

***Traitement des changements de qualité dans
les indices des prix des ordinateurs***

~ Examen des pratiques actuelles et des pratiques proposées ~

par Fred Barzyk et Matthew MacDonald

*Section des prix des dépenses d'investissement
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
matthew.macdonald@statcan.ca*

octobre 2001

Toutes les vues exprimées appartiennent aux auteurs et ne correspondent pas nécessairement à celles de la Division des prix ni de Statistique Canada.

1.0 Introduction¹

Depuis dix ans, les ordinateurs – et d’ailleurs les autres produits de haute technologie – ont eu un comportement unique sur le plan des prix en ce que les niveaux généraux de ces derniers n’ont pas suivi la progression des produits en qualité et en rendement. À la différence de produits classiques comme le bois, l’automobile ou l’acier, les ordinateurs ont connu une évolution technologique rapide et ininterrompue, d’où un incessant mouvement d’amélioration et d’accroissement de production qui n’est pas en rapport avec les niveaux correspondants des prix.

Vu l’importance des ordinateurs pour l’économie canadienne, la Division des prix a produit une série indiciaire des prix des ordinateurs pendant presque dix ans à l’aide de renseignements sur les prix pratiqués par les distributeurs. Cette série entre dans la composante des machines et du matériel de bureau de l’indice général des prix des machines et du matériel (IPMM) et sert à la détermination des mouvements des prix de la composante du matériel et des fournitures informatiques de l’Indice des prix à la consommation.

Notre étude a un triple but général. Notre objectif le plus important est d’examiner et d’évaluer la technique en usage de correction de qualité pour les ordinateurs, la méthode des échantillons appariés à remplacement hédonique et la solution de rechange que nous proposons avec un indice hédonique direct et un autre par échantillons appariés en chevauchement. Un deuxième objectif est de recommander une méthodologie pour l’avenir de cette série. Un dernier est d’évoquer, pour un meilleur traitement de la question, le contexte général ou historique du marché des ordinateurs et de broser brièvement un tableau synoptique des tendances et des faits les plus récents.

2.0 Source de données

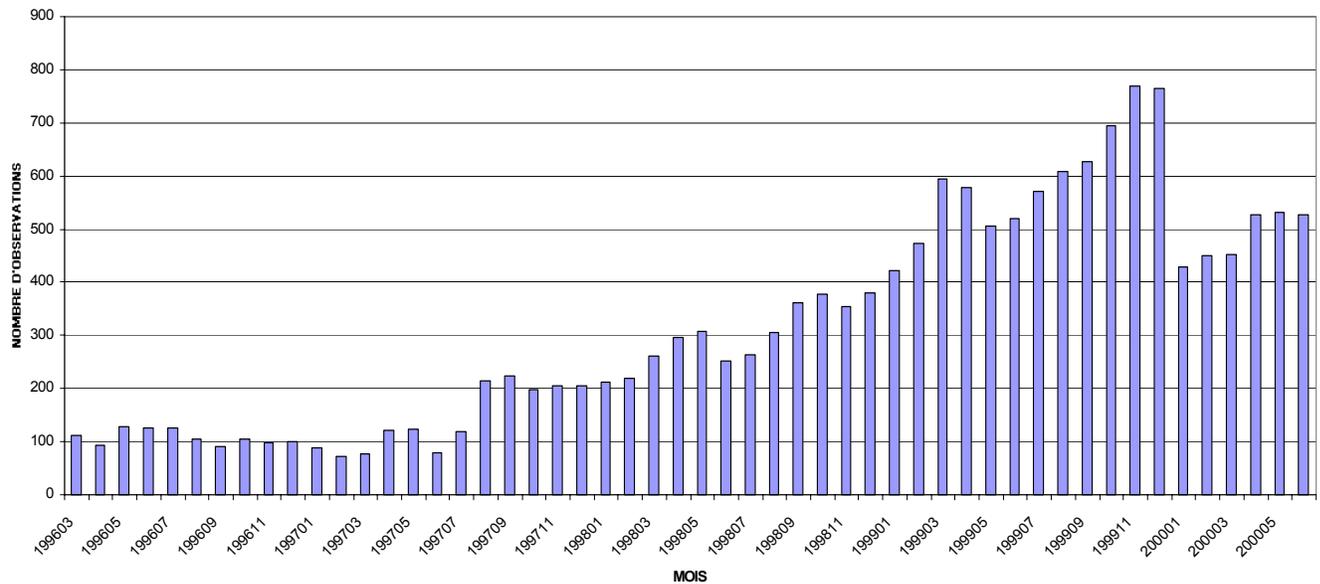
Depuis 1996, la Division des prix exploite les données mensuelles que lui fournit International Data Corporation of Canada (IDC Canada) afin de produire son propre indice des prix des micro-ordinateurs. Les données de notre étude viennent de la base d’information par laquelle on produit justement la série indiciaire des prix des ordinateurs, c’est-à-dire des prix des ordinateurs de bureau vendus au double secteur commercial et public pendant la période mars 1996-juin 2000. Les modèles qui figurent dans cette base sont à puces Intel. Ils sont écoulés par les fournisseurs qui dominent sur le marché (Compaq, IBM, Hewlett-Packard, Dell, etc.), puisqu’il se trouve que, sur ce marché, la plupart des micro-ordinateurs commercialisés sont munis de puces Intel et que les autres fournisseurs de processeurs (AMD, Cyrix, etc.) ont beaucoup de mal à s’implanter.

Comme l’indique la figure I.1, si on observait tous les mois un peu plus de 100 modèles en 1996, ce nombre dépassait les 500 en juin 2000. Les modèles se sont diversifiés davantage : le nombre moyen de modèles par catégorie de processeurs centraux a doublé, passant d’une dizaine

¹ Ce document ne doit pas être cité sans la permission expresse des auteurs. Ceux-ci désirent remercier de leurs observations et de leurs suggestions intéressantes Andy Baldwin, Robin Lowe, Marc Prud’homme et Kam Yu, de la Division des prix, ainsi qu’Erwin Diewert, Jack Triplett et Ralph Turvey. Les vues exprimées sont celles des auteurs, et non pas nécessairement celles de la Division des prix ni de Statistique Canada.

en 1996 à un peu plus de 20 en 2000. Cela indique bien la croissance de ce marché pour la quantité et la variété de modèles offerts par les fournisseurs dominants.

Figure I.1- DONNÉES MENSUELLES SUR LES PRIX DE IDC



3.0 Organisation de l'étude

Notre étude compte quatre chapitres. Le premier donne un aperçu des prix et du rendement des ordinateurs depuis cinq ans. Les chapitres 2 et 3 décrivent la méthode par échantillons appariés en chevauchement et la méthode hédonique permettant d'appréhender les changements de qualité des ordinateurs. Le chapitre 4 compare ces méthodes et, d'après les résultats de cet exercice, recommande une future méthodologie pour l'indice des prix des ordinateurs.

CHAPITRE 1 – Analyse historique

1.1 Introduction

Dans ce chapitre, nous ferons l'historique des prix et du rendement des ordinateurs. Nous dégagerons et examinerons les tendances de ces prix et de ce rendement (selon une mesure où interviennent principalement les catégories de processeurs, la mémoire vive et autres caractéristiques semblables), le but étant de montrer que, avant toute correction de qualité, le prix moyen des ordinateurs a été stable (ou a même décru), alors que leur rendement s'accroissait. Si tel est le cas, on peut dire que, dans le contexte de l'élaboration d'un indice des prix, la question des changements de qualité l'emporte d'emblée sur la question de la variation des prix, ce qui fera du traitement des variations de qualité le point de mire de la production d'un indice des prix des ordinateurs.

1.2 Grandes variables

Les grandes variables abordées dans ce chapitre sont celles des prix des ordinateurs, de la mémoire vive (RAM), des unités ou processeurs centraux (CPU), des catégories de processeurs et de puces, de la taille des disques durs (HD) et de l'antémémoire étendue (CACHE). La variable dépendante est le prix des ordinateurs, c'est-à-dire le prix moyen effectivement pratiqué pour un modèle ou un système informatique. Ce n'est pas encore le prix négocié, mais le prix effectif moyen en est très proche. Pour bien expliquer ce point, il convient de noter que, à l'heure actuelle, la Division des prix reçoit deux prix pour un modèle, à savoir un *prix courant* et un *prix effectif moyen*.

Le *prix courant* est le prix de vente proposé par le fournisseur (IBM, Dell, etc.). Par ailleurs, on calcule le prix effectif moyen en obtenant d'abord un prix moyen de revente par ordinateur à l'aide d'un relevé effectué aux divers maillons de la chaîne de distribution (revendeurs et distributeurs) et en augmentant ou diminuant ce prix pour ainsi dégager un prix effectif moyen au Canada. Les prix effectifs ne comprennent ni les taxes ni les frais d'expédition et de manutention.

Les prix courants et effectifs seront égaux ou inégaux. S'il s'agit de ventes commerciales d'ordinateurs aux administrations publiques et aux grandes et moyennes entreprises, les achats se font surtout en quantité avec un certain nombre de spécifications et d'exigences en matière de rendement. Ajoutons que de tels achats relèvent généralement de contrats, auquel cas les prix négociés ne sont pas immédiatement constatables (il n'y a pas d'« acte de vente » comme pour un ordinateur domestique acheté à un détaillant local et il n'y a donc pas non plus de données de lecture optique au comptoir de détail). Les contrats qui sont dressés comportent souvent un service après-vente, des garanties et des « extra » divers comme des progiciels. Enfin, dans le cas des marchés publics, les prix fixés dans des offres permanentes ou à commandes valent habituellement pour un an². Le plus souvent donc, le prix effectif moyen sera le prix effectivement négocié dans le cas du secteur public.

² Ce sont en particulier les achats du gouvernement fédéral.

Par le passé, les principaux fournisseurs fixaient ordinairement des « prix à ne pas dépasser », c'est-à-dire des prix plafonds auxquels devaient s'en tenir les revendeurs et les distributeurs. La situation a toutefois changé depuis l'avènement des fournisseurs directs (Dell en l'occurrence). La récente croissance du phénomène des « prix Web » (IBM et Compaq sont aujourd'hui dans la mêlée avec d'autres) est venue réduire le rôle des prix courants de catalogue, si bien qu'aujourd'hui la différence entre le prix courant et le prix effectif est nulle ou presque nulle dans la plupart des cas.

Des grandes variables que nous avons mentionnées, les plus importantes sont RAM et CPU. La capacité de mémoire vive (RAM) est essentiellement la capacité de stockage d'information d'un ordinateur et d'exécution d'instructions. Elle se mesure en mégaoctets (Mo). Dans notre étude, la variable des unités ou processeurs centraux (CPU) est représentée par une cote de rendement, c'est-à-dire un chiffre attribué à tous les processeurs ou à toutes les puces d'après des tests comparatifs de rendement à l'échelle de l'industrie³. Les cotes CPU sont importantes, parce qu'elles permettent d'évaluer les changements de qualité lorsque se présentent de nouvelles générations de puces informatiques. Une autre grande caractéristique des prix des ordinateurs est la taille des disques durs (qui se mesure aussi en mégaoctets), laquelle détermine combien de composants et d'applications un ordinateur pourra soutenir en stockage et en traitement. Enfin, l'antémémoire (CACHE), où l'ordinateur stocke les instructions et les données d'accès fréquent, complète la liste des éléments fondamentaux d'un système informatique (et se mesure en kilooctets (Ko)). Dans la perspective de notre analyse historique, nous décrivons les mouvements des diverses variables par des moyennes mensuelles.

1.3 Autres variables

Vu leur importance, les variables clés sont présentes, bien sûr, tout au long de la période étudiée, et il faut encore mentionner les variables qui ont influé (bien que leur incidence ait été relativement secondaire) sur les différences de prix entre modèles dans le temps. Le tableau A en fait mention là où, avant août 1997, des éléments comme les cartes mémoire vidéo, le nombre de fentes d'extension et les baies d'unités s'ajoutaient comme descripteurs des modèles dans les données sur les prix. Après 1997, ils ont été remplacés par des caractéristiques des modèles comme les modems, les cartes son et réseau et les unités de disque compact, qui ont gagné en importance. Il sera question de ces derniers éléments dans les chapitres qui suivent.

³ Nous avons pris les cotes de rendement des processeurs centraux au site Web www.cpuscorecard.com. Des entretiens qu'ont eus les auteurs de l'étude avec les représentants d'Intel – qui est actuellement le premier fabricant mondial de puces informatiques –, il ressort que, pour appréhender au mieux le rendement comparé entre générations de processeurs, on doit tirer des cotes de rendement (comme dans la présente étude) de résultats comparatifs pondérés. On met fréquemment ces cotes à jour pour rendre les nouveaux modèles comparables. Les cotes utilisées aux fins de notre analyse sont présentées à l'annexe I et datent du 17 juillet 2000.

Tableau A – Variables explicatives supplémentaires

Période	Variables
Mars 1996 à juillet 1997	Cartes mémoire vidéo, fentes d'extension, baies d'unités, types de boîtiers et éléments logiciels
Août 1997 à juin 2000	Modems, cartes son, cartes réseau, unités de disque compact et mode Small Computer System Interface (SCSI)

1.4 Comparaison historique

Prix

D'après les données, le prix moyen d'un ordinateur de bureau actuellement acheté par les entreprises ou les administrations publiques est demeuré dans la fourchette 2 500 \$-4 500 \$ pendant la période étudiée. Pour l'ensemble de nos observations, le prix moyen s'établissait à 3 057 \$.

Comme on peut le voir à la figure 1.1, le prix mensuel moyen d'un ordinateur a diminué de mars 1996 à juin 2000. Si on ajuste une courbe de tendance au logarithme naturel du prix, on dégage un taux mensuel de décroissance de 0,5 %. Ce n'est pas une pente raide dans l'ensemble, mais on peut relever des sous-périodes où les mouvements ont été plus prononcés. Ainsi, l'évolution à la baisse du prix moyen est plus marquée en 1996, en 1999 et au premier semestre de l'an 2000, et ce même prix moyen évolue en hausse en 1997 et se stabilise en 1998. Pour expliquer ces variations, on doit tenir compte de deux phénomènes importants, à savoir l'arrivée de nouvelles générations technologiques de processeurs centraux (surtout les processeurs Pentium I, Pentium II et Pentium III) et leur importance relative dans les données sur les prix tout au long de leurs durées utiles respectives. La figure 1.2 décrit le mouvement des prix moyens des diverses familles de processeurs avec les prix moyens totaux. La figure 1.3 indique l'évolution pendant la période de l'importance relative des divers processeurs en question dans les échantillons.

À partir de mars 1996, on observe une très forte tendance à la baisse du prix moyen des ordinateurs (voir la figure 1.4). Les prix moyens des ordinateurs de bureau décroissent d'environ 4 %⁴. Si on revient à la figure 1.3, on constate que tout l'échantillon est formé d'ordinateurs Pentium dans cette période en particulier, ces ordinateurs étant alors devenus la pièce maîtresse de l'informatique d'entreprise à la place des ordinateurs 486. Pendant ce temps, le prix moyen décroissait d'environ 1 200 \$, passant de 4 188 \$ à 3 023 \$. Ce mouvement s'explique par plusieurs événements connus dans l'industrie. D'abord, Intel a abaissé le prix de ses puces de processeurs à plusieurs reprises pour maintenir sa domination presque parfaite du marché et ménager une forte demande future de ses produits. Joint à ce phénomène, il y a eu celui du coût décroissant des mémoires (mémoire vive et disque dur). De plus, les fournisseurs de premier plan ont sabré les prix d'un grand nombre de modèles d'ordinateurs de bureau –souvent de plus de 20 % –, évoquant la baisse des prix des composants, un meilleur contrôle des stocks et une

⁴ Le taux mensuel d'accroissement (dont il est question ici comme dans le reste de notre étude) est le coefficient de pente obtenu par régression du logarithme naturel de la variable dépendante (prix, indice des prix, RAM, etc.) dans le temps.

situation d'intensification de la concurrence où ils devaient abaisser stratégiquement leurs prix pour accroître ou maintenir leurs parts de marché⁵. L'industrie croyait à l'époque que les fournisseurs seraient contraints de brader leurs vieux produits pour faire place aux nouveaux avec toutes les réductions de prix qui s'ensuivent⁶. L'introduction par Intel de ses Pentium 200 sur le marché en juin 1996 en a été la confirmation, ceux-ci offrant plus de possibilités par rapport au choix existant d'ordinateurs et devant constamment entraîner à la baisse les prix des modèles dans la fourchette des vitesses d'exécution 100-166 MHz⁷. Avec la diffusion prévue de la technologie MMX au début de 1997, ces facteurs ont considérablement fait reculer les prix des ordinateurs dans cette période.

En 1997, la tendance du prix moyen des ordinateurs faisait voir un renchérissement approximatif de 1,3 % en valeur mensuelle (voir la figure 1.5), phénomène surtout attribuable à l'arrivée de deux nouvelles puces de processeurs sur le marché. Le premier relèvement des prix a eu lieu dès janvier lorsqu'Intel a finalement diffusé son processeur MMX⁸. La majoration des prix n'a cependant été que légère (hausse moyenne de 125 \$ seulement), car l'industrie a pensé que le meilleur était encore à venir, la technologie Pentium II se profilant déjà à l'horizon⁹. Cette attente devait se vérifier : introduit en mai 1997, le processeur Pentium II d'Intel a été à l'origine de la plus grande augmentation des prix de toute la période¹⁰. De cette introduction à l'automne de 1997, le prix moyen des ordinateurs s'est élevé de 827 \$. Peu après l'arrivée du Pentium II, il a atteint un maximum approximatif de 5 800 \$ et, au début de 1998, il avait glissé sous la barre des 4 000 \$ (voir la figure 1.2). Pentium II s'est aussi taillé une place appréciable assez rapidement dans l'échantillon. En six mois seulement, il a représenté une proportion d'un peu moins de 40 % de tous les ordinateurs. Une vive demande s'attachait à ces machines, et la taille moyenne des mémoires vives et des disques durs a plus que doublé pour recevoir cette nouvelle génération technologique. Les capacités d'antémémoire ont également fait un bond moyen de 180 Ko. Phénomène fréquent dans cette industrie, les majorations de prix se sont émoussées peu après la consécration du Pentium II comme produit standard de l'industrie, et les prix ont repris leur tendance normale à la baisse. Les fournisseurs changeaient leur orientation en affaires et optaient pour une augmentation des quantités et une diminution des prix dans l'espoir qu'une montée de la demande serait là pour compenser la faiblesse des marges bénéficiaires. Le tout s'est révélé un obstacle pour les nouveaux concurrents, mais une aubaine par excellence pour les consommateurs¹¹.

En 1998, on n'a pas relevé de tendance effective de l'évolution des prix, mais plutôt observé une certaine instabilité. Le prix moyen a évolué en dents de scie, la moyenne se situant dans la zone des 3 100 \$ (voir la figure 1.6). Plusieurs facteurs ont causé cette oscillation du prix moyen dans l'année. D'abord, le Pentium II a commencé à supplanter le Pentium I de plus en plus désuet comme ordinateur de choix pour le consommateur, ainsi qu'à accroître son importance relative dans les modèles décrits par les données sur les prix. Même avec les baisses de prix de 1998, le

⁵ Voir www.newscan.com, *The Kitchener-Waterloo Record*, 10 juin 1996.

⁶ Voir www.newscan.com, *The Kitchener-Waterloo Record*, 10 juin 1996.

⁷ Voir le communiqué *Intel Introduces 200-MHz Pentium Processor* du 10 juin 1996 à www.intel.com.

⁸ Voir les archives de presse à www.intel.com, 8 janvier 1997.

⁹ Voir les articles à www.pcworld.com/hardware/desktop, février 1997.

¹⁰ Voir les archives de presse à www.intel.com, 7 mai 1997.

¹¹ Voir le communiqué de *The Toronto Star* à www.newscan.com, avril 1997.

Pentium II coûtait bien plus (près du double) que le Pentium qu'il remplaçait, d'où un amortissement de l'effet de la décroissance des prix des Pentium¹². Ajoutons que la prise en compte des appareils Celeron moins chers dans les données sur les prix en juillet 1998 n'a guère influé sur le prix moyen global, puisque le Pentium II conservait la part du lion parmi les ordinateurs commercialisés. Le seul renchérissement digne de mention a eu lieu lorsqu'a été diffusé le Pentium II à 400 MHz et introduite la nouvelle série Xeon de puces destinées aux postes de travail¹³.

La figure 1.7 indique pour la période début de 1999-juin 2000 une évolution générale à la baisse des prix, même avec l'arrivée de la génération Pentium III en janvier 1999. À cette époque, les acheteurs se sont faits plus avisés, constatant que les modèles haut de gamme à prix fort finiraient par se vendre moins cher¹⁴. Les distributeurs étaient aux prises avec une baisse des prix à long terme et l'avènement du Pentium III a eu pour effet net un faible renchérissement moyen de 100 \$. Le phénomène a duré quelques mois seulement et, à l'été de 1999, l'ordinateur de bureau moyen se vendait à moins de 3 000 \$, niveau où son prix se situe depuis lors. Les analystes jugent que ces tendances de l'industrie ne s'inverseront pas et que les prix continueront à descendre compte tenu surtout de la montée des attentes au sujet de la nouvelle génération technologique Pentium IV¹⁵.

Mémoire vive (RAM)

Nul doute que, de toutes les variables indépendantes, c'est la variable RAM qui influe le plus sur les prix des ordinateurs. D'après les données, la capacité moyenne de mémoire vive des modèles s'est accrue d'environ 4,2 % de mois en mois (voir la figure 1.8). En 1996, elle était d'environ 16 Mo. Aux deux premiers trimestres de l'an 2000, elle a octuplé pour approcher des 128 Mo. En règle générale, il y a deux facteurs qui jouent. D'abord, par suite de l'arrivée des Pentium II et III, il a fallu une meilleure capacité de base en mémoire vive pour soutenir les nouvelles puces et, ensuite, le coût de cette mémoire a considérablement diminué pendant la période étudiée¹⁶. L'ample bond de la capacité moyenne au début de 2000 tient au passage du Pentium II au Pentium III comme génération technologique de prédilection des entreprises et des administrations publiques. Les ordinateurs de bureau ont constamment gagné en qualité à cause de l'incidence de ce surcroît de mémoire accessible sur la vitesse d'exécution.

¹² En 1998, les prix mensuels respectifs du Pentium I, du Pentium II et du Celeron s'établissaient en moyenne à 1 884 \$, 3 862 \$ et 1 662 \$.

¹³ Voir www.pcworld.com, juin 1998. On remarquera le prix moyen bien plus élevé du Xeon (dans la zone des 7 600 \$). Ce sont de très puissantes machines (postes de travail à capacité moyenne de mémoire vive de 200 Mo) pour les utilisateurs spécialisés qui, à l'heure actuelle, tiennent une place relativement modeste dans l'échantillon. Voir www.smartcomputing.com, *Buying Computers - Desktop Systems*, septembre 1999, vol. 7, numéro 9.

¹⁴ Voir l'« actualité technologique » à www.intel.com/wire/story, 12 février 1999.

¹⁵ Voir l'*Ottawa Citizen*, 22 août 2000.

