix.

²³ Pour les modèles linéaire, semi-logarithmique et bilogarithmique, tous les coefficients estimés ont le bon signe, y compris celui de WGHT, qui est en rapport inverse avec le coefficient de PRICE.

²⁴ En fait, si on rapproche les résultats des modèles 3A et 3B, on constate que l'utilisation de la forme logarithmique des variables HD, SCORE, XCACHE, WGHT ET DSIZE n'améliore pas du tout l'ajustement. Le principe de parcimonie nous fera choisir le modèle le plus simple (3B).

modèle d'« identification de t », puisque nous n'avons pu calculer une valeur unique pour tous les coefficients²⁵.

4.4 - Comparaison des modèles estimés dans le cas des ordinateurs portatifs

Comme pour les ordinateurs de bureau, on peut pousser l'évaluation des modèles en question par une comparaison de rendement lorsque changent les caractéristiques d'un ordinateur portatif. Les résultats d'une telle évolution hypothétique figurent au tableau 18. La comparaison a porté sur un accroissement de mémoire vive (RAM) de 32 à 64 Mo, une augmentation de taille du disque dur (HD) de 3,2 à 4,0 Go et une conversion d'unité centrale (CPU) ou de classe de processeurs (SCORE) de Pentium 233 MMX à Pentium II 266. Le prix hypothétique de l'ordinateur avant ces mesures de mise à niveau était de 2 900 dollars canadiens²⁶.

On ne s'étonnera pas que les variations de prix déterminées par les versions bilogarithmiques 3A et 3B soient convergentes (avec une différence approximative de 98 \$). La variation selon le modèle semi-logarithmique 2 est sensiblement supérieure à celle du modèle bilogarithmique 3A (différence approximative de 219 \$) et semblable à celle du modèle 3B (différence approximative de 121 \$). Comme le modèle contenant le logarithme de RAM (3B) offre le meilleur ajustement, il est préférable au modèle semi-logarithmique 2.

5. Conclusion

Dans notre étude, nous avons estimé, analysé et comparé des formes fonctionnelles variant de linéaires à non linéaires pour trouver une forme au meilleur ajustement qui produise des résultats réalistes. Pour les ordinateurs de bureau et les ordinateurs portatifs, le modèle bilogarithmique (où RAM est la seule variable explicative transformée) semble le mieux répondre à ces critères. En guise de conclusion, disons que trois ans environ se sont écoulés depuis la dernière révision à Statistique Canada des équations hédoniques de régression pour les prix des ordinateurs. Vu le caractère dynamique de cette industrie, on doit nettement songer à rendre ces révisions plus fréquentes si on entend améliorer les corrections de qualité. Il n'en est pas question dans cette analyse, mais cette exigence n'en sous-tend pas moins tout le travail que nous avons fait. Nous avons l'intention de soumettre les régressions hédoniques à des actualisations plus fréquentes pour pouvoir appréhender la vaste évolution technologique que connaît l'industrie²⁷.

²⁵ Nous n'avons pas jugé bon d'inclure dans notre exposé les résultats propres au modèle non linéaire. Contentons-nous de préciser que nombreux ont été les modèles qui n'étaient pas en convergence et que, dans le cas des modèles convergents, nous avons trouvé un maximum local qui produisait des valeurs absurdes de coefficients.

²⁶ Nous avons calculé le prix après mise à niveau de la même manière que pour les ordinateurs de bureau.

²⁷ Nous prévoyons au moins des révisions annuelles.

Tableau 1 : Tableau synoptique des équations hédoniques estimées par la Division des prix