¹⁶ Au début de 1996, le prix de 4 Mo de mémoire vive est passé de 170 \$ à 70 \$ et se trouve toujours en décroissance. Pour plus de détails, voir le communiqué du *Toronto Star* du 2 mai 1996 à www.newscan.com. En mars 1999, le prix de la mémoire vive en dollars canadiens (selon les prix Internet) avoisinait 3 \$ le mégaoctet sans les frais d'installation ni les taxes.

Processeur central (CPU)

Avec la mémoire vive, l'unité centrale est un grand critère de distinction des prix des modèles d'ordinateurs. Dans l'un et l'autre cas, nous constatons une forte progression mensuelle du rendement moyen à un taux estimatif de 3,9 % (voir la figure 1.9)¹⁷. On doit cette solide ascension des rendements informatiques à la succession des générations technologiques : Pentium I, Pentium II, Pentium III. Fait intéressant, cette évolution du rendement des processeurs centraux semble obéir à la *loi de Moore*, puisque le rendement double à intervalles approximatifs de 18 mois¹⁸.

Disque dur (HD)

Avec les capacités de mémoire vive et d'unité centrale, la capacité de stockage sur disque dur est un facteur clé de distinction des prix des ordinateurs. Comme pour les autres éléments, cette capacité s'est constamment accrue pendant la période étudiée. Le taux d'accroissement mensuel s'est établi en moyenne à 4,4 % (voir la figure 1.10). C'est ainsi que, en juin 2000, les disques durs étaient d'une taille moyenne de 10,9 gigaoctets (Go), comparativement à 1,3 Go en mars 1996.

Antémémoire (CACHE)

L'antémémoire a moins progressé que les éléments déjà mentionnés. La figure 1.11 indique un accroissement mensuel de 1 %, mais la tendance n'est pas aussi forte que pour les trois autres éléments, et il y a plus d'oscillations autour de la valeur tendancielle, ce qui s'explique surtout par une progression qui est largement demeurée à deux crans (catégories dominantes de 256 et 512 Ko). Ce qu'indiquent foncièrement les chiffres, c'est que, jusqu'au milieu ou aux derniers mois de 1997, la capacité de l'antémémoire a été de 256 Ko dans plus de la moitié des observations de prix et que, depuis lors, elle est de 512 Ko pour une même proportion des observations.

1.5 Conclusion

Depuis cinq ans, l'industrie de l'informatique assiste à la genèse d'une double tendance de décroissance des prix et de constant accroissement du rendement. On pourrait peut-être le mieux résumer les données d'analyse historique que livre le présent chapitre en comparant des ordinateurs représentatifs du début (mars 1996) et de la fin (juin 2000) de la période. Cette représentativité est fondée sur les moyennes des variables examinées¹⁹. Les résultats figurent au tableau B. Il suffit de regarder les différences entre ces deux ordinateurs pour se rendre compte

¹⁷ Précisons au lecteur que l'axe des y à la figure 1.9 indique la catégorie correspondante de puces de processeurs plutôt que la cote réelle de rendement.

¹⁸ La loi de Moore dit en gros que la puissance ou le rendement de calcul des processeurs centraux doit doubler tous les 18 mois. Cette règle empirique dans l'industrie vient des déclarations de Gordon Moore, cofondateur de l'Intel Corporation. D'après les données, les cotes moyennes de rendement pour les trois majorations aux 18 mois sont 288,36, 572,17 et 1 095,67, ce qui représente des taux de croissance de 198 % et 191 %. Voir www.smartcomputing.com, **Ram - Need More? Then Buy More**, mars 1998, vol. 9, numéro 3.

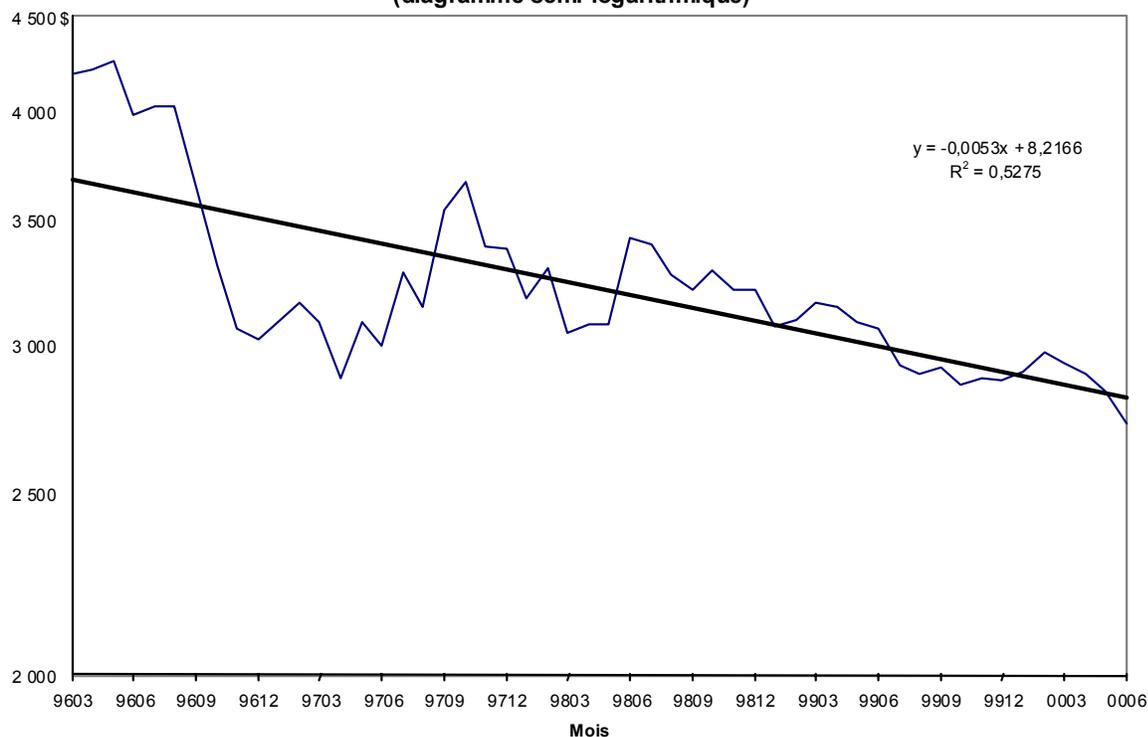
¹⁹ Les systèmes informatiques moyens dont parle ce tableau sont une **combinaison** des valeurs mensuelles moyennes de la variable du prix et des variables RAM, CPU, HD et CACHE (dans le cas des variables RAM et CACHE, les moyennes sont en arrondis à la valeur d'accroissement la plus proche).

que l'hypothèse que nous avons posée se vérifie, à savoir que le prix et la qualité sont de sens opposés et qu'un accroissement appréciable de la qualité l'emporte sur la décroissance du prix moyen en valeur absolue. Ainsi, le traitement des changements de qualité est une question primordiale à étudier lorsqu'on conçoit un indice des prix pour les ordinateurs, et peut-être pour d'autres produits de haute technologie au comportement analogue.

Tableau B – Systèmes moyens composés

Principaux éléments	Mars 1996	Juin 2000	Variation en pourcentage
Unité centrale	Pentium 133 MHz	Pentium III 600 MHz	+571,8 ²⁰
Mémoire vive	16 Mo	128 Mo	+700,0
Taille du disque dur	1,3 Go	10,9 Go	+738,5
Antémémoire étendue	256 Ko	512 Ko	+100,0
Prix	4 188 \$	2 727 \$	-34,8

**Figure 1.1 - Prix moyen
(diagramme semi-logarithmique)**



²⁰ Comparaison des cotes moyennes des unités centrales pour ces deux mois.

Figure 1.2 - Prix moyen des Pentium I, II et III et du Celeron

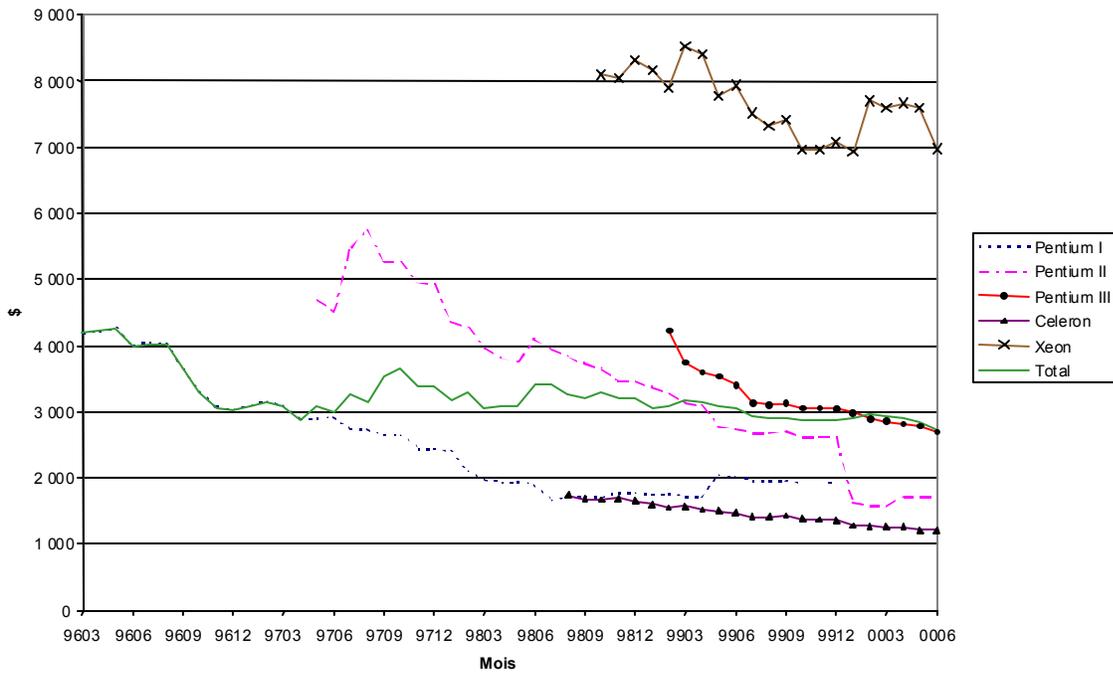
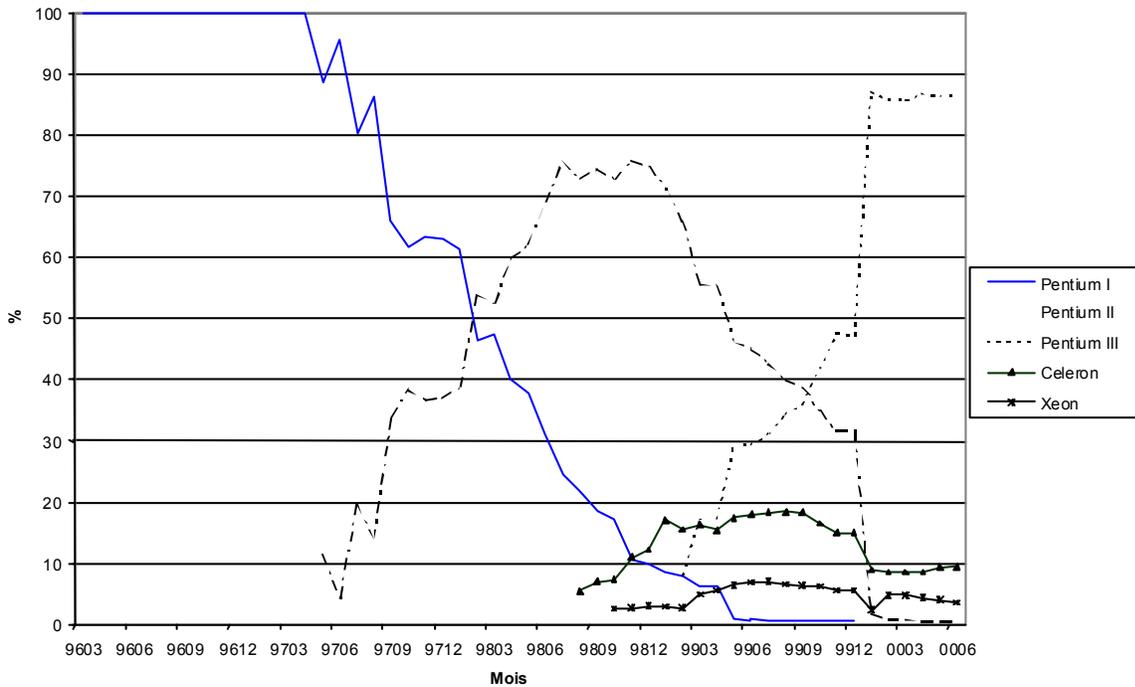
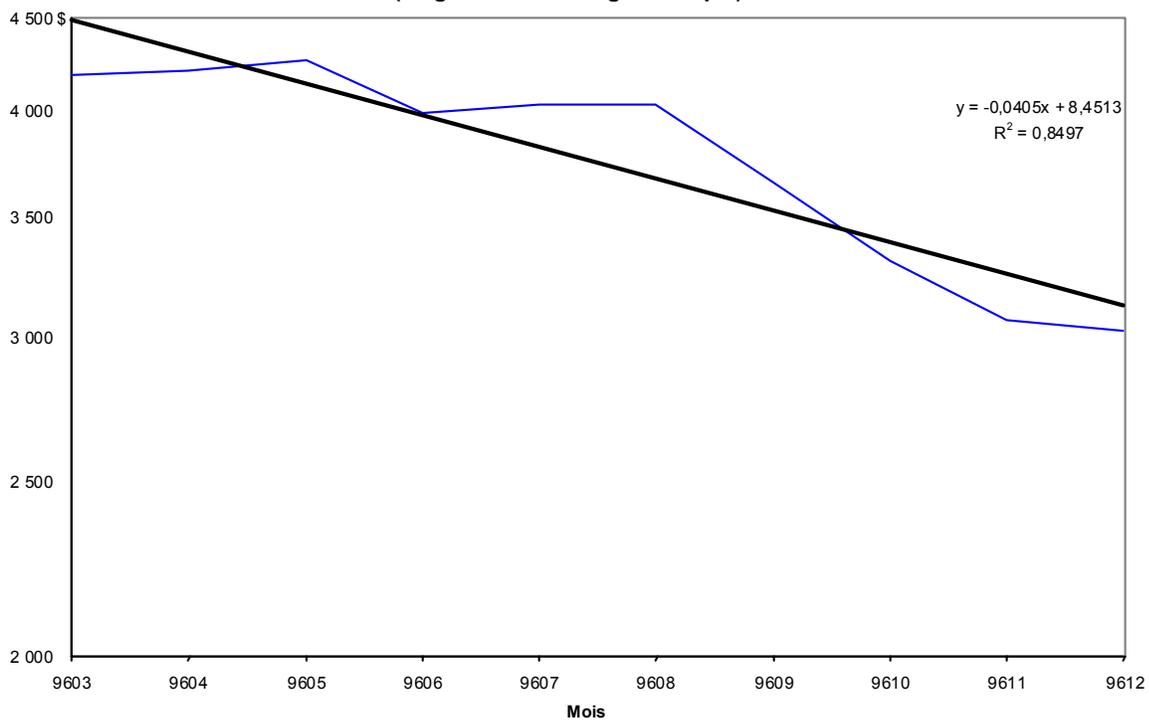


Figure 1.3 - Catégories de processeurs en proportion de l'échantillon mensuel



**Figure 1.4 - Prix moyen en 1996
(diagramme semi-logarithmique)**



**Figure 1.5 - Prix moyen en 1997
(diagramme semi-logarithmique)**

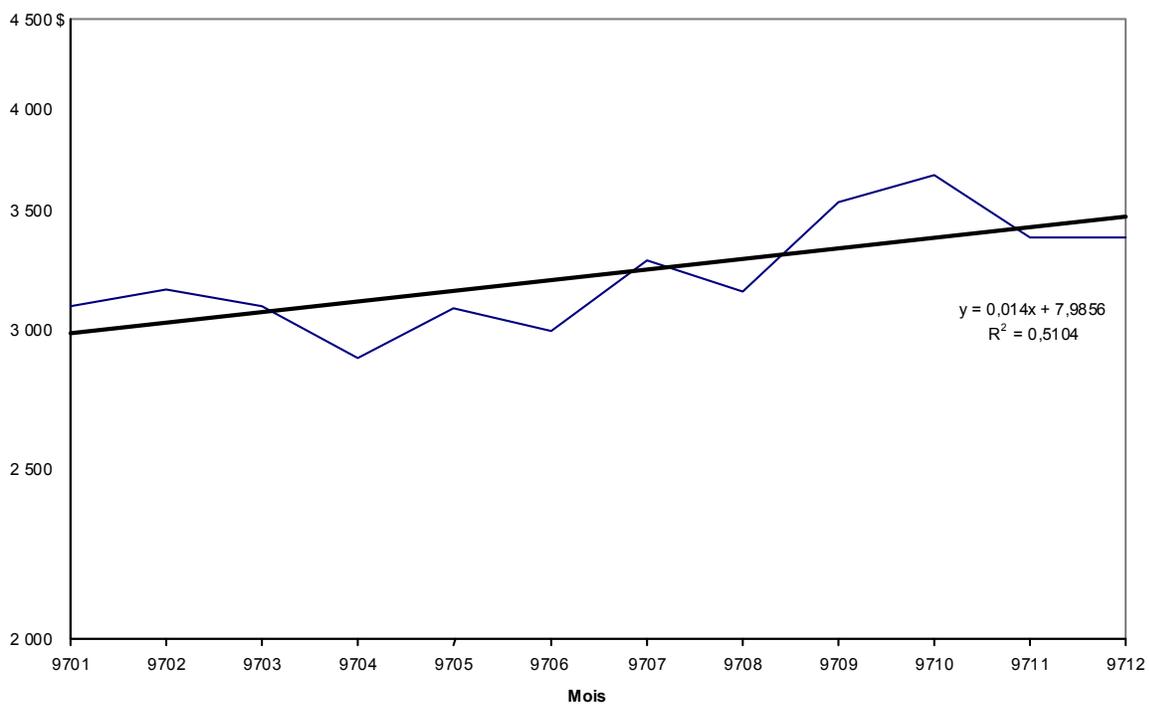


Figure 1.6 - Prix moyen en 1998
(diagramme semi-logarithmique)

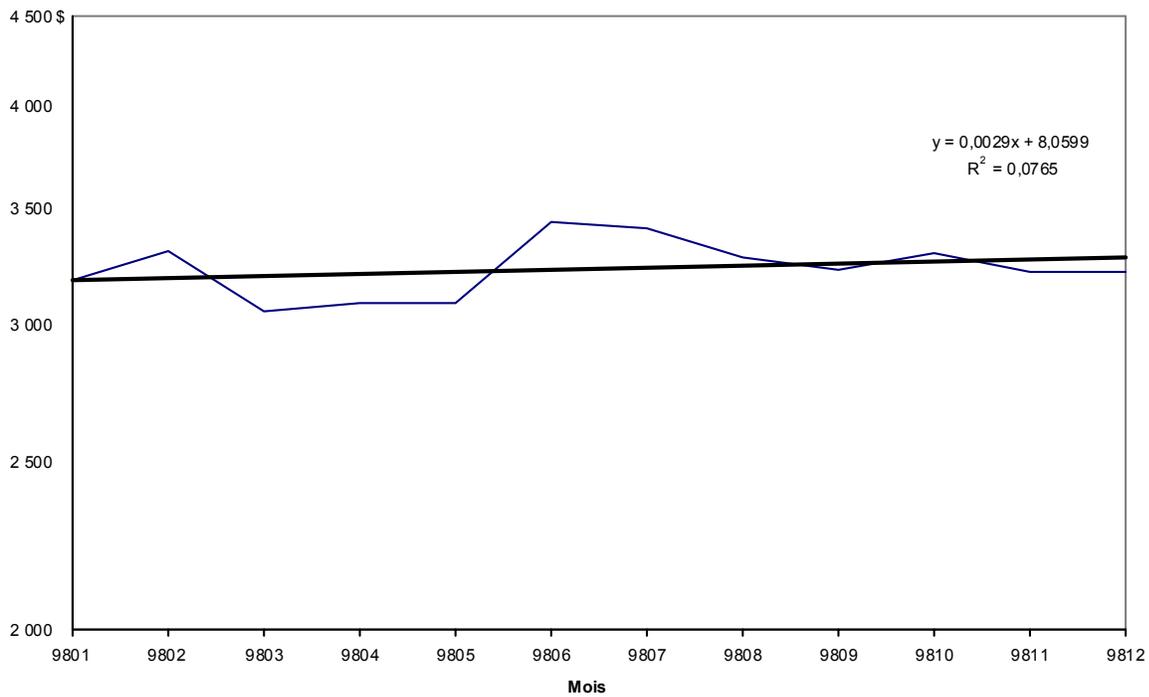


Figure 1.7 - Prix moyen de 1999 à juin 2000
(diagramme semi-logarithmique)

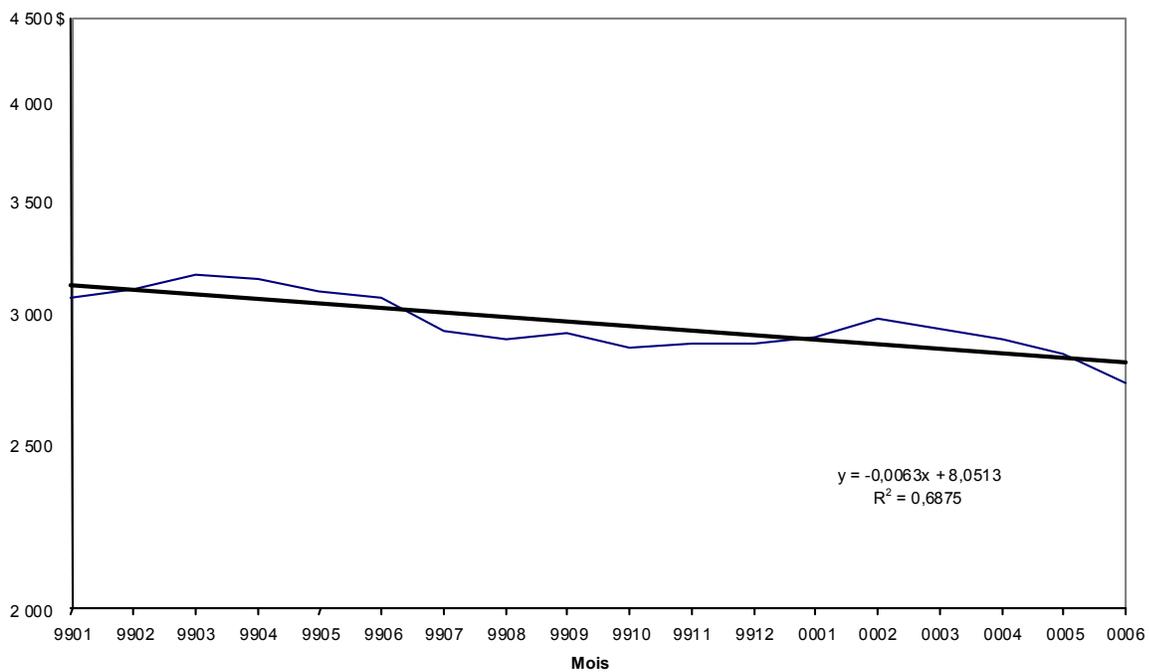


Figure 1.8 - Capacité moyenne de mémoire vive (RAM)
(diagramme semi-logarithmique)