	Première estimation		Deuxième estimation			Troisième estimation		
Date	1989-1990		1991			1995-1996		
Source des données	Rapports Data Pro		Rapports Data Pro			Guide des acquisitions de microinformatique de Travaux publics et Services gouvernementaux Canada		
Type d'ordinateurs	16 et 32 bits		16 et 32 bits			Ordinateur Pentium de bureau		
Nombre d'observations	185		89 et 64 ou 153 au total			41		
Forme fonctionnelle	Semi-logarithmique		Linéaire			Semi-logarithmique		
Modèle estimé pour les mini-ordinateurs	<i>Variable</i>	<i>Coefficient estimé</i>	<i>Variable</i>	16 bits <i>Coefficient estimé</i>	32 bits <i>Coefficient estimé</i>	Ensemble <i>Coefficient estimé</i>	<i>Variable</i>	<i>Coefficient estimé</i>
	Constante	5,968	Vitesse d'exécution	100,1	246,0	179,3	Constante	7,318
	Vitesse d'exécution	0,037	Mémoire vive	0,914	0,837	0,88	Disque dur	0,000058345
	Mémoire vive	0,069	Taille du disque dur	0,026	0,017	0,02	Mémoire vive	0,013819
	Taille du disque dur	0,032	Temps moyen d'accès	s.o.	-12,48	s.o.	Vitesse	0,003407
	Baies d'unités int.	0,036	N ^{bre} d'unités de disque	s.o.	s.o.	594,7	Moniteur	0,212789
	Entreprise	0,277	Ports série	-954,8	s.o.	-694,3	17 po	
	Commande postale	-0,563	Fentes d'ext. dispon.	215,6	385,3	281,3	Entreprise	0,212789
	Fentes d'ext. dispon.	0,024	Potentiel tâches	s.o.	s.o.	-1614,2		
	Nombre de bits	0,211	Service et fiabilité	408,7	s.o.	s.o.		
	R ²	plus de 0,80	R ² corrigé	0,86	0,77	0,91	R ² corrigé	0,78

* s.o. – sans objet

Tableau 2 : Liste de variables – ordinateurs de bureau

Variable	Source	Valeur	Nom de la variable
<i>A) Variables quantitatives</i>			
1. Mémoire vive	IDC	Valeur numérique (Mo) : 16, 32, etc.	RAM
2. Unité centrale	CPU Scorecard	Valeur numérique (classement) : 440, 494, 588, etc. (voir l'annexe A)	SCORE
3. Disque dur	IDC	Valeur numérique (Mo) : 2 000, 2 100, etc.	HD
4. Antémémoire étendue	"	Valeur numérique (Ko) : 0, 256, 512, etc.	XCACHE
<i>B) Variables fictives</i>			
5. Type de disque dur	"	Si HDTYPE = UWSCSI3, HDTYPU=1, 0 dans les autres cas	HDTYPU
		Si HDTYPE = IDEenh, HDTYPID=1, 0 dans les autres cas	HDTYPID
		Si HDTYPE = autre, HDTYPU=0 et HDTYPID=0	
6. Disque compact	"	Non = 0 Oui = 1	CD
7. Carte réseau	"	Non = 0 Ethernet = 1	NTWCD
8. Mode SCSI	"	Non = 0 Oui = 1	SCSI
9. Modem	"	Non = 0 Oui = 1	MDM
10. Type de boîtier	"	Bureau et demi-boîtier vertical = 0 Poste de travail = 1	CASE
11. Fournisseur	"	Si VENDOR = Hewlett-Packard ou Compaq, BQUAL=1 (BQUAL = meilleure qualité); BQUAL=0 dans les autres cas	BQUAL

Tableau 3 : Liste des variables – ordinateurs portatifs

Variable	Source	Valeur	Nom de la variable
<i>A) Variables quantitatives</i>			
1. Mémoire vive	IDC	Valeur numérique (Mo) : 16, 32, etc.	RAM
2. Unité centrale	CPU Scorecard	Valeur numérique (classement) : 440, 494, 588, etc. (voir l'annexe A)	SCORE
3. Disque dur	IDC	Valeur numérique (Mo) : 2 000, 2 100, etc.	HD
4. Antémémoire étendue	"	Valeur numérique (Ko) : 0, 256, 512, etc.	XCACHE
5. Poids	"	Valeur numérique (livres) : 2,0, 4,3, 7,9, etc.	WGHT
6. Taille de l'affichage	"	Valeur numérique (pouces) : 12,1, 13,3, etc.	DSIZE
<i>B) Variables fictives</i>			
7. Qualité du moniteur	"	DS = 0 (double balayage, faible qualité) AM = 1 (matrice active, haute qualité)	DQUAL
8. Fournisseur	"	Si VENDOR = IBM, HIEND=1 (HIEND = modèles haut de gamme); HIEND=0 dans les autres cas	HIEND

Tableau 4 : Ordinateurs de bureau – résultats du modèle linéaire 1

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	-350,10	0,099
RAM	20,4200	0,000
HARD	0,04514	0,041
SCORE	0,33720	0,114
XCACHE	1,03360	0,000
CASE	393,820	0,005
CD	-13,266	0,898
MDM	67,4540	0,751
NTWCD	50,3220	0,576
SCSI	977,230	0,000
BQUAL	432,430	0,000
HDTYPU	1502,20	0,000
HDTYPID	588,200	0,000
	$R^2 = 0,7787$	$R^2 \text{ corr.} = 0,7740$
	$\sigma = 946,33$	
	Fonct. log. de vraisemblance = -4 757,83	
	Statistique chi-carré DLR de Davidson et MacKinnon =	495,8390