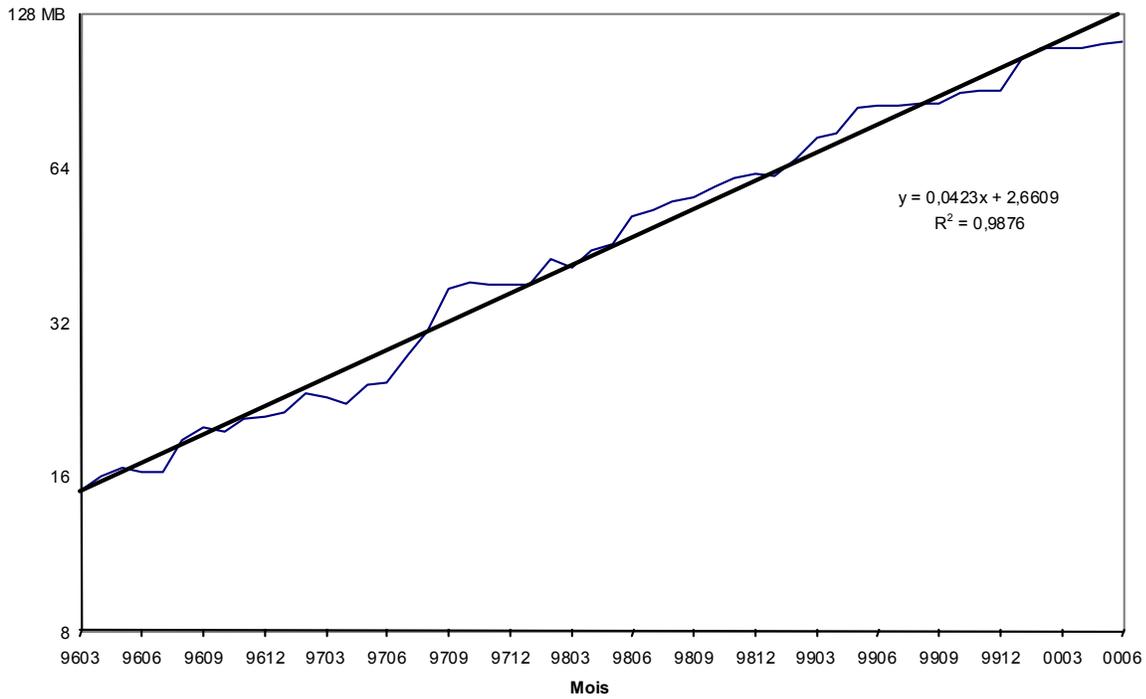


Figure 1.9 - Capacité moyenne d'unité centrale
(diagramme semi-logarithmique)

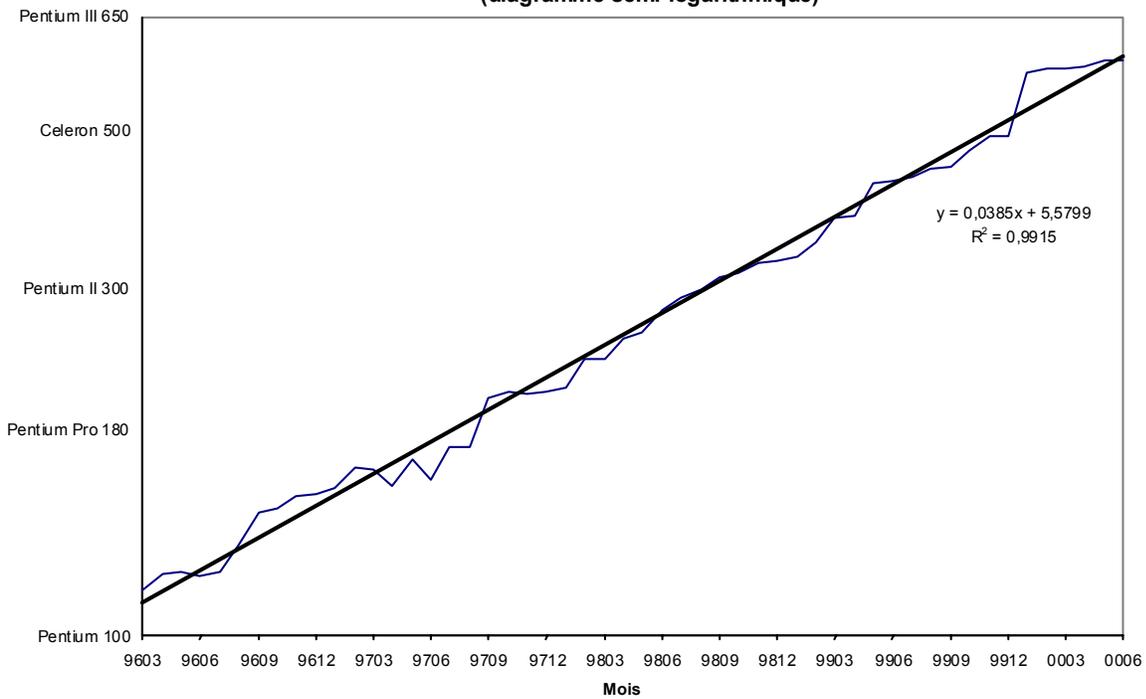


Figure 1.10 - Taille moyenne de disque dur
(diagramme semi-logarithmique)

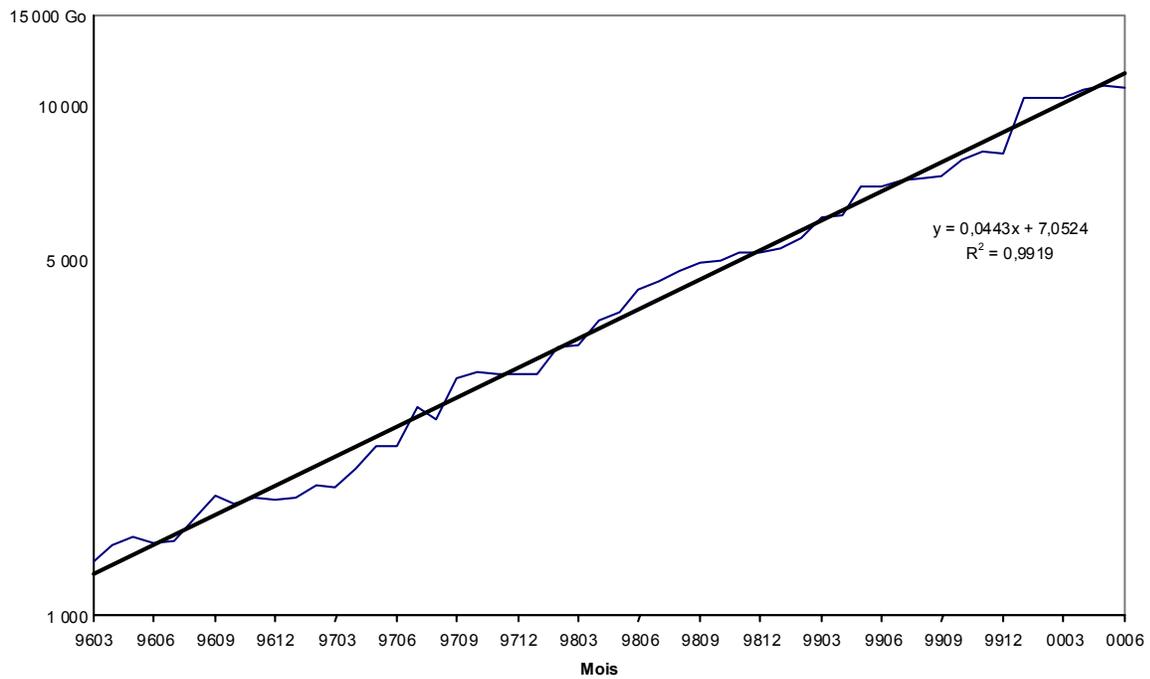
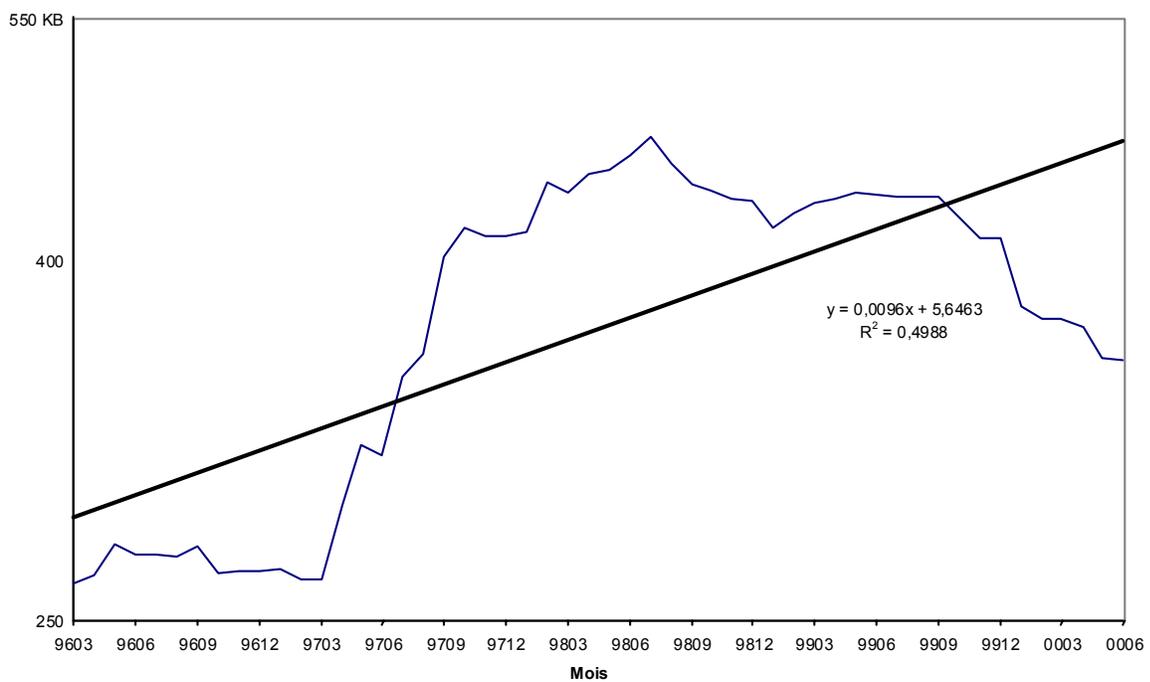


Figure 1.11 - Capacité moyenne d'antémémoire
(diagramme semi-logarithmique)



CHAPITRE 2 – Technique des échantillons appariés en chevauchement

2.1 Introduction

Dans ce chapitre, nous décrirons, appliquerons et évaluerons la méthode de détermination des prix par échantillons appariés en chevauchement aux fins de l'élaboration d'une série indiciaire des prix des ordinateurs. Cette méthode constitue une solution de rechange à la méthode actuelle de production à la Division des prix par échantillons appariés de modèles et correction hédonique de qualité pour les modèles de remplacement²¹.

Dans un grand nombre d'organismes statistiques, on procède dans bien des cas par échantillons appariés de modèles pour calculer des indices de prix²². Il s'agit de rapporter les uns aux autres les prix d'un même modèle dans des périodes successives. Comme un tel indice sera uniquement fondé sur des rapports de prix, l'appariement des modèles ne fera voir aucune évolution des caractéristiques des produits, dont on se trouve à garder la qualité constante. Cette méthode donne les meilleurs résultats dans le cas de produits homogènes dont les caractéristiques n'évoluent guère au fil des ans (on peut songer à des matières premières comme le carburant diesel).

Il se pose toutefois un problème lorsque des modèles disparaissent, disons par obsolescence. On peut alors décider soit d'accepter cette disparition et de maintenir l'échantillon de prix qui reste, soit de trouver un modèle de remplacement. Il faudra un jour trouver de ces modèles de remplacement pour que la série continue à exister²³. La perte de modèles peut ne pas être un enjeu de taille s'il s'agit d'industries fabriquant des produits peu technologisés (sciages, marteaux, rubans, etc.) où l'obsolescence est lente et où les articles de remplacement sont très proches en qualité des produits remplacés. Tel n'est pas le cas dans l'industrie de l'informatique, que l'on peut le mieux décrire comme *hétérogène*, les produits y évoluant très rapidement et la durée utile d'un ordinateur ou d'un modèle « type » se révélant très brève (elle n'est parfois que de quelques mois). Le remplacement des modèles n'est plus un exercice dénué d'intérêt.

Comment intègre-t-on alors les nouveaux modèles? Au prochain chapitre, nous examinerons la méthode hédonique où le prix d'un modèle de remplacement est en correction de qualité. Dans le présent chapitre, nous regarderons la méthode des échantillons appariés en chevauchement, où les modèles de remplacement sont tout simplement intégrés selon leur disponibilité dans des périodes d'échantillonnage en chevauchement. Un des grands avantages de cette méthode réside dans la simplicité de sa description et de son application. D'un point de vue opérationnel, on se trouve à tenir compte plus facilement des caractéristiques des nouveaux modèles que dans le cas d'une correction hédonique de qualité. Cette technique permet en outre une constante actualisation de l'échantillon de modèles (prix). La grande critique qu'on peut en

²¹ La méthode décrite et appliquée dans notre étude repose sur celle qui a été proposée par Turvey (1999). Le lecteur notera que la série indiciaire présentée est uniquement fondée sur des moyennes géométriques, ce qui permet d'éviter le dilemme **moyenne des valeurs relatives de prix-valeur relative des prix moyens**. Comme toutes les séries sont produites de cette manière, il n'y a aucune incidence sur leur comparaison.

²² Voir Triplett (2000).

²³ Les choix possibles sont notamment le rééchantillonnage et le remplacement un à un dans une base de sondage déterminée.

faire est qu'elle traite indirectement les changements de qualité, car elle fait appel à une correction de qualité implicite dont l'ordre de grandeur peut se révéler insuffisant et qui dépend d'une décroissance des prix des modèles.

Dans le reste de ce chapitre, nous donnerons un aperçu de la technique des échantillons appariés en chevauchement. Suivront les séries indiciaires estimées, une analyse de questions de méthodologie et les conclusions.

2.2 Aperçu méthodologique

Le tableau C résume la méthode des échantillons appariés en chevauchement. Si nous posons que la série commence au mois t , l'échantillon total de toute période peut se définir comme S_{t+n} et se composera des nouveaux modèles appariés (reportés de la période antérieure $t+n-1$) et non appariés, M_{t+n} et nm_{t+n} . À son tour, M_{t+n} peut se décomposer en nouveaux modèles appariés et en modèles existants appariés, nm_{t+n-1} et em_{t+n-1} , moins les modèles nouveaux et existants qui disparaissent, dnm_{t+n-1} et dem_{t+n-1} ²⁴. Si on l'exprime algébriquement, on obtient

$$M_{t+n} = nm_{t+n-1} + em_{t+n-1} - dnm_{t+n-1} - dem_{t+n-1},$$

où les modèles que contient M_{t+n} servent au calcul des valeurs relatives de prix entre $t+n$ et $t+n-1$ ²⁵.

Tableau C – Résumé de la méthode des échantillons appariés en chevauchement

Mois	t	t+1	t+2,....	t+n
Échantillon total	S_t	S_{t+1}	S_{t+2}	S_{t+n}
Modèles appariés	M_t	M_{t+1}	M_{t+2}	M_{t+n}
• Nouveaux modèles du dernier mois	nm_{t-1}	nm_t	nm_{t+1}	nm_{t+n-1}
• Autres modèles existants	em_{t-1}	em_t	em_{t+1}	em_{t+n-1}
• Nouveaux modèles qui disparaissent	$-dnm_{t-1}$	$-dnm_t$	$-dnm_{t+1}$	$-dnm_{t+n-1}$
• Modèles existants qui disparaissent	$-dem_{t-1}$	$-dem_t$	$-dem_{t+1}$	$-dem_{t+n-1}$
Nouveaux modèles non appariés	nm_t	nm_{t+1}	nm_{t+2}	nm_{t+n}

²⁴ Les nouveaux modèles qui disparaissent sont ceux de la période précédente $t+n-1$ qui n'étaient pas encore appariés à $t+n$.

²⁵ Par récursion, la formule de M_{t+n} peut se définir comme

$$M_{t+n} = M_t + \sum_{i=0}^{n-1} nm_{t+i} - dnm_{t+i} - dem_{t+i}, \text{ pour } n > 0 \text{ (à noter que, si } n=0, M_t \text{ se ramène à } nm_t).$$

Cette définition exprime M_{t+n} selon les modèles appariés de la période t et les apparitions et disparitions de modèles nouveaux et existants depuis lors. Les auteurs désirent remercier Andy Baldwin, de la Division des prix de Statistique Canada, de leur avoir fourni les bases du tableau C et les formules relatives à M_{t+n} .

En guise d'illustration, posons que, à t , on compte quatre modèles dans l'échantillon, A, B, C et D. À $t+1$, les modèles disponibles sont B, C, D, E et F, si bien que $nm_t = A, B, C$ et D, $dnm_t = A$ et $nm_{t+1} = E$ et F. Ainsi, $M_{t+1} = nm_t - dnm_t = B, C$ et D, et l'indice de prix de cette période se composera des valeurs relatives de prix des trois modèles en question.

Dans la période qui suit $t+2$, les modèles disponibles deviennent B, C, E et G, si bien que $nm_{t+1} = E$ et F, $em_{t+1} = B, C$ et D, $dnm_{t+1} = F$, $dem_{t+1} = D$ et $nm_{t+2} = G$. Ainsi, $M_{t+2} = nm_{t+1} + em_{t+1} - dnm_{t+1} - dem_{t+1} = B, C$ et E, et les valeurs relatives de ces modèles serviront au calcul de l'indice de prix entre $t+1$ et $t+2$. Et ainsi de suite.

Résumons en disant que les nouveaux modèles ne seront intégrés à l'échantillon que s'ils sont en appariement dans la période de chevauchement formée du mois en cours et du mois précédent. Comme l'illustre notre exemple, ce ne sont pas tous les nouveaux modèles qui seront nécessairement intégrés, puisqu'ils doivent être sur le marché pendant plus d'une période. Nous procédons ensuite au calcul d'un indice de prix pour les modèles de la période d'intérêt comme on le fait pour tout autre indice type de prix. Voici les résultats d'un tel exercice :

2.3 Résultats

Série indiciaire

De la manière que nous venons d'indiquer, nous avons calculé une série indiciaire des prix par moyenne géométrique en chaîne appliquée aux valeurs relatives de prix. C'est ce que présente la figure 2.1. Le tableau D livre les taux d'accroissement estimés pour toute la période, ainsi que pour les sous-périodes d'appariement dont nous avons parlé au chapitre 1. La série indiciaire fait voir un taux mensuel de décroissance de 2,85 % pour toute la période étudiée. Les prix ont le plus décliné en valeur relative dans les dix premiers mois de 1996²⁶.

Tableau D – Taux d'accroissement

Période	Valeur relative en chaîne de la moyenne géométrique (%)
Ensemble de la période étudiée	-2,85
1996	-5,58
1997	-3,31
1998	-2,33
1999 à juin 2000	-2,15

²⁶ Cette décroissance correspond à l'avènement des nouvelles puces informatiques qui a largement fait baisser les prix des ordinateurs (voir le chapitre 1).

Modification de l'échantillon

Un des avantages de la méthode d'intégration en période de chevauchement réside dans la forme que prend l'actualisation de l'échantillon. Pour le statisticien, deux questions se posent en réalité ici, celles de l'échantillonnage et des changements de qualité. Il ne faut pas *les confondre*, ainsi que nous allons le démontrer. S'il est question d'échantillonnage, il importe pour la représentativité de l'échantillon que les prix des ordinateurs courants soient déterminés pour toute période. Par sa constitution même, la méthode du chevauchement garantit le renouvellement de l'échantillon à mesure qu'apparaissent de nouveaux modèles. L'inconvénient cependant est que cette manière de procéder est peut-être trop objective, puisqu'on se trouvera à intégrer tous les modèles qui apparaissent sur le marché pour au moins deux périodes, bien qu'ils puissent ne pas être représentatifs. Bien sûr, on peut remédier à la chose en donnant plus de poids à certains modèles ou en ne sélectionnant que les modèles jugés représentatifs, mais alors notre manière de procéder se met à ressembler de plus en plus à la méthode actuelle de calcul indiciaire²⁷.

Par suite du calcul de la série indiciaire à la figure 2.1, le taux d'appariement observé en moyenne mensuelle est d'environ 88 %, ce qui veut dire que, de mois en mois, on calcule l'indice en se reportant à 9 modèles sur 10 du mois précédent (voir la figure 2.2)²⁸. C'est l'indice d'une ample continuité de l'échantillon, mais on doit garder à l'esprit que c'est là un taux d'appariement mensuel et que le phénomène des changements de modèles serait plus prononcé à intervalles trimestriels ou annuels. Le chiffre n'en demeure pas moins important, même à ce niveau. La figure 2.3 indique sommairement le nombre de nouveaux modèles en proportion de l'échantillon total de chaque mois. Bien que, en moyenne, les nouveaux modèles représentent 11,5 % de l'échantillon dans un mois quelconque, on sait bien que, dans certaines circonstances, cette proportion sera bien supérieure. Ainsi, pour la période avril-septembre 1997, la moyenne est de 51,7 %.

Comme nous l'avons mentionné, l'effet de déperdition d'échantillon est cumulatif. On le voit bien si on compare d'une année à l'autre les échantillons d'un même mois. Le taux moyen de déperdition sur 12 mois s'établit à 75,8 % (voir l'annexe II). Ainsi, 75,8 % des modèles disparaissent en l'espace de 12 mois. On voit donc le grand rôle que jouent les nouveaux modèles. Il est sûr que ceux-ci ont une incidence non négligeable sur l'échantillon en place.

En voici les principales conséquences : a) l'échantillon servant à la production de l'indice doit faire l'objet de mises à jour fréquentes; b) la substitution rapide de nouveaux modèles aux anciens doit influencer sur les caractéristiques de l'échantillon qui définissent la qualité des produits et, de là, le mouvement des prix en correction de qualité. La première est une importante conséquence quelle que soit la technique que l'on utilise, puisqu'elle fait partie de la stratégie d'échantillonnage de toute activité d'enquête. Quant à la question de la correction de qualité, nous allons la regarder de plus près.

²⁷ Voir le chapitre 4. On peut choisir de retarder l'intégration de nouveaux modèles jusqu'à ce qu'ils deviennent représentatifs, mais plus cette intégration acquiert un caractère subjectif, plus on s'écarte de la méthode des échantillons appariés en chevauchement. On finit par se retrouver avec la méthode actuelle, moins la correction de qualité.

²⁸ Il y a eu quatre mois sans nouveaux modèles en appariement.

Correction de qualité

Conceptuellement, une des hypothèses de base au microniveau de l'élaboration d'un indice de prix est que l'on doit garder constantes la qualité et/ou la quantité d'un bien ou d'un service pour pouvoir dégager les purs mouvements des prix, que notre comparaison soit temporelle ou spatiale. Idéalement, le statisticien recherchera la permanence du bien ou du service au microniveau. La pomme d'une période ou d'un lieu doit se comparer en nature, en qualité et en grosseur à la pomme d'une autre période ou d'un autre lieu si on entend parler logiquement d'une différence de prix dans son cas. Bien sûr, à des niveaux supérieurs d'agrégation, les paniers de biens ou de services ou les plans de pondération sont nombreux selon l'usage que l'on prévoit faire des données²⁹. Mais il demeure que, pour que des comparaisons significatives soient possibles, un produit doit se définir et se décrire de la même manière dans le temps et l'espace, d'où la difficulté pour le statisticien qui doit composer avec le phénomène d'évolution dans le temps de presque tous les produits (amélioration de leur qualité) et avec une situation où la version ou le modèle du moment en vient à disparaître et à faire place à un produit d'une qualité (habituellement) supérieure. Si notre échantillon est représentatif de la période en cours, la qualité évoluera et une correction de qualité quelconque s'imposera³⁰.

Le statisticien se retrouve devant l'alternative a) d'une *méthode de correction directe* (correction hédonique ou détermination des prix de caractéristiques facultatives) dans un remplacement un à un et b) d'une *méthode indirecte* (par échantillons appariés en chevauchement) en cas de remplacement en masse. Dans une correction directe, on connaît l'ampleur du changement de qualité, et l'incidence peut en être facilement évaluée. Dans un appariement d'échantillons en chevauchement, il n'y a aucun traitement direct de changements de qualité. Il s'agit plutôt de changements implicites par une évaluation marchande de la qualité des nouveaux modèles, ce qui rend la chose moins visible par rapport à la méthode de correction directe. Il reste que, comme le signale Turvey (2001), il est possible d'estimer l'évolution de la qualité qui s'opère par la méthode du chevauchement, c'est-à-dire par le calcul d'un *indice de qualité implicite* qui rapporte simplement un indice non corrigé (en fonction de la qualité) à l'indice effectif produit pour la période pour ainsi dégager la pure variation des prix. Dans le cas de la moyenne géométrique, le tout se présenterait sous la forme suivante :

$$Q_{t/t-1} = \frac{P_{t/t-1}^{non\ corrigé}}{P_{t/t-1}^{effectif}}$$

où

$$P_{t/t-1}^{non\ corrigé} = \frac{\prod_{i=1}^n (P_t^i)^{1/n}}{\prod_{j=1}^m (P_{t-1}^j)^{1/m}}, \quad P_{t/t-1}^{effectif} = \prod_{i=1}^n \left(\frac{P_t^i}{P_{t-1}^i} \right)^{1/n},$$

²⁹ Pour agréger des produits en indices, on peut prendre des indices à pondération fixe ou mobile, des indices de Fisher, etc.

³⁰ On a toujours la faculté de ne pas changer l'échantillon et de continuer à déterminer les prix des mêmes modèles, quelque peu représentatifs qu'ils puissent devenir. On évite le problème des changements de qualité pour un temps seulement, car les anciens modèles finissent par disparaître.

ce qui se ramène à

$$Q_{t/t-1} = \frac{\prod_{i=1}^n (P_{t-1}^i)^{1/n}}{\prod_{j=1}^m (P_{t-1}^j)^{1/m}}$$

Dans ce cas, $Q_{t/t-1} > 1$ lorsque les prix des modèles effectifs l'emportent sur ceux des modèles de l'échantillon non corrigé dans la même période. Cela veut dire que la qualité s'améliore à mesure que les nouveaux modèles – dans une hypothèse « prix supérieur à qualité supérieure » – se substituent à des modèles moins chers qui disparaissent. Toutefois, si cette dynamique des prix n'a pas lieu et que les prix des nouveaux modèles sont égaux ou inférieurs à ceux des modèles qui disparaissent, $Q_{t/t-1} \leq 1$, d'où l'implication d'une absence d'amélioration ou d'une détérioration de la qualité. Cela peut être le cas ou ne pas l'être du tout selon que l'hypothèse « prix supérieur à qualité supérieure » se vérifie dans les faits. Comme l'indique Turvey (2001), l'indice sera entaché d'un biais si l'arrivée de nouveaux produits d'une meilleure qualité ne fait pas reculer le prix des modèles en disparition ou que les nouveaux modèles qui entrent sur le marché sont plus chers sans que la qualité ait évolué outre mesure.

Pour nous faire une idée du degré de correction de qualité effective par la méthode du chevauchement, nous avons calculé un indice de qualité implicite et l'avons comparé à la série indiciaire en moyenne géométrique pour les variables RAM et CPU. Les chiffres se rapportent aux modèles de l'échantillon servant au calcul de l'indice effectif³¹. Si cette méthode permet une correction de qualité suffisante, on peut prévoir que les trois séries auront des mouvements assez convergents, bien que sachant que d'autres caractéristiques qualitatives comme la taille du disque dur et la capacité de l'antémémoire ne sont pas incluses dans l'analyse.

Ces séries indiciaires sont présentées à la figure 2.4. Les résultats font voir une divergence croissante entre les indices RAM et CPU et l'indice de qualité implicite. Sauf pour le premier mois ou les deux premiers, le niveau de qualité dégagé semble en sous-estimation par la méthode du chevauchement et le problème s'aggrave à mesure qu'avance la période. On peut expliquer ce phénomène en regardant les comportements des nouveaux modèles en soi. Si nous élaborons et comparons des séries indiciaires semblables pour les nouveaux modèles (voir la figure 2.5), nous constatons que l'indice de prix n'évolue guère en hausse, oscillant plutôt autour d'une valeur moyenne de 111,3. Quant aux séries indiciaires RAM et CPU, elles sont en très forte progression dans le sens même des tendances dont nous avons parlé au chapitre 1. Dans le temps, les prix des nouveaux modèles ne traduisent pas nécessairement l'évolution de leur qualité, ce qui introduit un biais par défaut dans la mesure de la correction de qualité.

³¹ Ces trois séries indiciaires sont produites en chaîne. Ajoutons que, pour les quatre mois sans arrivée de nouveaux modèles, les valeurs du mois précédent ont été substituées dans le calcul indiciaire en chaîne, ce qui n'influera pas outre mesure sur l'analyse puisque le mouvement de l'indice n'a pas changé dans les trois séries pendant cette période.

Comparaison des modèles nouveaux et anciens

Pour mieux cerner ce biais, nous avons comparé les nouveaux modèles qui entrent dans l'échantillon aux anciens modèles qui le quittent pour la variable du prix et les variables RAM et CPU. Dans l'application de la méthode du chevauchement, le rapport de prix entre les modèles nouveaux de remplacement et les modèles anciens en disparition devrait refléter la différence foncière de qualité dans un remplacement en masse. Pour pousser le raisonnement, disons que la variation en pourcentage du prix moyen entre modèles nouveaux et anciens devrait correspondre à la variation en pourcentage de leur qualité moyenne. Ainsi, si la mémoire vive et l'unité centrale constituent les principaux indicateurs de qualité, on devrait s'attendre à des valeurs semblables de ces variables pour les périodes où des modèles nouveaux et anciens sont observés. Ainsi, si le prix moyen des nouveaux modèles est de 10 % supérieur dans un mois à celui des anciens modèles remplacés, on devrait voir une augmentation correspondante – du même ordre de grandeur environ – des capacités de mémoire vive et d'unité centrale. La correspondance ne sera pas parfaite et, en fait, les différences en pourcentage devraient être moindres pour les variables RAM et CPU, car on ne se trouve pas à tenir compte d'autres caractéristiques qualitatives³².

Nous avons calculé ces variations en pourcentage par la moyenne géométrique de la variable du prix et des variables RAM et CPU pour les modèles nouveaux et anciens dans les 38 mois où on a observé des modèles tant nouveaux qu'anciens. Nous présentons les résultats à la figure 2.6³³. En ce qui concerne le mouvement de l'indice, l'évolution des prix en pourcentage ressemble le plus souvent à celle des capacités de mémoire vive et d'unité centrale³⁴. La différence de prix en pourcentage le cède en réalité aux différences en pourcentage des variables RAM et CPU dans la majorité des mois (27 et 26 mois ou 71 % et 68,4 %). Ces différences sont aussi décrites aux figures 2.7 et 2.8 avec leur importance et leur variation. Comme le font voir ces résultats, la différence est négative dans la plupart des cas et peut être appréciable dans certains, dépassant les -20 % dans 14 périodes pour la variable CPU (ou 36,8 %) et se chiffrant à 42,1 % dans 16 périodes pour la variable RAM. Il est clair que cette méthode implicite de correction de qualité sous-estime l'ampleur des changements de qualité.

2.4 Conclusion

Voici en bref les principaux avantages de la méthode d'appariement d'échantillons en chevauchement que nous exposons dans cette étude : elle est simple à appliquer, elle actualise constamment l'échantillon et – nous n'en avons pas beaucoup parlé ici – intègre bien plus facilement les nouveaux modèles aux caractéristiques différentes que la méthode hédonique.

Malgré ces avantages, elle accuse une grande faiblesse d'ordre conceptuel en ce que les différences de qualité entre modèles ne font nullement l'objet d'un traitement direct, mais d'un

³² Si les variations effectives en pourcentage des variables RAM et CPU étaient, disons, toutes deux de 7 %, on pourrait dire qu'une variation supérieure des prix surestime le changement de qualité, mais tel pourrait ne pas être le cas s'il y avait aussi eu amélioration de la taille ou du type de disque dur.

³³ Aux figures 2.6, 2.7 et 2.8, nous prenons des périodes plutôt que des mois pour présenter les données, car nous avons retranché les mois où il n'y avait pas de modèles nouveaux qui s'ajoutaient aux anciens.

³⁴ Les coefficients estimés de corrélation entre la variable du prix, d'une part, et les variables respectives RAM et CPU étaient de 0,83856 et 0,63892.

traitement implicite qui peut faire sous-estimer les changements de qualité. Qui plus est, la méthode d'échantillonnage est peut-être trop objective. Tous les nouveaux modèles sont-ils suffisamment représentatifs pour être inclus si tôt dans l'échantillon? Avec cette méthode, il faut répondre oui, mais vu l'importance de ces nouveaux modèles dans l'ensemble de l'échantillon, on doit se demander quand ou si on doit tous les inclure (on peut peut-être songer à une période de chevauchement plus longue). Disons enfin de la question des nouvelles caractéristiques des modèles de remplacement que le problème n'a rien d'insurmontable dans le cas d'un traitement hédonique et qu'il faut tout simplement y mettre un plus grand effort. Ce seul point ne suffit malheureusement pas à recommander la méthode de l'appariement d'échantillons en chevauchement par rapport à la méthode hédonique dans le cas des indices de prix des ordinateurs.

Figure 2.1 - Indice par échantillons appariés en chevauchement
(diagramme semi-logarithmique)

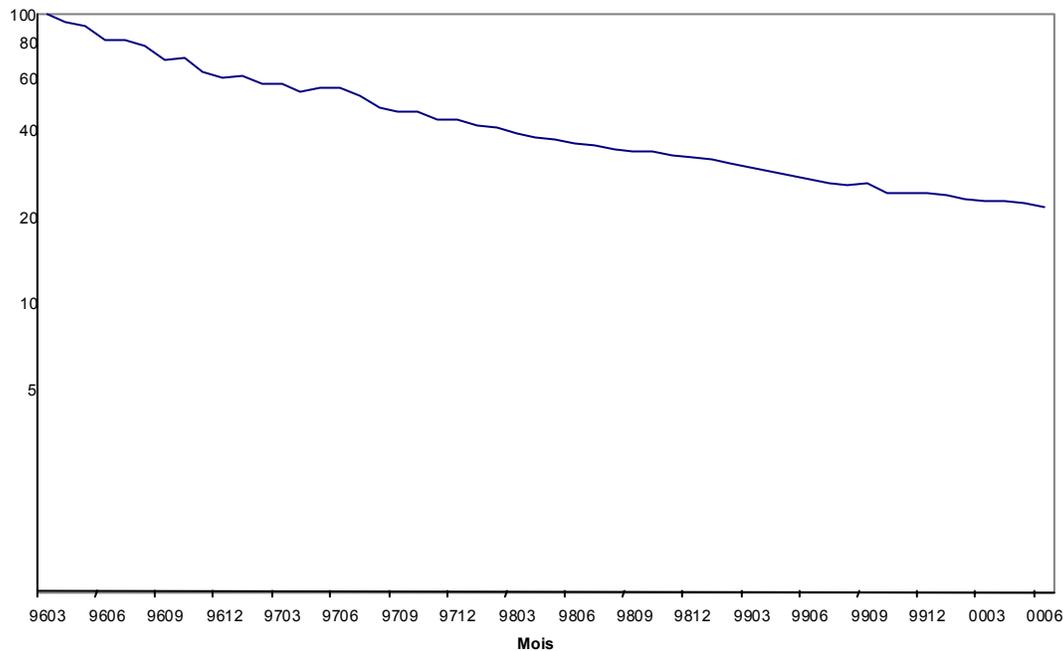


Figure 2.2 - Résultats d'appariement

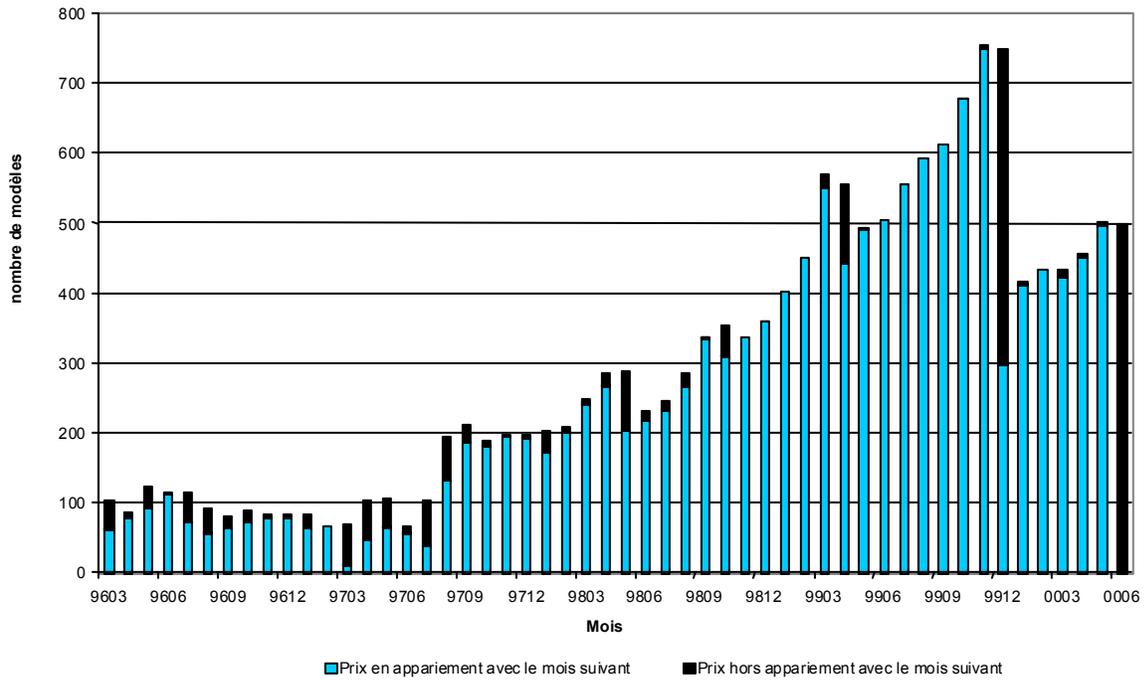


Figure 2.3 - Nouveaux modèles en proportion de tout l'échantillon mensuel

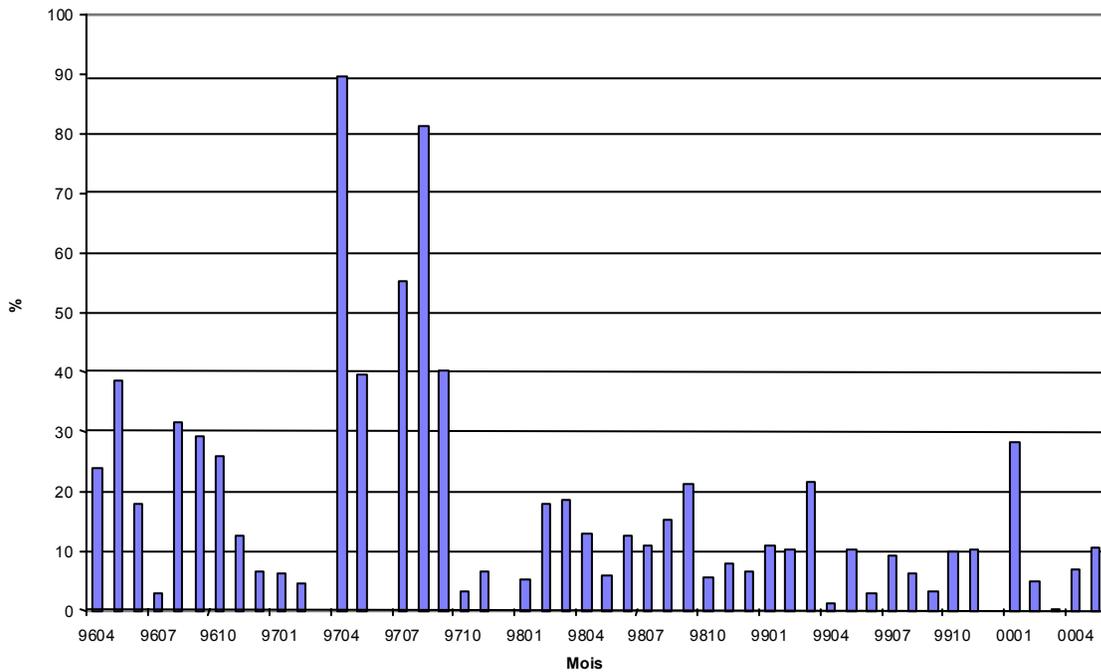


Figure 2.4 - Indice de qualité implicite et indices RAM et CPU

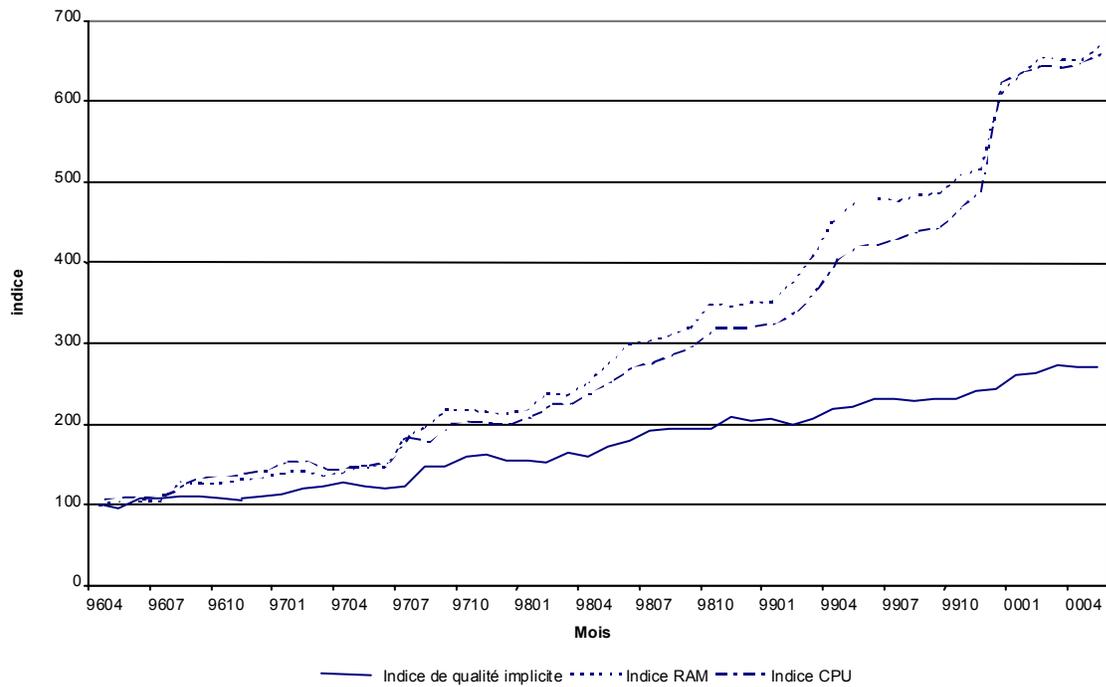


Figure 2.5 - Indice de prix et indices RAM et CPU – nouveaux modèles

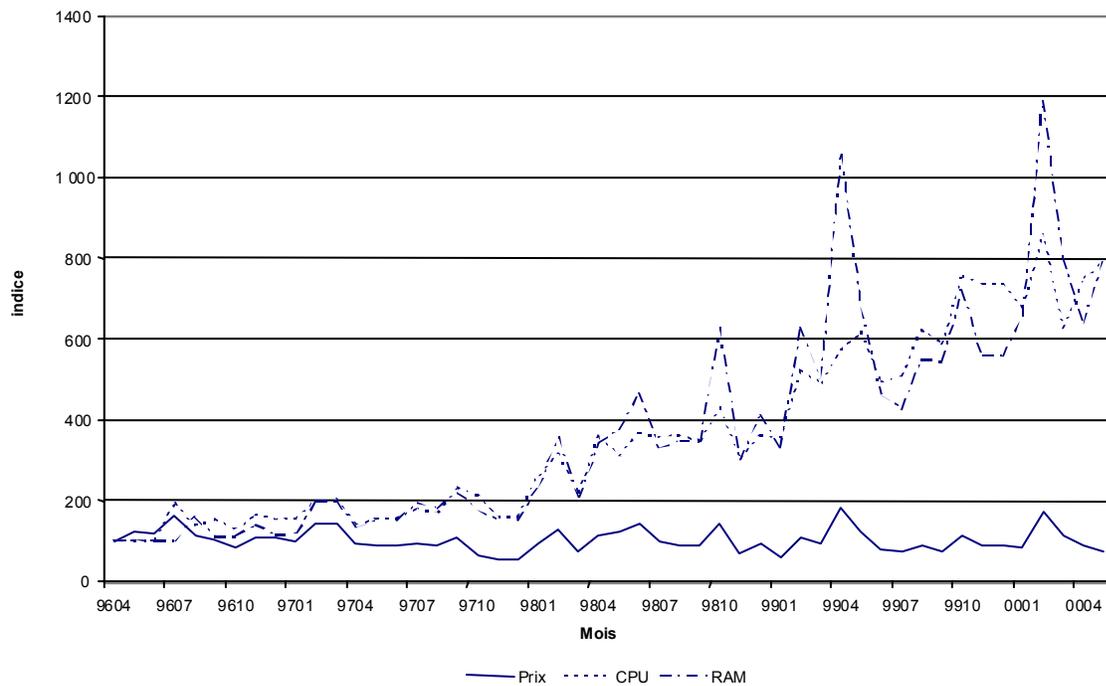


Figure 2.6 - Prix, CPU et RAM des modèles nouveaux et anciens (en pourcentage)