Tableau 5 : Ordinateurs de bureau – résultats du modèle semi-logarithmique 2

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	6,678100000	0,000
RAM	0,003390100	0,000
HARD	0,000023039	0,000
SCORE	0,000250930	0,000
XCACHE	0,000585340	0,000
CASE	0,171610000	0,000
CD	0,053291000	0,015
MDM	0,055399000	0,219
NTWCD	0,032204000	0,092
SCSI	0,346070000	0,000
BQUAL	0,105000000	0,000
HDTYPU	0,312810000	0,000
HDTYPID	0,171750000	0,000
	$R^2 = 0,8550$	$R^2 \text{ corr.} = 0,8519$
	$\sigma = 0,20045$	
	Fonct. log. de vraisemblance = -4 416,09	
	Statistique chi-carré DLR de Davidson et MacKinnon =	0,06114

Tableau 6 : Ordinateurs de bureau – résultats du modèle bilogarithmique 3A

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	3,974000000	0,000
LRAM	0,355170000	0,000
LHARD	0,119800000	0,000
LSCORE	0,139720000	0,005
XCACHE	0,000585557	0,000
CASE	0,165080000	0,000
CD	0,018118000	0,417
MDM	0,077837000	0,085
NTWCD	0,024055000	0,209
SCSI	0,247370000	0,000
BQUAL	0,099690000	0,000
HDTYPU	0,356850000	0,000
HDTYPID	0,195650000	0,000
	R ² = 0,8551	R ² corr. = 0,8520
	σ = 0,20039	
	Fonct. log. de vraisemblance = -4 415,92	
	Statistique chi-carré DLR de Davidson et MacKinnon =	10,2105

Tableau 7 : Ordinateurs de bureau – résultats du modèle bilogarithmique 3B, variable RAM seulement

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	5,641200000	0,000
LRAM	0,342100000	0,000
HARD	0,000019173	0,000
SCORE	0,000188440	0,000
XCACHE	0,000569590	0,000
CASE	0,161490000	0,000
CD	0,019959000	0,360
MDM	0,070749000	0,112
NTWCD	0,022961000	0,224
SCSI	0,276120000	0,000
BQUAL	0,106460000	0,000
HDTYPU	0,348220000	0,000
HDTYPID	0,201540000	0,000
	R ² = 0,8589	R ² corr. = 0,8559
	σ = 0,19774	
	Fonct. log. de vraisemblance = -4 408,27	
	Statistique chi-carré DLR de Davidson et MacKinnon =	0,8053

Tableau 8 : Ordinateurs de bureau – résultats des estimations de Box-Cox du modèle 4

Modèles		λ estimé	Fonction logarithmique de vraisemblance pour λ estimé	Fonction logarithmique de vraisemblance pour $\lambda = 0$
Cas non restreint	P et RAM	-0,36	-4 379,92	-4 408,27
Cas restreint ($\lambda = 0$ pour P)	RAM, HARD et SCORE	0,47	-4 405,84	-4 415,92
	RAM et HARD	0,40	-4 404,04	-4 410,85
	RAM	0,34	-4 403,88	-4 408,27
	HARD	-0,06	-4 413,93	-4 413,94
	RAM et logarithme de HARD	0,41	-4 404,60	-4 410,85

Tableau 9 : Ordinateurs de bureau – résultats de diverses estimations du modèle non linéaire

1.) Convergence des modèles	
<i>Modèle</i>	<i>Résultats</i>
RAM + HARD + SCORE	Non-convergence après 1 500 itérations
RAM + HARD	Convergence après 101 itérations
RAM	Convergence après 148 itérations
HARD	Non-convergence après 1 500 itérations
2. Test F : modèle non linéaire de RAM	
<i>Modèle</i>	<i>Statistique</i>
RAM + HARD+ SCORE	-11,6212
RAM + HARD	-0,1811
3. Test de rapport de vraisemblances : modèle non linéaire de RAM	
<i>Modèle</i>	<i>Statistique</i>
RAM + HARD+ SCORE	-5,9324
RAM + HARD	-,01752