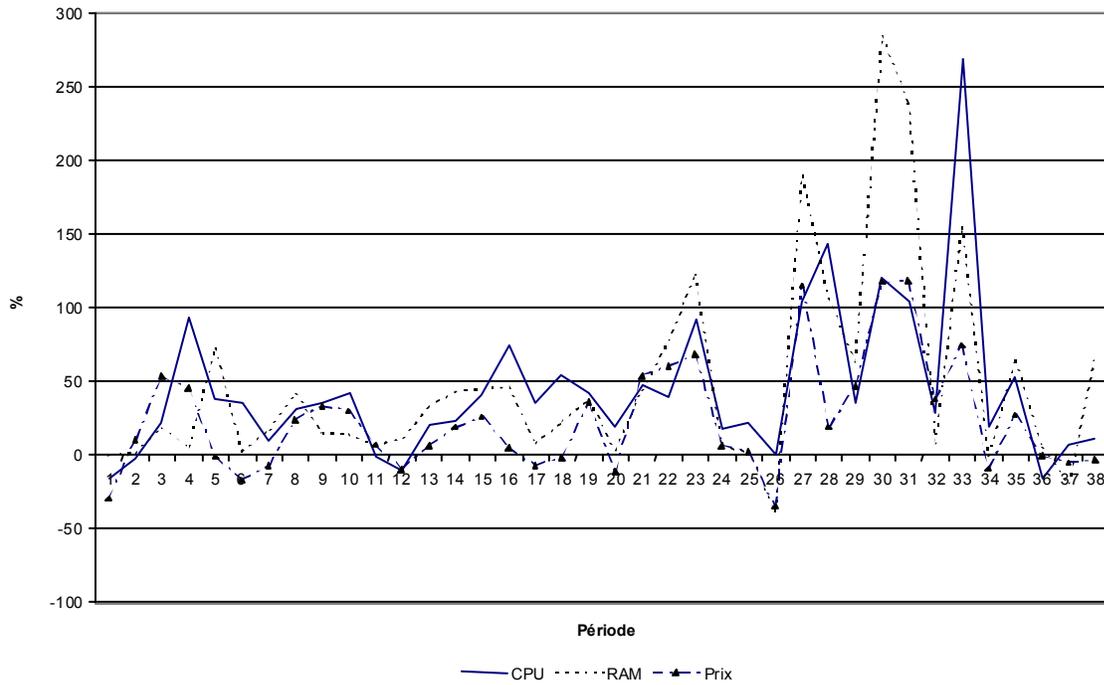


Figure 2.7 - Différence de prix entre les modèles nouveaux et anciens (en pourcentage) – variable RAM (en pourcentage)

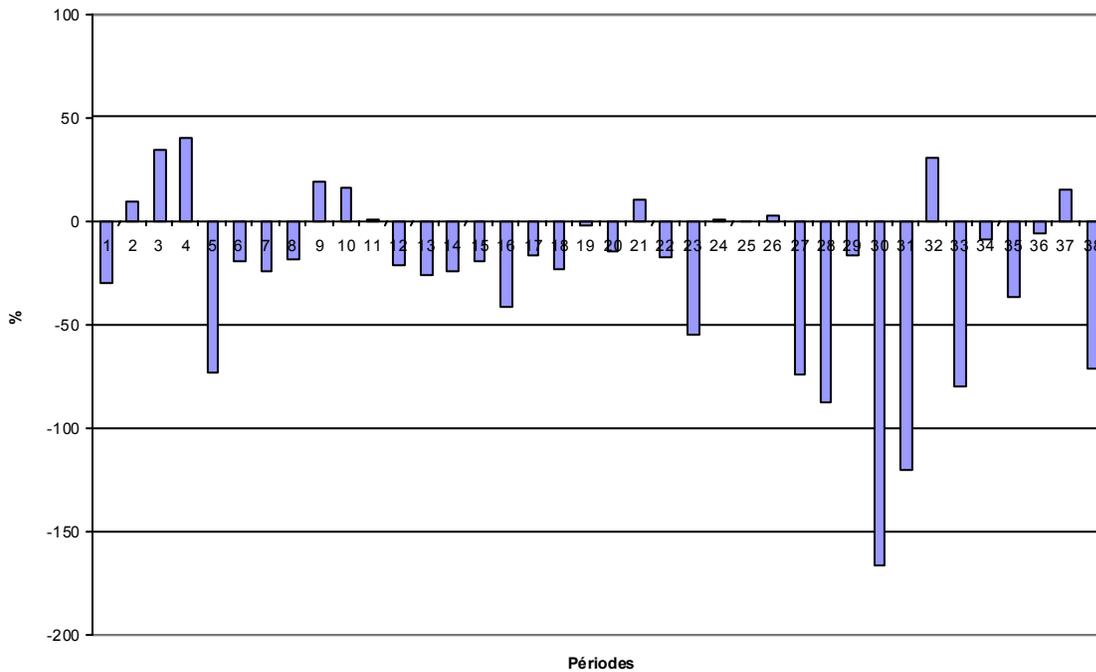
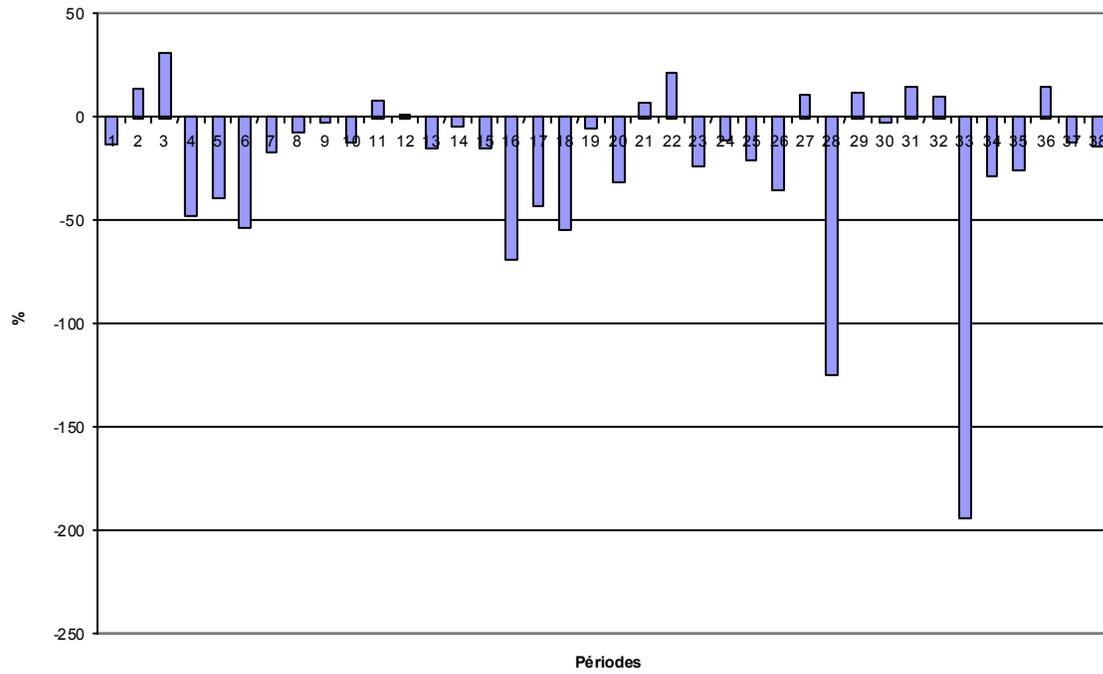


Figure 2.8 - Différence de prix entre les modèles nouveaux et anciens (en pourcentage) – variable CPU (en pourcentage)



CHAPITRE 3 – Méthodologie hédonique

3.1 Introduction

Dans ce chapitre, nous examinerons l'application de la méthodologie hédonique au traitement des changements de qualité des ordinateurs, le but étant d'évaluer le rendement de ce cadre méthodologique et de voir comment il influe sur le calcul de l'indice des prix des ordinateurs. Voici les grandes questions à se poser : quel est le degré d'applicabilité de cette technique de traitement de la qualité dans la pratique? peut-on s'en servir régulièrement et, si oui, dans quelles conditions?

Généralement parlant, la méthode hédonique vise à expliquer les variations du prix d'un bien ou d'un service en les rapportant aux variations des caractéristiques de ce bien ou de ce service³⁵. Dans le cas des ordinateurs, le lien s'établit entre leur prix et leurs divers éléments ou caractéristiques (mémoire vive, vitesse d'exécution de l'unité centrale, taille du disque dur, etc.) avec un certain nombre de variables qualitatives (fournisseur, présence d'un disque compact ou d'une unité de disque compact, d'un modem, etc.). Le principal avantage de cette méthode est qu'elle aborde de front la question des changements de qualité et repose sur une théorie. D'autres techniques visent à un traitement indirect des changements de qualité (par des corrections d'échantillon, par exemple) et sont plus axées sur une procédure à appliquer³⁶. Un des grands inconvénients de la méthode hédonique est qu'elle dépend étroitement d'une grande qualité des données et d'un nombre suffisant d'observations pour que ses résultats soient statistiquement significatifs. Il y a aussi le fait que, avec des méthodes d'analyse de régression servant à l'estimation des équations, on court toujours le risque d'obtenir des résultats médiocres ou inexploitable. En réalité, on peut avoir à disposition de ressources appréciables pour produire de bons résultats si on compare cette méthode à d'autres techniques de traitement de la qualité. Il reste qu'elle est actuellement employée ou envisagée par plusieurs autres organismes statistiques (Bureau of Labor Statistics, Banque du Japon, Institut national de la statistique et des études économiques (INSEE), Australian Bureau of Statistics, etc.).

Les modes d'emploi de la méthodologie hédonique sont divers pour le traitement des changements de qualité, mais ils se rangent généralement dans deux catégories, celles des indices hédoniques directs et des indices par appariement de modèles et correction hédonique de qualité. Les premiers sont des séries indiciaires qui se calculent à l'aide des estimations directement tirées d'une équation de régression. Dans le traitement par modèles appariés, il y a appariement des modèles de période en période, et on recourt à des équations hédoniques pour imputer les prix de modèles de remplacement en correction de qualité seulement lorsque des modèles disparaissent de l'échantillon. La Division des prix de Statistique Canada procède actuellement par appariement. Dans notre étude, il sera question de l'une et l'autre de ces techniques hédoniques.

³⁵ Pour un examen approfondi de la correction hédonique de qualité, le lecteur devrait consulter le chapitre 4 de *The Practice of Econometrics, Classic and Contemporary* de Ernst R. Berndt (1991).

³⁶ Au chapitre précédent, nous avons évoqué une technique semblable, à savoir la **méthode d'appariement d'échantillons en chevauchement**.

Dans le reste du chapitre, nous examinerons en général le traitement hédonique sous l'angle des résultats d'ajustement de modèle, puis présenterons deux méthodes de calcul d'un indice hédonique direct. Suivront une illustration de la méthode d'appariement de modèles, l'élaboration d'un indice hypothétique par cette méthode et nos conclusions.

3.2 Techniques hédoniques – observations générales et résultats

Avec la méthodologie hédonique, il y a la réelle difficulté d'estimer un modèle approprié, c'est-à-dire un modèle qui appréhende la relation entre les prix et les éléments ou caractéristiques des produits en toute précision. L'estimation d'équations hédoniques pose le grand problème du choix d'une forme fonctionnelle, lequel déterminera la qualité des résultats (pertinence, exactitude, fiabilité, vraisemblance, etc.). D'habitude, ce choix se limite aux modèles linéaire, semi-logarithmique et bilogarithmique ou à une version quelconque de ce dernier modèle (il y aura, par exemple, une seule variable explicative qui sera en expression logarithmique), bien qu'on ait aussi proposé des formes appelant une estimation non linéaire³⁷. Pour bien choisir, on peut appliquer plusieurs critères : analyse graphique, signes et valeurs des coefficients estimés, statistiques logarithmiques de vraisemblance, R^2 corrigé, statistiques de test formel de forme fonctionnelle comme la statistique de régression artificielle dite de double longueur (DLR ou DLAR), résultats d'une estimation de Box-Cox, résultats de « jeux d'essai » hypothétiques, etc.³⁸

Pour notre propos, nous voulions principalement juger du degré de difficulté d'ajustement de modèle pour chaque mois de la période, de la faisabilité de l'exercice et de l'effort à consentir. Nous désirions aussi savoir si les modèles visés produiraient des résultats exploitables. En comparant le degré de difficulté d'emploi de cette méthode aux résultats qu'on en obtient, nous établissions en réalité les compromis liés à la méthode hédonique dans la pratique.

Résultats

Comme nous l'avons mentionné, il s'agissait d'abord d'essayer d'ajuster une équation pour chaque mois d'observation. Nous l'avons fait au départ avec les deux grandes variables explicatives RAM et CPU. Nous avons estimé diverses formes fonctionnelles pour les équations mensuelles (versions linéaire, semi-logarithmique, loglinéaire et bilogarithmique)³⁹. Nous avons ensuite estimé des modèles par plusieurs autres variables importantes pour voir si l'ajustement ne

³⁷ Voir Triplett (1987).

³⁸ Voir Davidson et MacKinnon (1993).

³⁹ Ces formes fonctionnelles se définissent ainsi :

modèle linéaire
$$P_i = \beta_0 + \sum_{j=1}^n \beta_{ij} X_{ij}$$

modèle semi-logarithmique
$$\ln P_i = \beta_0 + \sum_{j=1}^n \beta_{ij} X_{ij} \text{ ou } P_i = \beta_0 + \sum_{j=1}^n \beta_j \ln X_{ij}$$

modèle loglinéaire
$$\ln P_i = \beta_0 + \sum_{j=1}^n \beta_j \ln X_{ij} + \sum_{k=1}^l \gamma_k Z_{ik}$$

modèle bilogarithmique
$$\ln P_i = \beta_0 + \sum_{j=1}^n \beta_{ij} \ln X_{ij}$$

où Z_k représente un jeu de variables linéaires.

pourrait pas être meilleur. Il s’agissait en second lieu d’appliquer les modèles dégagés au traitement d’un changement de qualité hypothétique dans chaque mois, et ce, en traitant un changement réaliste dans ces modèles. Nous avons enfin voulu évaluer le réalisme des changements de qualité en question.

Ajustement de modèle par les seules variables RAM et CPU

Les résultats de la seule utilisation de ces deux variables clés sont impressionnants. Non seulement on trouve un modèle pour chacun des 52 mois de l’ensemble de données, mais le degré d’ajustement, comme il est grossièrement mesuré par le R^2 corrigé, est relativement élevé (voir la figure 3.1)⁴⁰. En moyenne mensuelle, le R^2 corrigé est de 0,7282 avec un minimum de 0,5491 et un maximum de 0,8782, ce qui indique que, dans le pire des cas, la seule prise en compte des différences des variables RAM et CPU est de nature à expliquer dans une proportion de 54,91 % les différences de prix entre modèles d’ordinateurs dans un mois. Le niveau d’explication s’établit en moyenne à 72,82 %.

Le tableau E présente la répartition des formes fonctionnelles choisies. Celle qui domine est la transformation loglinéaire où la variable du prix est transformée et où soit la variable RAM soit la variable CPU est transformée par le logarithme naturel. Cette forme fonctionnelle rend compte de 31 mois sur 52 (proportion approximative de 60 %).

Tableau E – Formes fonctionnelles pour les variables RAM et CPU

Forme fonctionnelle	Mois	Pourcentage
Linéaire	7	13,5
Semi-logarithmique	5	9,6
Loglinéaire	31	59,6
Bilogarithmique	9	17,3
Total	52	100,0

Ajustement du modèle avec des variables supplémentaires

Après avoir ajusté un modèle chaque mois pour les seules variables RAM et CPU, nous avons ajouté des variables, surtout HD et CACHE, mais d’autres aussi comme le type de fournisseur ou de disque dur. Le tableau F indique la fréquence des modèles. Le nombre de variables qui revenait le plus était quatre, bien que certains modèles offrent une bonne qualité d’ajustement avec trois variables seulement et que quelques-uns exigent jusqu’à six variables explicatives⁴¹. Bien sûr, les variables ainsi ajoutées élevaient nettement le degré d’ajustement (voir la figure 3.2). Ici, le R^2 corrigé moyen devient 0,8230 et varie de 0,7549 à 0,9066. Une fois de plus,

⁴⁰ La caractérisation de l’ajustement d’un modèle particulier est souvent subjective, mais il est raisonnable de s’attendre à ce que les signes des coefficients estimés correspondent aux « espérances » *a priori* et que l’équation estimée puisse expliquer au moins dans une proportion « significative » les variations de la variable dépendante. Que « **significatif** » veuille dire 50 %, 60 % ou 80 % de la variance des prix est laissé au jugement du lecteur.

⁴¹ Il n’y a qu’un modèle qui ait demandé plus de six variables explicatives (huit en l’occurrence).

le modèle loglinéaire, où au moins une des variables importantes est transformée, a été le plus souvent choisi. Il rendait compte de 43 mois sur 52 (proportion de 82,7 %).

Tableau F – Formes fonctionnelles avec des variables supplémentaires

Forme fonctionnelle	Mois	Pourcentage
Linéaire	1	1,9
Semi-logarithmique	8	15,4
Loglinéaire	43	82,7
Bilogarithmique	0	0,0
Total	52	100,0

Changements de qualité

Une fois trouvés des modèles d'un assez bon ajustement pour chacun des mois, il fallait juger du réalisme de ces résultats en se servant des équations estimées pour dégager un changement hypothétique de qualité dans chacun des mois en question. Nous avons choisi un modèle initial et un modèle de remplacement pour illustrer un changement de modèle qui se produirait normalement dans un mois, de sorte que la différence de qualité entre les deux modèles ne soit pas aussi grande que s'il s'agissait d'un intervalle annuel. Dans ce dernier cas, il faudrait prévoir une plus grande différence de qualité en raison de la rapidité de l'évolution technologique.

Le tableau G présente les résultats de ce changement hypothétique de qualité. Il contient les variables ou les caractéristiques qui évoluent entre le modèle initial et le nouveau modèle de remplacement, ainsi que le prix du premier et du second, le prix du nouveau modèle en correction de qualité et les variations en pourcentage entre le nouveau et l'ancien modèle et pour le nouveau modèle avant et après correction de qualité. Comme il ressort des résultats, le prix en correction de qualité fait voir une plus grande décroissance ou un plus petit accroissement que si on avait tout simplement négligé le changement de qualité. En moyenne, la variation de prix est de -5,8 % pour toute la période en correction de qualité et la moyenne est de 34,5 % sans cette correction. Les mouvements sont donc fort différents. Si nous considérons un certain nombre de changements individuels de modèles à des fins d'illustration, nous constatons qu'une augmentation de mémoire vive, d'unité centrale et de nombre de baies d'unités en août 1996 donne lieu à un renchérissement de 0,7 % en correction de qualité, bien moins qu'un renchérissement nominal de 86,5 %. De même, un changement d'unité centrale, de disque dur et de type de fournisseur en juin 1997 donne une décroissance de 32,7 % après correction de qualité; c'est plus que la diminution « nominale » de 13,4 %. La différence entre ces deux exemples peut sembler importante (85,8 % et 19,3 % respectivement), mais on doit l'examiner en sachant que plusieurs grandes caractéristiques du modèle ont aussi changé, ce qui en fait une correction de qualité appréciable. Un dernier exemple : si dans un modèle de remplacement arrivé en avril 1997 il n'y a que l'unité centrale qui ait changé, la différence sera bien moindre, soit une diminution de 0,6 % par rapport à une augmentation de 5,8 % pour une différence totale de 6,4 %. Dans l'ensemble, les résultats de l'utilisation de ces modèles sont les résultats prévus, l'ordre de grandeur et le sens des changements de qualité variant en fonction directe de l'importance et du sens des changements de caractéristiques, ce dont on ne s'étonnera pas compte tenu du degré d'ajustement des modèles au départ.

Tableau G – Récapitulation des changements hypothétiques de qualité

Date	Évolution des variables de l'ancien au nouveau modèle					Prix de l'ancien modèle	Prix du nouveau modèle	Variation en pourcentage du nouveau à l'ancien	Prix du nouveau modèle en correction de qualité	Variation en pourcentage pour le nouveau modèle avant et après correction de qualité
	CPU	RAM	CACHE	HD	Autres variables					
199603				x		4 258	3 596	-15,5	4 336	-17,1
199604	x			x		3 754	3 456	-7,9	4 271	-19,1
199605	x		X			2 767	4 078	47,4	3 619	12,7
199606	x	x				3 547	5 947	67,7	5 782	2,9
199607	x		X			5 570	7 272	30,6	8 165	-10,9
199608	x	x			N ^{bre} de baies d'unités	3 406	6 353	86,5	6 306	0,7
199609	x				N ^{bre} de baies d'unités	2 643	2 632	-0,4	3 241	-18,8
199610	x	x				3 597	3 585	-0,3	4 454	-19,5
199611	x	x				2 850	3 561	24,9	3 782	-5,8
199612	x	x				2 131	3 023	41,9	2 905	4,1
199701	x	x			Type de disque dur	2 935	5 799	97,6	4 884	18,7
199702	x	x				3 430	3 893	13,5	4 563	-14,7
199703	x	x				2 867	3 884	35,5	3 864	0,5
199704	x					3 267	3 455	5,8	3 477	-0,6
199705	x		x			1 812	2 711	49,6	2 471	9,7
199706	x			x	Fournisseur	3 975	3 443	-13,4	5 117	-32,7
199707	x			x	Fournisseur	2 813	2 875	2,2	3 443	-16,5
199708	x	x				2 233	3 327	49,0	3 603	-7,6
199709	x				Modem	2 111	3 318	57,2	2 905	14,2
199710	x			x		1 819	2 449	34,6	2 228	9,9
199711	x	x		x	Modem	1 836	4 741	158,2	3 628	30,7
199712	x			x	Modem	1 848	3 822	106,8	3 215	18,9
199801	x			x		1 519	2 199	44,8	1 826	20,4
199802	x	x				1 826	3 536	93,6	2 707	30,6
199803	x		x		Modem, type de disque dur	1 621	2 442	50,6	3 222	-24,2
199804	x				Modem	2 230	3 606	61,7	3 483	3,5
199805	x		x			5 981	6 311	5,5	7 671	-17,7
199806	x	x				2 630	3 965	50,8	3 769	5,2
199807	x	x			Type de disque dur	2 423	2 552	5,3	3 685	-30,7
199808	x		x			1 866	2 351	26,0	2 668	-11,9
199809	x	x	x		Type de disque dur	1 799	2 513	39,7	3 005	-16,4
199810	x	x				2 533	3 661	44,5	3 658	0,1
199811	x	x				7 568	8 875	17,3	10 827	-18,0
199812	x	x				1 806	2 112	16,9	2 591	-18,5
199901	x		x			1 684	2 316	37,5	2 194	5,5
199902	x		x			1 644	2 128	29,4	2 158	-1,4
199903	x	x			Type de disque dur	1 741	2 939	68,8	2 672	9,9
199904	x	x				2 058	2 053	-0,2	2 839	-27,7
199905	x	x				2 997	4 652	55,2	4 077	14,1
199906	x				Type de disque dur	3 918	4 639	18,4	5 637	-17,7
199907	x	x				1 108	1 275	15,1	1 501	-15,1
199908	x	x	x			2 793	3 266	16,9	4 939	-33,9
199909	x		x			6 574	10 086	53,4	9 762	3,3
199910	x	x				1 779	1 734	-2,5	2 537	-31,7
199911	x				Fournisseur	1 889	1 547	-18,1	2 259	-31,5
199912	X	x				1 554	2 121	36,5	2 165	-2,1
200001	X		x			2 066	2 150	4,1	2 297	-6,4
200002	x	x	x			1 606	2 646	64,8	2 566	3,1
200003	x	x				2 186	2 314	5,9	3 241	-28,6
200004	x		x			1 984	2 336	17,7	2 511	-7,0
200005	x				Type de disque dur	3 251	4 233	30,2	4 784	-11,5

3.3 Indices hédoniques directs

Méthode à variable fictive et à groupement

La façon la plus directe de produire un indice de prix en correction de qualité consiste surtout à estimer une série indiciaire à l'aide de méthodes économétriques. Il y a plusieurs manières de procéder. Une des plus anciennes est la méthode à variable fictive où on estime un modèle comme le suivant :

$$\ln P_{it} = \alpha_0 + \sum_{t=1}^T \delta_t D_t + \sum_{k=1}^n \beta_{ki} X_{ki} + u_i .$$

Dans cette représentation, P_{it} est le prix du i^{e} modèle dans la période t , D_t est une variable fictive de temps égale à 1 dans la période t et à 0 dans les autres cas et X_{ki} représente les diverses caractéristiques d'un ordinateur comme les capacités respectives de la mémoire vive, de l'unité centrale et du disque dur. Le coefficient δ_t décrit la variation du prix en pourcentage de période en période, la qualité étant constante. Dans cette méthode de groupement, D_t représentera plusieurs périodes subséquentes (mois, trimestres ou années) qui sont toutes groupées dans une même estimation. On prend alors l'antilogarithme des valeurs estimées des divers δ_t pour construire l'indice de prix en correction de qualité.

Comme autre possibilité, la méthode de groupement offre effectivement l'avantage d'une maximisation de la taille d'échantillon, mais accuse deux grands inconvénients en raison de la stabilité des coefficients et du peu d'applicabilité de cette technique à l'élaboration d'un indice de prix en toute continuité. D'abord, s'il y a groupement, c'est qu'on suppose que les coefficients des divers X_{ki} seront stables tout au long de la période d'observation, ce qui est hautement improbable pour deux raisons à mesure qu'avance la période T . D'abord, plus le temps passe, plus il devient probable que les caractéristiques changent au gré de l'évolution des produits et de l'introduction de nouvelles caractéristiques. Il y a aussi le fait que, sur une période suffisamment longue, on peut s'attendre à une évolution des relations entre le prix d'un produit et les différents X_{ki} , principalement à cause de l'évolution et du comportement des marchés secondaires d'éléments comme la mémoire vive et l'unité centrale⁴². Enfin, d'un point de vue pratique, le recours à une telle méthode pour la production d'une série indiciaire continue ferait continuellement réviser toute la série par réestimation de l'équation de départ à mesure que s'accroît l'ensemble de données en groupement.

Méthode à variable fictive et à périodes adjacentes

Comme solution de rechange à la méthode de groupement, il y a les régressions de périodes adjacentes. Au lieu de grouper d'un coup toutes les périodes et d'estimer une seule équation, on groupe seulement deux périodes adjacentes à la fois. Les estimations dégagées servent ensuite à produire un indice de prix en chaîne par multiplication sur la suite d'estimations de l'antilogarithme pour δ_t (Berndt, 1991). Avec cette méthode hédonique à variabilité temporelle, l'hypothèse de la stabilité des coefficients se relâche et on laisse ceux-ci varier dans le temps, particulièrement dans le cas des données mensuelles. On écarte en outre le problème d'une

⁴² Dans leur étude, Berndt, Dulberger et Rappaport (2000) ont constaté que tel était le cas.

révision de toute la série indiciaire après chaque estimation. C'est pour ces raisons que nous avons calculé un indice hédonique direct par la méthode des périodes adjacentes.

La figure 3.3 présente les résultats avec la série indiciaire tirée de la méthode d'appariement d'échantillons en chevauchement à des fins de comparaison. Comme on peut le voir au graphique, l'indice à variable fictive et à périodes adjacentes décroît plus rapidement que sa contrepartie. Le taux mensuel moyen de variation est de -4,23 %; c'est une différence de 1,38 % (en valeur absolue) d'avec la série en chevauchement.

Méthode de Prud'homme-Yu

Les solutions de rechange à la méthode à variable fictive abondent pour le calcul d'indices hédoniques directs. Dans une méthode récemment proposée par Prud'homme et Yu (2001), on se sert des valeurs prévues de périodes adjacentes pour établir de simples valeurs relatives de prix⁴³. Pour dégager une valeur relative entre t et $t+1$ (ou $p_{t/t+1}$), on doit en première étape estimer deux modèles hédoniques pour l'une et l'autre de ces périodes. On groupe ensuite les observations des deux périodes pour former un ensemble unique de données, et celui-ci permet alors de produire deux jeux de valeurs prévues pour l'une et l'autre des équations :

$$\hat{p}_k^t = f(X_i^t, X_j^{t+1}, \hat{\beta}^t)$$

$$\hat{p}_k^{t+1} = f(X_i^t, X_j^{t+1}, \hat{\beta}^{t+1})$$

où $k = i + j$ (données groupées)⁴⁴. Dans l'ensemble de données en groupement, chaque modèle a prévu les prix des périodes t et $t+1$. On peut alors se reporter à ces prix pour calculer la base indiciaire en fonction de la moyenne arithmétique des valeurs relatives de prix, de la valeur relative de prix de la moyenne arithmétique ou de la moyenne géométrique. Il semblerait que cette méthode est principalement privilégiée par rapport à la méthode à variable fictive, parce qu'elle procède par appariement de modèles d'ordinateurs, sous une forme faible cependant puisque les prix sont imputés pour les modèles manquants (en fait, tous les prix font l'objet d'une imputation).

Résultats

La figure 3.4 présente les résultats de l'application de cette méthode où l'indice est issu du calcul des valeurs relatives de prix par la moyenne géométrique, opération suivie d'un chaînage. Le taux estimé de variation mensuelle de cet indice est de -4,07 %. Nous avons inclus à des fins de comparaison la série indiciaire calculée par la méthode à variable fictive et la méthode de chevauchement. On ne s'étonnera pas que les séries produites par la méthode à variable fictive et

⁴³ En fait, la méthode de Prud'homme-Yu dont nous parlons dans cette étude n'est pas une méthode hédonique directe à proprement parler, mais vu le grand rôle que jouent les estimations de régression dans l'élaboration des indices, elle est plus proche d'une méthode directe que d'une méthode indirecte.

⁴⁴ Il s'agit tout simplement d'une reformulation de la description donnée à l'origine dans Prud'homme et Yu (2001).

la méthode de Prud'homme-Yu se ressemblent le plus des trois, bien que le premier de ces indices décroisse à un rythme un peu plus rapide⁴⁵.

Indices hédoniques directs – observations

D'un point de vue pratique, l'emploi de ces méthodes hédoniques directes pose un certain nombre de problèmes, dont le principal est un degré supérieur d'intervention dans le calcul indiciaire. Avec la méthode à variable fictive et à périodes adjacentes, les estimations de δ_t peuvent subir l'influence de facteurs normalement liés à toute estimation de régression (médiocrité de l'ajustement de modèle, présence de formes aggravées de multicollinéarité, variables omises, etc.), ce qui peut à son tour influencer sur la « robustesse » de la série indiciaire. Ajoutons que les résultats présentés dans ce chapitre font appel à toutes les observations d'un mois quelconque, qu'elles soient représentatives ou non⁴⁶. Ce facteur est aussi de nature à introduire un certain biais dans le calcul indiciaire, particulièrement dans les mois où les modèles ne changent pas beaucoup. Le lecteur se rappellera que, au chapitre qui précède, nous avons dit qu'environ 9 modèles sur 10 sont en appariement d'un mois à l'autre et que, par conséquent, l'évolution des modèles demeure un phénomène modeste à intervalles mensuels. Comme le taux d'appariement est si élevé, on peut s'interroger sur la nécessité d'employer un indice hédonique direct si on dispose des prix effectifs (il faut aussi savoir que l'évolution est cumulative et qu'une certaine forme d'ajustement est donc nécessaire).

De même, dans la méthode de Prud'homme-Yu où tous les prix font l'objet d'une imputation, le biais pourrait être plus grand que lorsqu'on se reporte aux prix effectifs ou directs de modèles en appariement. Là encore, le degré de biais dépendra du degré d'ajustement du modèle estimé selon ce que détermine le choix de variables explicatives, de formes fonctionnelles, etc., d'où la conclusion que les modèles à piètre ajustement produiront des prix à piètre imputation. Une autre difficulté se présente lorsque les modèles hédoniques ajustés changent légèrement de mois en mois pour les définitions des variables qualitatives (fictives)⁴⁷. On se doit en pareil cas d'imposer des définitions communes pour les deux périodes visées au détriment même d'un bon degré d'ajustement d'un des deux modèles (ou des deux).

3.4 Méthodologie actuelle – échantillons appariés et correction hédonique

Pour produire un indice des prix des ordinateurs, Statistique Canada applique actuellement une méthode analogue à celle qu'emploie le Bureau of Labor Statistics (BLS). Dans les deux cas, on procède par échantillons appariés et on se sert d'équations hédoniques en

⁴⁵ La différence de taux mensuels estimés de décroissance entre la méthode de Prud'homme-Yu et la méthode à variable fictive est de -0,16 %; la différence entre cette même méthode de Prud'homme-Yu et la méthode d'appariement d'échantillons en chevauchement s'établit à 1,41 %.

⁴⁶ Au chapitre suivant, nous aborderons la question de la composition d'échantillon.

⁴⁷ Pour le type de disque dur par exemple, le nombre de variables fictives représentant cette caractéristique et leur définition différaient entre deux mois avec pour résultat que la valeur relative de prix qu'on calculait était nettement non représentative. Pour résoudre le problème, on a groupé les données et ajusté un nouveau modèle. À l'aide des variables contenues dans l'ensemble de données groupées, on a réestimé les modèles d'ordinateurs pour chaque mois et dégagé des valeurs prévues de prix. Les valeurs relatives de prix ainsi obtenues s'alignaient davantage sur le reste de la série. Ce problème s'est présenté à quelques reprises, mais il n'était pas omniprésent dans l'analyse.

ajustement des prix déclarés à mesure qu'évoluent les caractéristiques des modèles (voir Holdway, 2000).

On estime foncièrement des modèles transversaux à intervalles réguliers. Les résultats les plus récents servent alors au calcul du prix en correction de qualité d'un modèle de remplacement dans l'échantillon par application des différences de caractéristiques entre l'ancien et le nouveau modèle⁴⁸.

Considérons l'exemple hypothétique du tableau H où le nouveau modèle disponible en mai est sur le point de remplacer le modèle initial disponible jusqu'en avril. La différence nominale de prix entre les deux modèles est de 450 \$ et, si le nouveau modèle devait remplacer l'ancien en mai sans autre considération d'un changement de qualité, la valeur relative de prix dégagée serait de 1,286, soit un renchérissement de 28,6 %. Il est clair que les modèles ne sont pas directement comparables, puisqu'ils diffèrent pour des caractéristiques essentielles de rendement. Par une équation hédonique qui appréhende les apports respectifs de ces éléments à l'explication des différences de prix des ordinateurs, on prend l'évolution des caractéristiques du modèle pour établir un prix fictif du nouveau modèle si celui-ci avait existé en avril.

Tableau H – Changement hypothétique de modèles d'ordinateurs

Caractéristiques du modèle	Modèle initial en avril	Modèle de remplacement en mai
Processeur central	Compaq Pentium III Deskpro EP 500 MHz	Compaq Pentium III Deskpro EP 600 MHz
Mémoire	64 Mo de mémoire vive	128 Mo de mémoire vive
Antémémoire étendue	128 Ko	256 Ko
Capacité de stockage ou taille du disque dur	6,4 Go	10,0 Go
Inclusion d'une carte Ethernet	Aucune carte Ethernet (ou réseau)	Carte Ethernet (ou réseau) incluse
Prix	1 573 \$	2 023 \$
Prix fictif du nouveau modèle en avril		2 847 \$

En guise d'illustration, posons qu'une telle équation existait et que son utilisation a donné un prix fictif de 2 847 \$ du nouveau modèle pour avril. On se trouverait alors à comparer ce prix au prix effectif du modèle de remplacement (2 023 \$). La variation de prix dégagée est une diminution (-28,9 %) de 2 847 \$ à 2 023 \$. L'ordre de grandeur du changement de qualité se calcule comme la différence entre le prix fictif et le prix du modèle initial, soit $2\ 847\ \$ - 1\ 573\ \$ = 1\ 274\ \$$ (augmentation de 81 % par rapport au modèle initial). La valeur

⁴⁸ Le BLS a soumis ses équations à une révision trimestrielle, alors que, à Statistique Canada, on a révisé moins fréquemment. Au moment où nous rédigeons ces lignes, les équations avaient fait l'objet de deux révisions annuelles. Auparavant, celles-ci étaient occasionnelles (voir Barzyk, 1999).

relative de prix obtenue indique une *diminution* de 28,9 % d'avril à mai, et non pas l'augmentation de 28,6 % calculée au départ⁴⁹.

Cette méthode joint les deux extrêmes du modèle d'appariement d'échantillons en chevauchement et des méthodes hédoniques directes. En soi, elle offre l'avantage de s'ancrer dans la méthode d'appariement de modèles, ce qui idéalement constitue l'objet même de cette mesure indiciaire. S'il n'y a pas d'évolution des modèles, la question des changements de qualité disparaît et toutes les méthodes se ramènent à un simple calcul indiciaire à appariement de modèles. Toutefois, comme il y a bel et bien évolution des modèles, la méthode hédonique sert à ajuster la qualité des modèles de remplacement, si bien que les différences de qualité font l'objet d'un traitement direct, autre grand avantage de cette orientation. Dans une optique historique, il convient de mentionner qu'une des grandes raisons pour lesquelles on a établi et appliqué ce cadre méthodologique en le préférant à un indice hédonique direct est qu'on manquait des données nécessaires pour produire un tel indice. Tel n'est plus le cas, car la base de données en place est une riche source d'observations.

Dans une méthode d'appariement d'échantillons et de correction hédonique de qualité cependant, la gestion d'échantillon tient une grande place dans la production d'un indice. Dans tous les cas où on procède par échantillons appariés, le choix et la mise à jour d'un échantillon représentatif est de la plus haute importance si on entend produire une série « robuste ». Maintenant que nous avons décrit en général cette orientation en matière d'élaboration d'indices, nous allons construire une série indiciaire hypothétique à l'aide de cette technique.

Série indiciaire à échantillons appariés et à correction hédonique de qualité

Comme nous l'avons mentionné, cet indice de prix hypothétique se calculera principalement par une méthode tirée du cadre méthodologique déjà en place. Pour l'essentiel, on établit les valeurs relatives de prix pour un échantillon de modèles en appariement d'un mois à l'autre. On remplace des modèles en appliquant la méthode hédonique pour tenir compte des changements de qualité. En voici le détail :

Mode d'échantillonnage

La série commence en mars 1996. Ce mois-là et les 18 mois qui suivent approximativement, l'échantillon est d'une taille mensuelle d'une centaine de modèles. L'ensemble initial était formé de 30 modèles et l'intention était de représenter le groupe principal de processeurs centraux (en gros les 50 % intermédiaires de l'échantillon). À mesure que se multipliaient les modèles disponibles, l'échantillon s'est accru de 15 modèles deux fois dans la période, soit en octobre 1997 et en mars 1998, ce qui devait porter l'échantillon total à 60 modèles. La sélection des modèles supplémentaires s'est répartie également sur toute la tranche principale en fonction de la distribution courante de ceux-ci à l'époque et dans un souci de ne pas introduire de biais dans l'échantillon. Enfin, on a pondéré les fournisseurs représentés dans l'échantillon d'après les proportions de modèles attribuables à chacun dans tout l'ensemble de données d'un mois. Ainsi,

⁴⁹ Selon le mode actuellement appliqué de calcul d'ajustement de qualité dans le programme des indices de prix dans l'industrie de la Division des prix, la différence de prix entre avril et mai se calculerait pour le **nouveau** modèle, et non pas pour l'ancien.

si les modèles d'un fournisseur constituaient 20 % de tous les modèles d'un mois, ils constituaient aussi 20 % de l'échantillon du mois⁵⁰.

Remplacement de modèles

On a remplacé les modèles selon ce que déterminait l'échantillon (remplacements forcés) et on a intégré les modèles de remplacement en appliquant la méthode de correction hédonique de qualité. Lorsqu'un modèle courant bas de gamme ou de gamme intermédiaire disparaissait, on sélectionnait un modèle haut de gamme dans la tranche principale et, si on n'en trouvait pas, on descendait le long des catégories d'unités centrales jusqu'à ce qu'on découvre un modèle de remplacement approprié. Si un modèle haut de gamme était retiré, on sélectionnait un modèle de remplacement dans le même groupe et, si on n'en trouvait aucun, la tranche était étendue à la catégorie de processeurs centraux immédiatement supérieure. On procédait ainsi jusqu'à ce qu'on découvre un modèle de remplacement. En réalité, le choix d'un modèle de remplacement a représenté un arbitrage entre la longévité dans l'échantillon et l'ampleur de la correction de qualité correspondante. Pour expliquer ce moyen terme, disons que, le plus souvent, les modèles qui disparaissent dans un mois appartiennent à une ou plusieurs catégories d'unités centrales où la technologie est moins avancée ou plus dépassée. Ce sont foncièrement des catégories qui disparaîtront un jour ou l'autre. Ainsi, si on remplace un modèle par un autre dans la même catégorie d'unités centrales ou dans une catégorie analogue, il faut normalement songer à des remplacements plus fréquents – puisque les modèles de remplacement sont susceptibles de disparaître assez vite eux aussi –, mais comme les modèles de remplacement sont d'une technologie semblable, les différences de qualité seront ténues. À l'autre extrême, le choix de modèles de remplacement dans des catégories d'unités centrales où la technologie est très avancée garantit directement que ces modèles resteront longtemps dans l'échantillon, puisqu'ils arrivent tout juste sur le marché. Les changements de qualité seront donc moins fréquents, mais plus importants, car la technologie sera différente. Dans notre étude, notre compromis a été de sélectionner les modèles de remplacement dans la partie supérieure de la tranche principale, de sorte que les avantages et les coûts propres aux deux extrêmes se réalisent et s'équilibrent dans une certaine mesure.

Mode de correction de qualité

Les équations hédoniques estimées plus tôt dans ce chapitre ont servi à la correction de qualité des modèles de remplacement. Cela s'est fait à intervalles trimestriels et, dans ce cas, l'équation du premier mois d'un trimestre (mars, juin, septembre 1996, etc.) a été appliquée trois mois durant avant d'être remplacée. Nous disposions certes d'équations mensuelles aux fins de notre étude, mais les changements trimestriels de modèles représentent un bon compromis entre l'actualité des modèles et les contraintes de ressources que connaîtra la production effective de l'indice. Là où il n'existait aucune variable dans l'équation hédonique pour une caractéristique

⁵⁰ C'est plus un effet artificiel de l'exercice, et il ne faut pas y voir nécessairement un avantage. Tout au long de la production de l'indice des prix des ordinateurs, on a fait ample usage de renseignements et de recherches à jour sur le marché pour établir la composition de l'échantillon. Pour notre étude, nous avons expérimenté une méthode plus objective. Nous en discuterons plus en détail au prochain chapitre.

unique d'un modèle de remplacement, nous avons cherché un modèle analogue et, faute d'en trouver, nous avons écarté le changement en question⁵¹.

Résultats

La figure 3.5 présente la série indiciaire obtenue avec les deux autres indices hédoniques. Le taux mensuel estimé de variation en est de -4,49 %, très près des valeurs de -4,29 % et -4,07 % dégagées pour la méthode à variable fictive et la méthode de Prud'homme-Yu.

Changements de modèles

Sur les 52 mois de la série indiciaire, 307 modèles au total ont servi à la production de l'indice. Si on prend les 30 modèles initiaux et les 15 modèles qui se sont ajoutés à deux reprises, on constate que 247 remplacements de modèles ont été nécessaires en 51 mois. C'est une moyenne approximative de 5 changements de modèles par mois.

Correction de qualité

Les 247 remplacements ont tous été effectués par l'équation hédonique applicable à la période, mais pour 89 d'entre eux (proportion de 36 %), il y avait un changement pour au moins une caractéristique supplémentaire non représentée dans l'équation. Le tableau I résume les changements apportés. Il faut dire que la majeure partie de ceux-ci visent une seule variable supplémentaire (proportion de 68,5 % ou 61 sur 89). Pour la plupart de ces remplacements, les différences de variables ont été gardées le plus petites possible compte tenu de la disponibilité de modèles. En fait, 61 changements de modèles sur 89 ont eu lieu dans les deux premières années de la série (1996 et 1997) au moment où l'échantillon était le plus petit et le nombre de remplacements, limité. Il faut noter que, pour les caractéristiques en question, on dispose généralement des prix de caractéristiques facultatives. Disons enfin que le R^2 corrigé moyen était de 0,8270 pour les modèles hédoniques ayant servi à produire cette série. On peut voir que, en moyenne, ces modèles rendaient compte de 82,7 % de la variabilité des prix, ce qui veut dire que l'incidence de ces variables supplémentaires sur les calculs était infime au mieux. Les résultats vont dans le sens de ce que nous avons fait valoir plus tôt dans ce chapitre au sujet des caractéristiques supplémentaires et des équations hédoniques, à savoir que ce n'est pas là un grand obstacle à l'application de cette méthode de traitement de la qualité.

⁵¹ Aux fins de notre étude, nous n'avons pas établi de prix d'éléments ou de caractéristiques facultatifs, car les données chronologiques sur les prix étaient restreintes. Dans la pratique, si un tel cas se présente, le prix d'un élément ou d'une caractéristique « en option » est relativement facile à trouver.

Tableau I – Récapitulation des corrections de qualité

Catégorie	Nombre	Répartition en pourcentage
Nombre total de changements de qualité	247	100,0 %
1. Changements de modèles dont rend entièrement compte l'équation hédonique	↓ 158	↓ 64,0 %
2. Changements de modèles où l'équation hédonique ne rend pas compte de caractéristiques supplémentaires et qui se composent de :	89	36,0 %
2a) Changements d'une variable seulement	61	24,7 %
- taille du disque dur	↓	↓
- antémémoire	36	14,6 %
- type de disque dur	10	4,0 %
	15	6,1 %
2b) Changements de plusieurs variables	28	11,3 %

3.5 Conclusion

Nos conclusions dans ce chapitre se limiteront à des observations générales sur le cadre hédonique. Au prochain chapitre, nous soumettrons à une comparaison plus structurée les séries issues des diverses méthodes dont nous avons parlé jusqu'ici.

Cela étant dit, une des grandes critiques adressées à la méthode hédonique est celle de contraintes d'exécution comme la disponibilité des données et l'abondance des ressources nécessaires à l'estimation des modèles. Si on veut obtenir des résultats solides et exploitables, il faut disposer d'une bonne source de données tant en qualité qu'en quantité. On peut constater ensuite que l'application du mode d'ajustement de modèle a tout d'un exercice laborieux. Tout le raisonnement à la base de la méthode hédonique est celui du rapport à établir entre les différences de prix et les différences de caractéristiques clés des modèles d'ordinateurs. Les modèles à piètre ajustement seraient l'indice que, au mieux, ce rapport est douteux ou aléatoire.