Tableau 10 : Ordinateurs de bureau – résultats du modèle non linéaire 5A

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	262,68000	0,418
RAM - β_{Ram}	0,17436000	0,159
α_{Ram}	0,38563000	0,000
HARD - β_{Hard}	0,00002217	0,681
α_{Hard}	0,99131000	0,000
SCORE - β_{SCORE}	-257,540000	0,427
α_{SCORE}	-0,68669000	0,431
XCACHE	0,00058533	0,000
CASE	0,155610000	0,000
CD	0,030314000	0,164
MDM	0,067347000	0,133
NTWCD	0,026200000	0,053
SCSI	0,284390000	0,166
BQUAL	0,107780000	0,000
HDTYPU	0,329090000	0,000
HDTYPID	0,189860000	0,000
	$\sigma = 0,19502$	
	Fonct. log. de vraisemblance = 124,2577	
	SCE = 21,906	

Tableau 11 : Ordinateurs de bureau – résultats du modèle non linéaire 5B

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	6,05590000	0,000
RAM - β	0,23419000	0,200
α	0,341860000	0,001
HARD	0,000018943	0,000
SCORE	0,000199520	0,000
XCACHE	0,000577370	0,000
CASE	0,153620000	0,000
CD	0,031326000	0,152
MDM	0,065922000	0,136
NTWCD	0,025065000	0,163
SCSI	0,300520000	0,000
BQUAL	0,108840000	0,000
HDTYPU	0,322470000	0,000
HDTYPID	0,191910000	0,000
	$\sigma = 0,19402$	
	Fonct. log. de vraisemblance = 127,2194	
	SCE = 21,682	

Tableau 12 : Ordinateurs de bureau – comparaison des variations de prix par les modèles estimés

Modèle	Prix initial (\$)	Prix après mise à niveau (\$)	Variation en pourcentage
1. Semi-logarithmique (2)	2 000	2 945,52	47,3
2. Bilogarithmique avec logarithme de RAM (3B)	2 000	2 905,16	45,3
3. « Identification de t » (5B)	2 000	6 060,18	203,0

Mise à niveau :

1. ΔRAM . De 64 à 128 Mo = 64 Mo.
2. $\Delta HARD$. De 8,4 à 12,9 Go = 4,5 Go.
3. $\Delta SCORE$. De 865 (Pentium II 300) à 1 130 (Pentium II 400) = 265.

Tableau 13 : Ordinateurs portatifs – résultats du modèle linéaire 1

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	-3 485,8	0,000
RAM	6,45120	0,006
HARD	0,19805	0,000
SCORE	1,29620	0,001
XCACHE	2,15810	0,000
WGHT	-179,33	0,000
DSIZE	365,490	0,000
DQUAL	228,900	0,172
HIEND	653,430	0,000
	$R^2 = 0,8058$	$R^2 \text{ corr.} = 0,7964$
	$\sigma = 525,21$	
Fonct. log. de vraisemblance =	-1 332,17	
Statistique chi-carré DLR de Davidson et MacKinnon =		6,76084

Tableau 14 : Ordinateurs portatifs – résultats du modèle semi-logarithmique 2

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	6,176600000	0,000
RAM	0,001554800	0,000
HARD	0,000039409	0,000
SCORE	0,000427790	0,000
XCACHE	0,000602900	0,000
WGHT	-0,054339000	0,000
DSIZE	0,104110000	0,009
DQUAL	0,105920000	0,000
HIEND	0,156210000	0,000
	$R^2 = 0,8215$	$R^2 \text{ corr.} = 0,8129$
	$\sigma = 0,12655$	
Fonct. log. de vraisemblance =	-1 319,29	
Statistique chi-carré DLR de Davidson et MacKinnon =		0,04535

Tableau 15 : Ordinateurs portatifs – résultats du modèle bilogarithmique 3A

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	-0,05425000	0,911
LRAM	0,117630000	0,001
LHARD	0,257100000	0,000
LSCORE	0,291580000	0,000
LXCACHE	0,176980000	0,000
LWGHT	-0,232140000	0,000
LDSIZE	1,140200000	0,000
DQUAL	0,093781000	0,024
HIEND	0,156650000	0,000
	$R^2 = 0,8157$	$R^2 \text{ corr.} = 0,8067$
	$\sigma = 0,12861$	
Fonct. log. de vraisemblance = -1 322,10		
Statistique chi-carré DLR de Davidson et MacKinnon =		0,04461

Tableau 16 : Ordinateurs portatifs – résultats du modèle bilogarithmique 3B avec RAM seulement