Il ressort toutefois de cet exercice que, avec un minimum d'efforts, on peut trouver des modèles assez bien ajustés qui produisent des résultats réalistes et exploitables. Notre base de données se prête très bien à ce genre d'estimation. Outre les prix, nous suivons certaines caractéristiques des modèles (mémoire vive, unité centrale, etc.) uniformément dans le temps, ce qui facilite une analyse de ce genre.

Une autre question que nous n'avons pas abordée vraiment est celle de savoir ce qu'il advient si une caractéristique n'entre pas dans les changements d'équations hédoniques. Qu'en est-il, par exemple, si une carte son est incluse dans le modèle de remplacement (elle ne se trouvait pas dans le modèle initial) et que ce changement de qualité n'est pas représenté dans l'équation hédonique que nous estimons? En fait, le statisticien a quelques possibilités : a) il peut choisir un autre modèle de remplacement où cette caractéristique est inchangée; b) il peut trouver d'autres renseignements pour estimer le coût de cette caractéristique qualitative (par le prix d'une caractéristique offerte à titre facultatif, par exemple); c) il peut enfin négliger ce changement de qualité s'il n'a pas d'importance (s'il ne s'agit pas d'une variable clé, l'incidence devrait en être infime). Si le modèle est d'un bon ajustement et que les variables essentielles sont représentées (mémoire vive, unité centrale, disque dur, etc.), les autres changements devraient être considérés comme secondaires et traités à ce titre. Il suffit de dire que ce problème ne nuit pas outre mesure à l'utilité de la méthode hédonique dans le traitement des changements de qualité.

Un mot enfin sur le vieillissement des estimations d'un modèle hédonique. Il n'y a pas eu d'étude en bonne et due forme qui ait été consacrée à la longévité de ces modèles estimés, mais par la répartition déglagée des formes fonctionnelles et par l'expérience que peut avoir la Division des prix de l'utilisation de ces modèles, on peut penser que leur degré d'efficacité dépend principalement de la fréquence de leur révision. Plus ces estimations vieillissent, moins leurs résultats sont représentatifs. En définitive, ce problème est plus primordial que celui des variables manquantes : si on a le choix, il semblerait de loin préférable d'estimer des modèles simples plus fréquemment (modèles comportant les seules variables clés) que d'estimer occasionnellement seulement des modèles d'un parfait ajustement, c'est-à-dire en laissant s'écouler de longues périodes entre les révisions. Comme notre étude le démontre jusqu'ici, les changements de qualité tiennent en gros à une poignée de caractéristiques principales et l'appréhension de cette dynamique devrait être notre but réel en matière méthodologique.

Figure 3.1 - R² corrigé pour les modèles avec les seules variables RAM et CPU

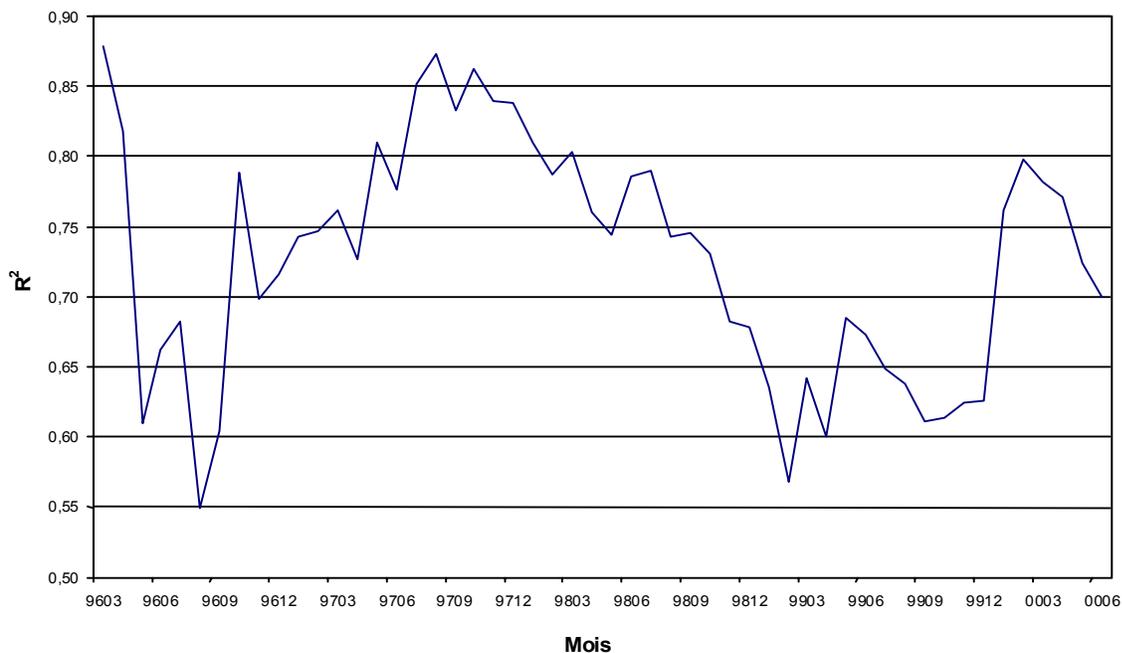


Figure 3.2 - R² corrigé pour les modèles avec variables supplémentaires

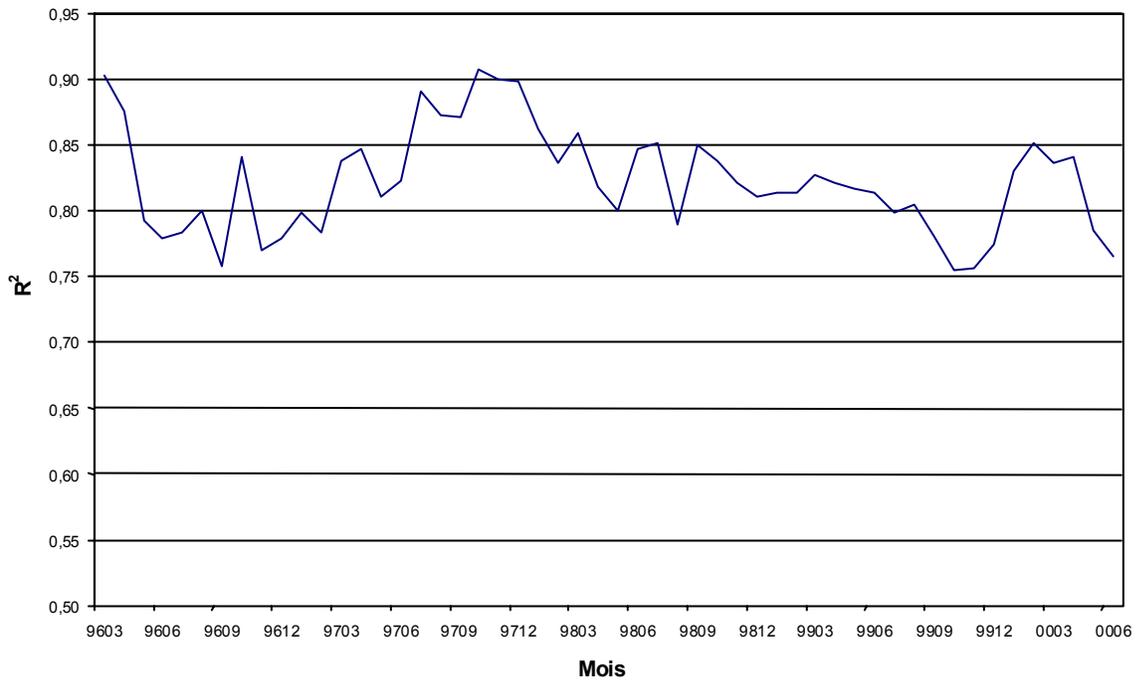


Figure 3.3 - Indices de prix par la méthode à variable fictive et à périodes adjacentes et par la méthode du chevauchement (diagramme semi-logarithmique)

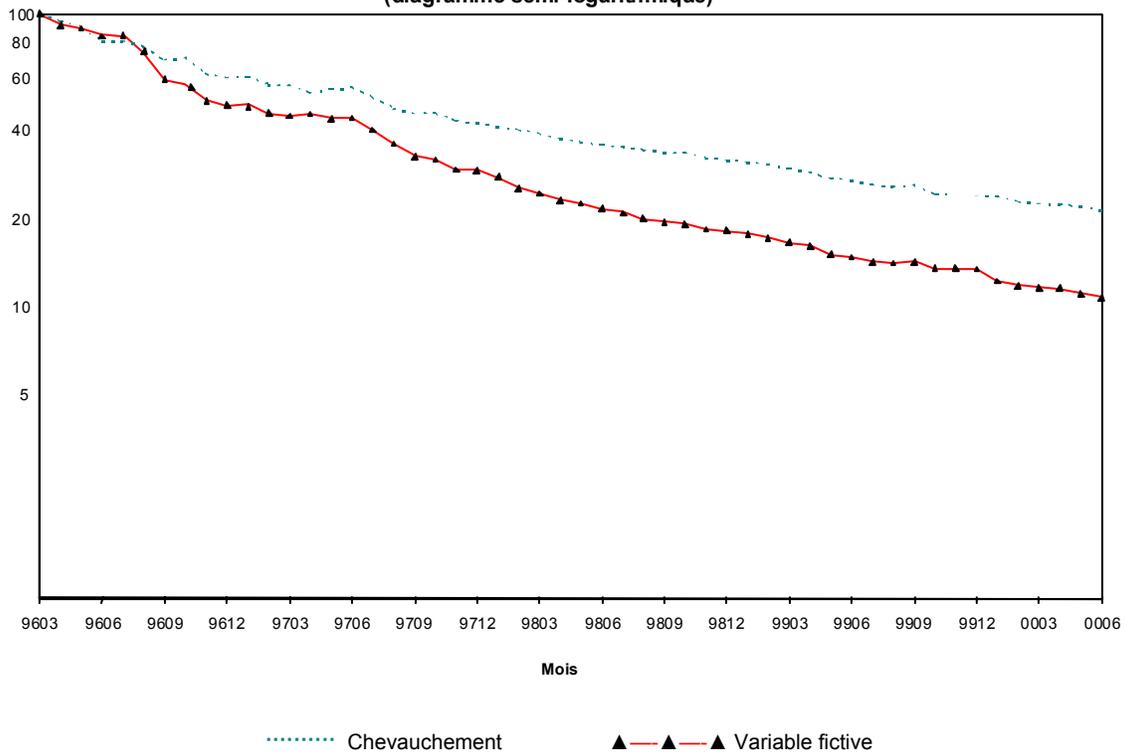


Figure 3.4 - Indices de prix par la méthode du chevauchement, la méthode à variable fictive et à périodes adjacentes et la méthode de Prud'homme-Yu (diagramme semi-logarithmique)

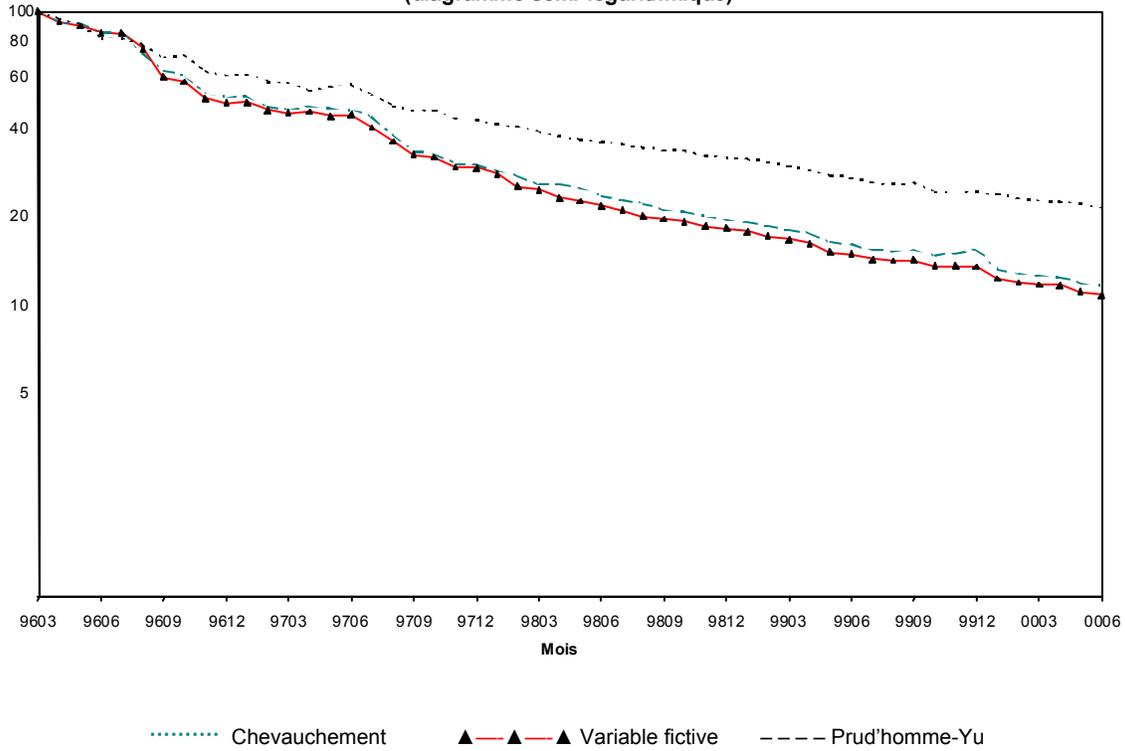
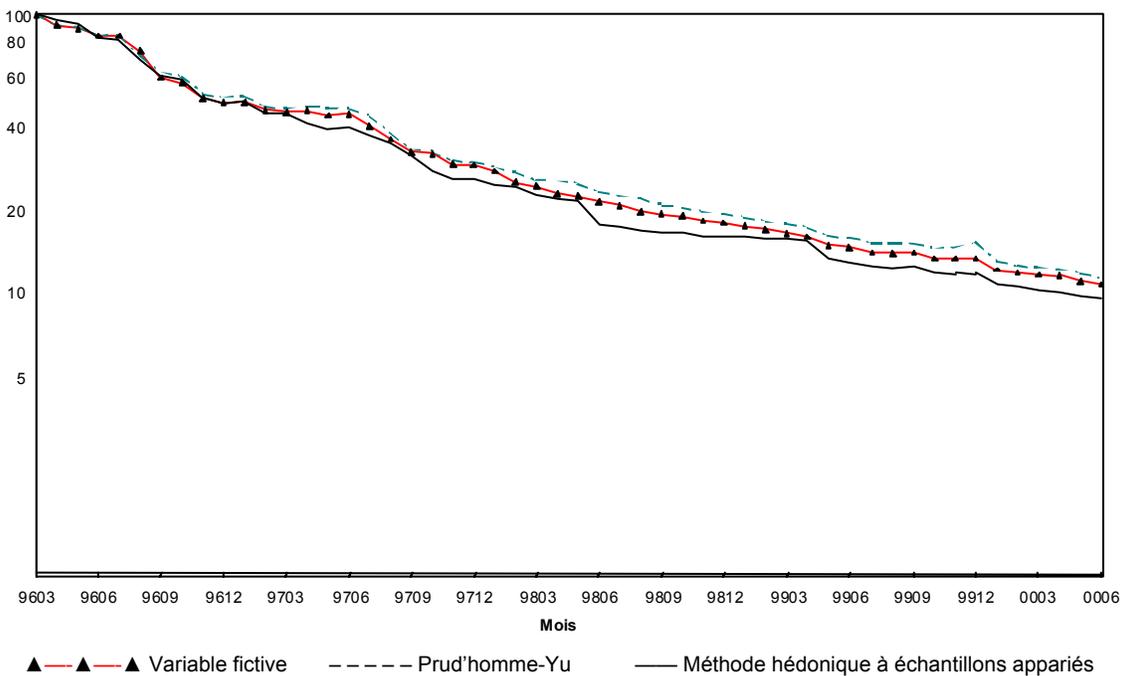


Figure 3.5 - Indices de prix par la méthode à variable fictive et à périodes adjacentes, la méthode de Prud'homme-Yu et la méthode hédonique à échantillons appariés (diagramme semi-logarithmique)



CHAPITRE 4 - Comparaison des séries et méthodologie proposée

4.1 Introduction

Dans ce chapitre, nous avons deux grands objectifs, le premier étant de comparer les diverses séries indiciaires produites jusque-là dans notre étude (méthodes d'appariement d'échantillons en chevauchement, à variable fictive, de Prud'homme-Yu et de correction hédonique à échantillons appariés) et le second de recommander une méthodologie de production d'un indice courant des prix des ordinateurs, ainsi que de plusieurs séries récemment proposées à la Division des prix⁵². La méthodologie envisagée vient d'une diversité de comparaisons et d'analyses présentées dans cette étude, tout comme de la connaissance et de l'expérience acquises depuis près de dix ans dans la production d'un tel indice.

4.2 Comparaison des séries

Séries initiales

En première étape, nous avons comparé les quatre séries indiciaires déjà mentionnées à un niveau général et noté les différences de mouvements et de taux mensuels moyens de variation. Par niveau général, nous entendons les séries initiales décrites dans les chapitres qui précèdent comme elles seraient traitées dans un cadre méthodologique ordinaire. La figure 4.1 présente ces versions. On peut y voir que les trois séries hédoniques sont pour ainsi dire identiques et sûrement bien plus proches en valeur et en mouvement que dans le cas de la série issue de la méthode d'appariement d'échantillons en chevauchement. Ce résultat se trouve encore confirmé par une comparaison de taux mensuels moyens de variation portant sur les quatre séries du tableau J (lignes A et B), où l'indice hédonique à échantillons appariés sert de point de comparaison. La différence est ténue entre les trois indices hédoniques (0,26 % et 0,42 % en valeur absolue), mais bien supérieure dans le cas de l'indice tiré de la méthode de chevauchement (1,64 % en valeur absolue). Il est clair que les indices hédoniques décroissent plus rapidement que la série indiciaire par la méthode de chevauchement.

⁵² Comme nous l'avons déjà mentionné, on produit les séries d'indice courant des prix des ordinateurs à l'aide de données relatives aux ordinateurs de bureau et aux ordinateurs portatifs que doivent acheter le secteur commercial et le secteur des administrations publiques. La Division des prix a l'intention de produire une série équivalente pour le secteur des ménages. Il en sera plus amplement question dans la seconde partie de ce chapitre.

Tableau J – Comparaison des taux mensuels moyens de variation

Catégorie	Indice par échantillons appariés en chevauchement	Indice à variable fictive	Indice de Prud'homme-Yu	Indice hédonique à échantillons appariés
A) Taux moyen de variation des séries initiales	-2,85 %	-4,23 %	-4,07 %	-4,49 %
B) Différence de taux entre les diverses séries initiales et la série hédonique à échantillons appariés	-1,64 %	-0,26 %	-0,42 %	0,00 %
C) Taux moyen de variation des séries produites à l'aide des modèles de la série hédonique à échantillons appariés	-2,32 %	-3,94 %	-4,26 %	-4,49 %
D) Différence due à l'effet de changement de qualité = (série hédonique à échantillons appariés) – (séries de la ligne C)	-2,17 %	-0,55 %	-0,23 %	0,00 %
E) Différence due à l'effet d'échantillonnage = (séries de la ligne B) – (séries de la ligne D)	0,53 %	0,29 %	-0,19 %	0,00 %

Séries issues de la méthode hédonique à échantillons appariés

Compte tenu des méthodes différentes qui s'appliquent, ces résultats n'ont rien d'étonnant. Pour l'essentiel, deux effets jouent, à savoir un *effet d'échantillonnage* et un *effet de changement de qualité*. Dans le premier cas, la différence entre les séries peut tenir en partie à ce que des échantillons différents soient choisis et utilisés dans ces cadres méthodologiques différents, d'où bien sûr une incidence tout aussi différente sur l'indice, soit un *effet d'échantillonnage*. Pour l'*effet de changement de qualité* – même une fois que l'on garde l'échantillon constant pour toutes les séries –, on relèvera des différences imputables au mode de traitement du changement de qualité pour la nature et la portée de la correction effectuée (directe ou indirecte).

Pour cerner ces deux effets, nous avons recalculé les séries des méthodes d'appariement d'échantillons en chevauchement, à variable fictive et de Prud'homme-Yu avec les mêmes modèles (ou le même échantillon) que dans le calcul de l'indice hédonique à échantillons appariés. La figure 4.2 présente ces séries recalculées. Ajoutons que le tableau J indique les taux mensuels moyens de variation correspondants tant pour les séries initiales que pour les séries recalculées (lignes A et C respectivement). À la ligne D du tableau J, on peut trouver le degré estimé de l'*effet de changement de qualité* calculé comme le taux mensuel moyen de variation pour l'indice hédonique à échantillons appariés, moins le taux correspondant pour chacun des autres indices recalculés. Enfin, l'*effet d'échantillonnage* peut être l'objet d'un calcul résiduel comme la différence des séries initiales (ligne B), moins la différence due à l'*effet de changement de qualité* (ligne D).

Des résultats de la figure 4.2 et du tableau J, on peut tirer les grandes observations suivantes : 1) la série de Prud'homme-Yu et la série hédonique à échantillons appariés sont encore plus proches en valeur et en mouvement qu'auparavant; 2) la divergence s'est accrue aussi bien entre la série par la méthode de chevauchement et la série hédonique à échantillons appariés et, ce qui est encore plus intéressant, entre la série par la méthode à variable fictive et cette même série hédonique. On peut voir ces résultats à la troisième ligne du tableau J.

Série par la méthode d'appariement d'échantillons en chevauchement

L'effet de changement de qualité estimé est de -2,17 % dans le cas de la série par chevauchement. Il est fort important si on le compare à ceux des autres séries. Si on considère les résultats présentés au chapitre 2, ce résultat n'a cependant pas de quoi étonner. Le lecteur se rappellera que, à ce chapitre, nous avons montré que le traitement des changements de qualité lié à la méthode de chevauchement est entaché d'un biais par excès, d'où une prévision de décroissance plus rapide des séries hédoniques. On notera avec intérêt que l'effet d'échantillonnage est positif (0,53 %) et aide à compenser la différence globale par rapport à l'indice hédonique à échantillons appariés. Il semblerait que, lorsque tout l'échantillon est utilisé, des modèles s'ajoutent là où les baisses de prix sont plus prononcées, ce qui rend plus proches en valeur la série par chevauchement et la série hédonique à échantillons appariés.

Série par la méthode à variable fictive

Pour l'indice à variable fictive, les effets d'échantillonnage et de changement de qualité ont le même sens, bien qu'étant moindres que les effets correspondants de la série par chevauchement. Le second est de -0,55 % et le premier, de 0,29 %. On doit sans doute l'un et l'autre au nombre bien moindre d'observations sur lesquelles est maintenant fondée l'estimation de δ_t , ce qui réduit le degré moyen d'ajustement des modèles et la « robustesse » des estimations. C'est ce que montre dans une certaine mesure la figure 4.3, où sont comparés les R^2 corrigés des séries initiales et des séries venant des modèles de l'indice hédonique à échantillons appariés. Comme on peut le voir, le degré d'ajustement est bien moindre dans le second cas (pour toute la période, le R^2 corrigé moyen est de 0,6802 comparativement à 0,8092 pour les séries initiales). La différence s'accroît tôt dans les séries. Pour la période mars 1996-mars 1998, le R^2 corrigé moyen est de 0,6076 pour les séries recalculées et de 0,8092 pour les séries initiales. Le reste de la période, les degrés d'ajustement sont bien plus proches en valeur (les R^2 corrigés moyens sont maintenant de 0,7447 et 0,8161 respectivement).