Variable	Estimation	Valeur des probabilités
Valeur à l'origine	5,893600000	0,000
LRAM	0,113760000	0,006
HARD	0,000038954	0,000
SCORE	0,000422990	0,000
XCACHE	0,000605140	0,000
WGHT	-0,05049700	0,000
DSIZE	0,096929000	0,000
DQUAL	0,103450000	0,000
HIEND	0,158190000	0,000
	$R^2 = 0,8245$	$R^2 \text{ corr.} = 0,8160$
	$\sigma = 0,12548$	
Fonct. log. de vraisemblance = -1 317,82		
Statistique chi-carré DLR de Davidson et MacKinnon =		0,02786

Tableau 17 : Ordinateurs portatifs – résultats des estimations de Box-Cox du modèle 4

Modèle	λ estimé	Fonction logarithmique de vraisemblance pour λ estimé	Fonction logarithmique de vraisemblance pour $\lambda = 0$	
Cas non restreint	P seulement (semi-logarithmique)	-0,04	-1 319,26	-1 319,29
	P et RAM, HARD, XCACHE, SCORE, WGHT, DSIZE (modèle bilogarithmique 3A)	0,04	-1 322,08	-1 322,10
Cas restreint ($\lambda = 0$ pour P)	RAM (modèle bilogarithmique 3B)	0,00	-1 317,82	-1 317,82

Tableau 18 : Ordinateurs portatifs – comparaison des variations de prix par les modèles estimés

Modèle	Prix initial (\$)	Prix après mise à niveau (\$)	Variation en pourcentage
4. Semi-logarithmique (2)	2 900	3 715,21	28,1
5. Bilogarithmique (3A)	2 900	3 496,13	20,6
6. Bilogarithmique avec logarithme de RAM (3B)	2 900	3 593,78	23,9
<i>Mise à niveau :</i>			
4.	<i>ΔRAM.</i> De 32 à 64 Mo = 32 Mo.		
5.	<i>ΔHARD.</i> De 3,2 à 4,0 Go = 0,8 Go.		
6.	<i>ΔSCORE.</i> De 546 (Pentium 233 MMX) à 793 (Pentium II 266) = 247.		

Annexe A – Classement des processeurs centraux

UNITÉ CENTRALE	CLASSEMENT
1. Pentium 150 MMX	398
2. Pentium 166 MMX	440
3. Pentium 200 MMX	494
4. Pentium Pro 200	588
5. Pentium 233 MMX	546
6. Pentium 266 MMX	634
7. Pentium 300 MMX	709
8. Pentium Celeron 266	571
9. Pentium Celeron 300	603
10. Pentium Celeron 300A	776
11. Pentium Celeron 333	830
12. Pentium Celeron 366	998
13. Pentium Celeron 400	1 092
14. Pentium Celeron 433	1 183
15. Pentium II 266PE	815
16. Pentium II 300PE	879
17. Pentium II 233	704
18. Pentium II 266	793
19. Pentium II 300	865
20. Pentium II 333	949
21. Pentium II 350	1 000
22. Pentium II 366	1 053
23. Pentium II 400	1 130
24. Pentium II 450	1 240
25. Pentium II Xeon 450	1 370*
26. Pentium III 450	1 500
27. Pentium III 500	1 650
28. Pentium III Xeon 500	1 815*
29. Pentium III Xeon 550	1 997*

*Dans le cas de la série Xeon, nous avons estimé le classement à l'aide de données comparatives figurant au site Internet www.intel.com et à la suite de communications avec les gens chargés de classer les processeurs centraux au site Internet « The CPU Scorecard ».

Bibliographic

Berndt, Ernst R. (1991), *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley Publishing Company, Reading Massachusetts.

Box, G.E.P. and D. Cox (1964), 'An Analysis of Transformations', *Journal of the Royal Statistical Society, Series B*, 211-264.

Darnell, Adrian C. (1994), *A Dictionary of Econometrics*, Vermont: Edward Elgar Publishing Limited.

Davidson, R., and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Greene, William H. (1990), *Econometric Analysis*, New York: Macmillan Publishing Company.

Judge, G. G., R.C. Hill, W.E. Griffiths, H. Lutkepohl and T.Lee (1988), *Introduction to the Theory and Practice of Econometrics*, New York: John Wiley & Sons Inc.

Microsoft Press (1991) *Computer Dictionary*, Macmillan of Canada.

PC WORLD, (1998), "Problem-free PCs - Who Makes Them?", San Francisco, November.

PC WORLD, (1996), "Best PCs for Windows 95", San Francisco, March.

Triplett, Jack E. (1987), "Price and Technological Change In A Capital Good: A Survey Of Research On Computers" *Bureau of Economic Analysis, Discussion Paper 23*.