Série par la méthode de Prud'homme-Yu

L'effet de changement de qualité est bien moindre dans le cas de l'indice de Prud'homme-Yu (-0,23 %). Il reste qu'on doit probablement cette différence à un degré d'imputation bien supérieur, puisque *tous* les prix font l'objet d'une imputation tant *prospective* que *retrospective* par cette méthode. Les modèles hédoniques ont beau être d'un bon ajustement en général, celui-ci n'est pas parfait si bien que l'imputation pourrait avoir un effet d'amortissement sur les mouvements observés des prix. Pour l'effet d'échantillonnage, la différence est de -0,19 %, l'explication étant que tous les modèles sont utilisés dans la méthode de Prud'homme-Yu et que les prix de ces modèles supplémentaires sont tous imputés, ce qui oblige à tenir compte une fois de plus d'un effet d'amortissement.

4.3 Recommandations

D'après les résultats de notre étude, nous recommandons que la Division des prix continue à produire sa série indiciaire des ordinateurs par la méthodologie conçue il y a presque dix ans déjà. Il est sûr que la méthode d'appariement d'échantillons en chevauchement est une solution de rechange viable là où les méthodes hédoniques ne peuvent être employées, mais comme nous le démontrons dans notre étude, un biais par excès paraît lié à ce genre de correction de qualité implicite. Ajoutons que l'on peut surmonter le problème des caractéristiques manquantes en se reportant aux prix de caractéristiques facultatives.

Pour des indices hédoniques directs comme ceux que donnent la méthode à variable fictive et périodes adjacentes et la méthode de Prud'homme-Yu que nous avons examinées, les résultats obtenus ressemblent fort à ceux de la méthode hédonique à échantillons appariés; il n'y aurait donc guère de différence entre ces trois méthodes. Toutefois, notre étude montre qu'une méthode d'appariement de modèles est certainement applicable, puisqu'un tel appariement de mois en mois donne des résultats généralement bons. Et si on accepte la prémisse d'une minimisation du degré d'imputation lorsqu'on dispose de données effectives pour le calcul d'une valeur relative de prix, une méthode hédonique directe serait considérée comme plus utile surtout là où une grande proportion de l'échantillon a disparu et qu'un degré élevé d'imputation s'impose. Ce serait le cas avec un indice annuel.

Bref, si nous prenons une grille comme celle de la figure 4.4 où les deux critères d'évaluation des cadres méthodologiques sont la théorie (ou le caractère souhaitable de ces méthodes pour des raisons théoriques) et la pratique, le classement des diverses méthodes porterait toutes les séries hédoniques dans le quadrant supérieur gauche et la série par échantillons appariés en chevauchement dans le quadrant supérieur droit. Dans cette représentation, nous pouvons voir que les séries hédoniques seraient théoriquement plus souhaitables et que, par ailleurs, les considérations pragmatiques liées à ce cadre méthodologique ne différeraient pas outre mesure par rapport à la méthode d'appariement d'échantillons en chevauchement. Des trois séries hédoniques, la série par échantillons appariés serait la plus souhaitable, suivie à cet égard de la série de Prud'homme-Yu et de la série à variable fictive. Si cette distinction existe, c'est à cause du degré d'application du concept des échantillons appariés dans l'indice qui décroît d'une technique hédonique à l'autre.

Il reste que la méthodologie actuelle de production d'une série indiciaire nécessite un affinement et une normalisation (qu'il s'agisse d'une révision plus fréquente des équations hédoniques ou d'une amélioration des procédures d'échantillonnage), aussi proposons-nous une méthodologie que nous présenterons dans le reste de ce chapitre. Celle-ci est largement tirée de la méthodologie d'élaboration de l'indice hédonique à échantillons appariés que nous exposons au chapitre 3. Nous avons toutefois ajouté des concepts et des procédures en fonction de la disponibilité actuelle de données.

4.4 Méthodologie proposée pour les indices des prix des ordinateurs

Objectifs généraux

La création d'un indice des prix des ordinateurs vise en général à la production de tout un train de séries indiciaires des ordinateurs et du matériel informatique pour plusieurs secteurs de

l'économie, à savoir ceux des ménages, des administrations publiques et des entreprises commerciales⁵³. Le but est de produire un groupe d'indices qui soit représentatif et théoriquement solide.

Stratégie d'échantillonnage

Il y a deux volets à une stratégie d'échantillonnage, à savoir la sélection et le remplacement de l'échantillon.

Sélection de l'échantillon

La stratégie d'échantillonnage vise principalement à concevoir, à constituer et à mettre à jour un échantillon représentatif de modèles d'ordinateurs compte tenu des sources d'information disponibles. Dans la mesure du possible, on se reportera aux valeurs des livraisons pour déterminer par trimestre la composition correspondante de fournisseurs et de générations de processeurs. On choisira les modèles dans un groupe principal correspondant à la tranche intermédiaire de 50 % à 60 % des modèles dont les prix sont actuellement fournis à la Division des prix.

Remplacement de l'échantillon

On remplacera les modèles selon ce que détermine l'échantillon. À l'heure actuelle, la situation ne justifie pas des remplacements forcés là où les prix de modèles dépassés ou obsolètes sont encore relevés. L'expérience acquise à ce jour indique que ces modèles s'éliminent de la source de données en temps opportun. Lorsqu'un modèle courant bas de gamme ou de gamme intermédiaire est perdu, il sera remplacé par un modèle choisi dans la partie supérieure de la tranche principale et, si on n'en trouve pas, la sélection descendra le long des catégories de processeurs centraux jusqu'à ce qu'on découvre un modèle de remplacement approprié. Si un modèle courant haut de gamme disparaît, on trouvera un modèle de remplacement dans le même groupe et, si on n'en découvre pas, la tranche principale sera étendue à une catégorie supérieure à la fois jusqu'à ce qu'on repère un modèle de remplacement.

Correction de qualité

À mesure que des modèles sont remplacés, la méthode hédonique servira à la prise en compte des changements de qualité. On révisera les équations hédoniques tous les trimestres en s'appuyant sur les données les plus à jour. À chaque révision, il s'agira de vérifier la forme fonctionnelle jugée appropriée et d'évaluer l'ajustement des modèles et la « robustesse » des résultats (inclusion des principales variables indépendantes, bons signes des coefficients, valeurs prévues, etc.). Si un modèle de remplacement offre des caractéristiques qui ne sont pas représentées dans l'équation hédonique, trois possibilités s'offrent :

- (1) trouver un modèle de remplacement approprié qui a des caractéristiques représentées;
- (2) déterminer le prix de caractéristiques facultatives pour estimer les valeurs des caractéristiques manquantes;

⁵³ Dans la terminologie du Système de comptabilité nationale, ces trois secteurs sont ceux des **particuliers et des entreprises individuelles**, des **administrations publiques** et des **sociétés et entreprises publiques**.

(3) négliger ces caractéristiques supplémentaires.

Formules et pondération

Selon la disponibilité de données de pondération, on mettra tout en œuvre pour produire un ensemble d'indices à pondération fixe et mobile, ainsi qu'établir une moyenne géométrique des deux (ce qu'on appelle fréquemment les formules indiciaires de Laspeyres, de Paasche et de Fisher).

Séries proposées

Il s'agira d'indices des prix des ordinateurs de bureau et des ordinateurs portatifs pour chacun des trois secteurs visés. On combinera les deux indices en question en un indice d'ensemble pour chacun de ces secteurs. Enfin, on réunira les secteurs dans un autre grand indice d'ensemble.

Diffusion

Les séries seront d'une périodicité mensuelle, et on prévoira des révisions aux six mois. On publiera ces séries et les mettra à la disposition du public par les moyens habituels de diffusion de Statistique Canada.

Figure 4.1 - Indices par les méthodes de chevauchement, à variable fictive, de Prud'homme-Yu et hédonique à échantillons appariés (diagramme semi-logarithmique)

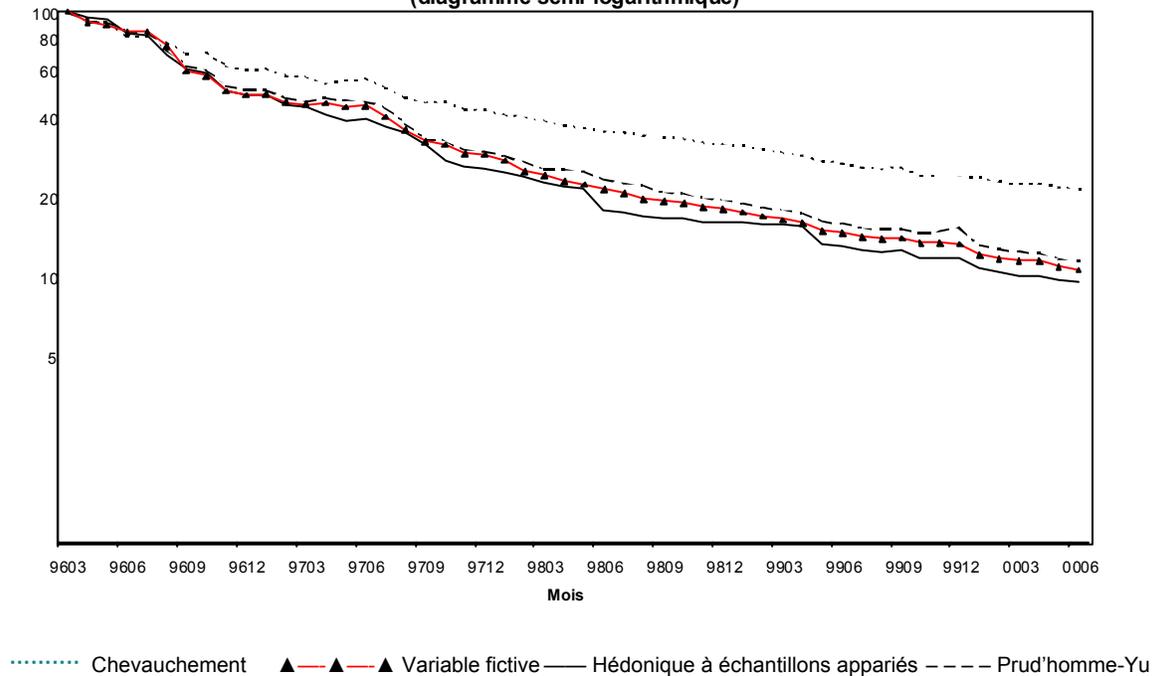


Figure 4.2 - Indices par les méthodes de chevauchement, à variable fictive, de Prud'homme-Yu et hédonique à échantillons appariés à l'aide de données hédoniques à échantillons appariés (diagramme semi-logarithmique)

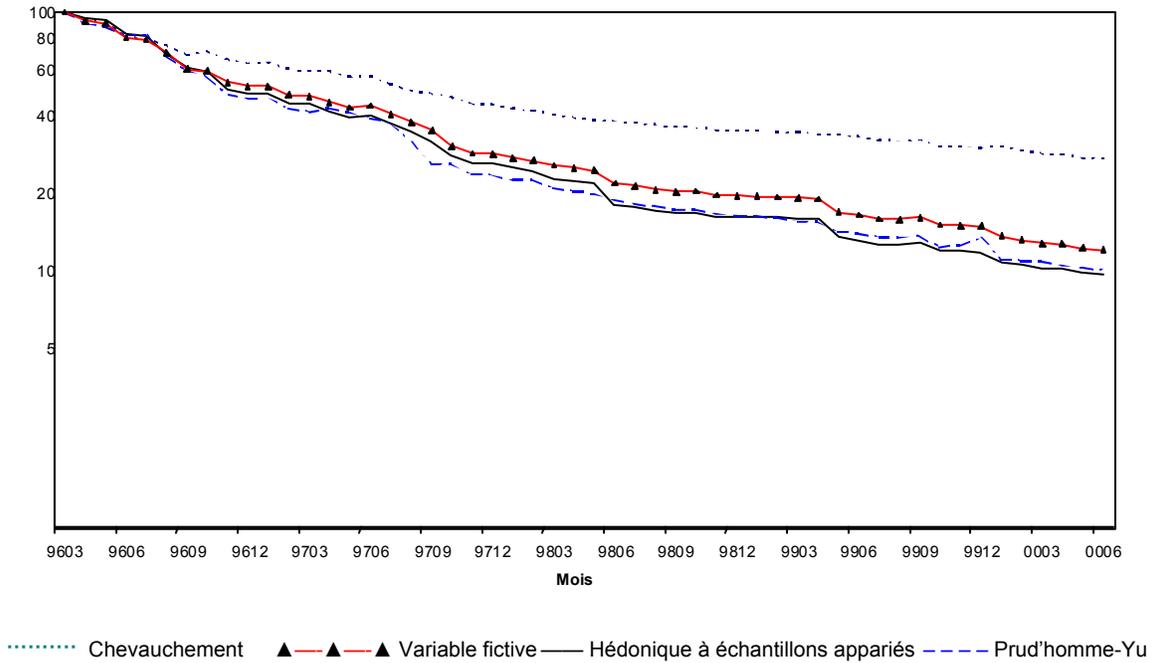


Figure 4.3 - Comparaison des R² corrigés obtenus pour les séries initiales et recalculées à variable fictive

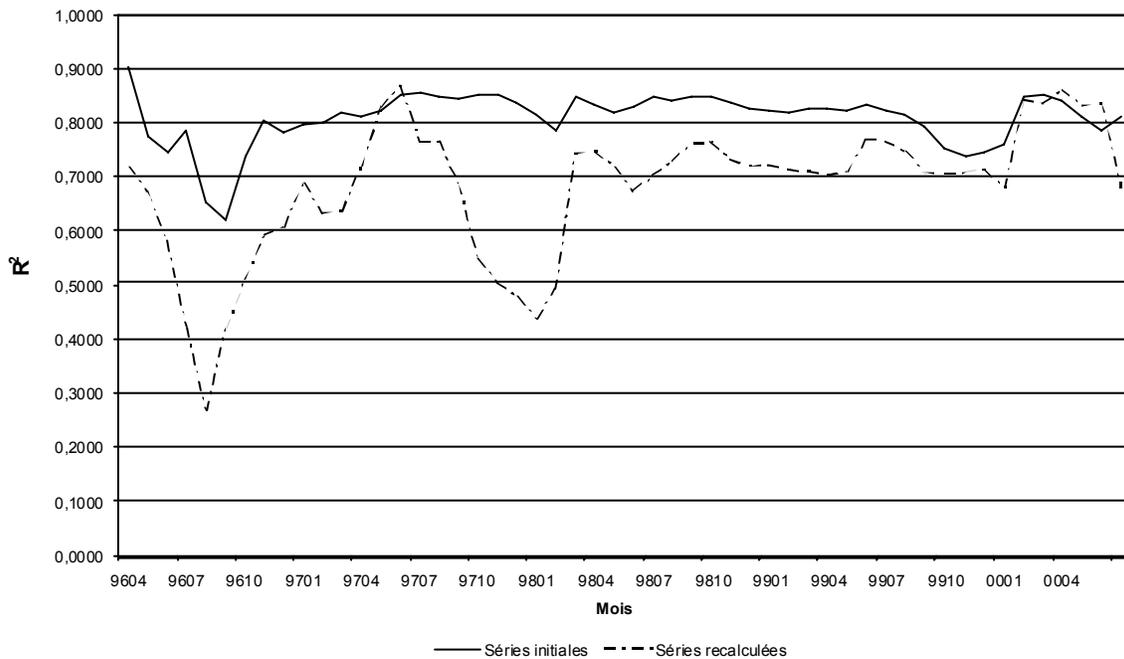
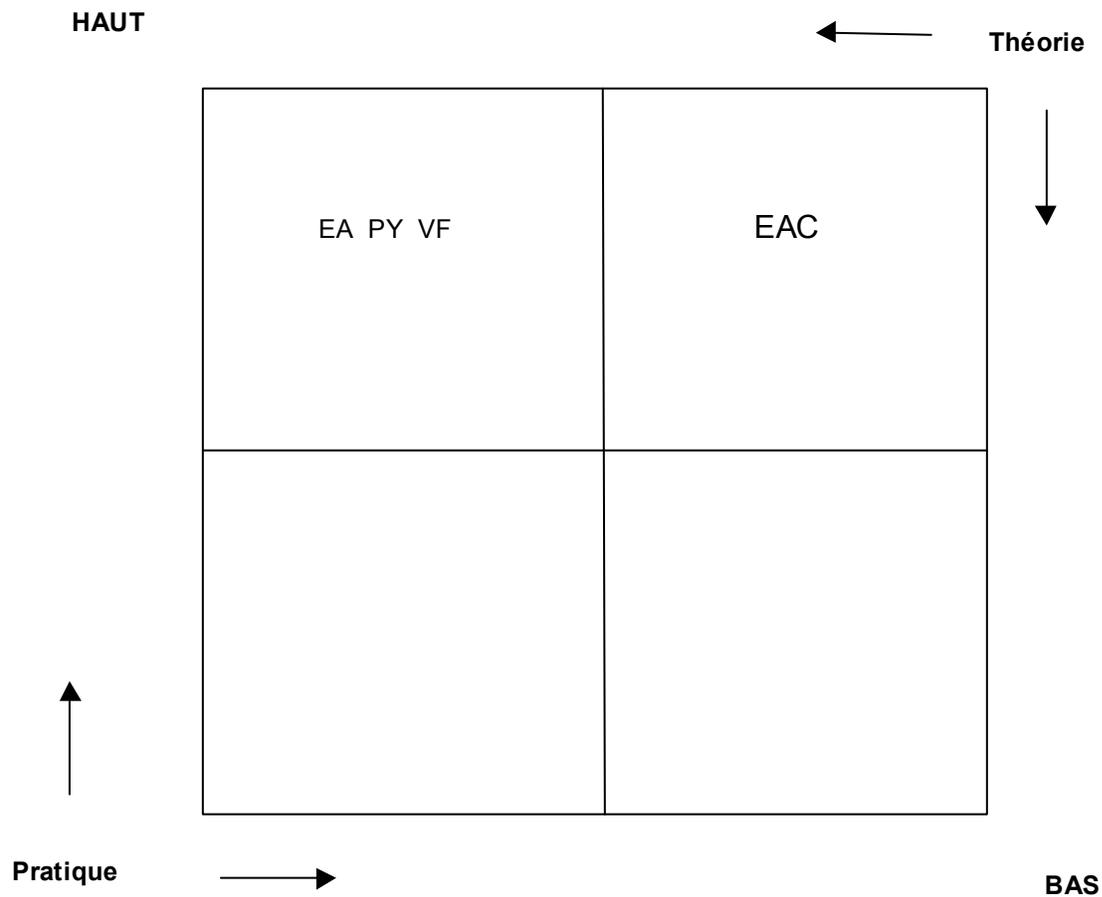


Figure 4.4 – Récapitulation des possibilités du point de vue de Statistique Canada



Où :

- EA - Indice hédonique à échantillons appariés
- PY - Indice de Prud'homme-Yu
- VF - Indice à variable fictive
- EAC - Indice par échantillons appariés en chevauchement

Bibliographie

- Barzyk, Fred (1999), “Updating the Hedonic Equations for the Price of Computers”, Working Paper, Prices Division, Statistics Canada.
- Berndt, Ernst R. (1991), *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley Publishing Company, Reading Massachusetts.
- Berndt, Ernst R., Ellen R. Dulberger and Neal J. Rappaport (2000), “Price and Quality of Desktop and Mobile Personal Computers: A Quarter Century of History”, National Bureau of Economic Research, paper presented at the conference on *Price, Output and Productivity Measurement, Summer Institute 2000*.
- Computer Consulting Ltd. *The CPU Scorecard*. URL: http://www.cpuscorecard.com/all_cpus.htm. Last updated July 17, 2000. (Read July 17, 2000).
- Cranford-Petelle, Buffy (1999), “Buying Computers – Desktop Systems”, *Smart Computing Magazine*. 7, 9.
- Davidson, Russell and James G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.
- Edwards, Cliff. “Intel has High Hopes for Pentium IV Processor”, *The Ottawa Citizen*, August 22, 2000, p. C4.
- Gujarati, Damodar N. (1988), *Basic Econometrics: Second Edition*, Toronto: McGraw-Hill Book Company.
- Hastings, Bryan (1997), “Top Power at Budget Prices”, *PC World Magazine*. 15, 2: pp. 179–200.
- Holdway, Michael (2000), “Quality-Adjusting Computer Prices in the Producer Price Index: An Overview,” US Bureau of Labor Statistics, Last Modified: November 27, 2000, URL: <http://stats.bls.gov/ppicomqa.htm>.
- Intel Corporation. “Intel Delivers the Next Level of Computing with the New Pentium II Processor,” *Press release*. URL: <http://www.intel.com/pressroom/archive/releases/dp050797.htm>. Released May 7, 1997. (Read September 5, 2000).
- Intel Corporation. “Intel Introduces 200-Mhz Pentium Processor”, *News Release*. URL: <http://www.intel.com/pressroom/archive/releases/pp200.htm>. Released June 10, 1996. (Read September 5, 2000).
- Intel Corporation. “Intel Introduces the Pentium Processor with MMX Technology”, *Press Release*. URL: <http://www.intel.com/pressroom/archive/releases/dp010897.htm>. Released January 8, 1997. (Read September 5, 2000).

- Kenedy, Kristen. “Intel Launching Pentium III with a Bang,” *Technology News*. URL: <http://www.techweb.com/wire/story/TWB1999012s0011.htm>. Released February 12, 1999. (Read September 5, 2000).
- Mainelli, Tom (1998), “RAM – Need More? Then Buy More”, *Smart Computing Magazine*. 9, 3.
- McLaughlin, Laurianne (1998), “400-Mhz Pentium II’s: The Great Leap Forward”, *PC World Magazine*. 16,6: p. 98–104.
- Morocho, Richard. “Compaq Bets on Growth”, *The Toronto Star*. April 24, 1997, p. G5.
- Prud’homme, Marc and Kam Yu (2001), “Trends in Internet Access Prices”, Unpublished paper, Prices Division, Statistics Canada.
- Reuter. “Intel, Dell make hefty price cuts”, *The Toronto Star*. May 2, 1996, p. B8.
- Starnes, Richard. “Tumbling Computer Prices Boost Sales Across Canada”, *The Kitchener-Waterloo Record*. June 10, 1996, p. B5.
- 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.
- Triplett, Jack E. (2000), “Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products”, *Industry Committee*, OECD Directorate for Science, Technology and Industry, Revised Draft.
- Turvey, Ralph (1999), “Incorporating new models into a CPI: PCs as an example”, Working Paper presented at *The Measurement of Inflation Conference*, University of Cardiff.
- Turvey, Ralph (2000), “Ralph Turvey CPI Manual”, URL: <http://www.turvey.demon.co.uk/index.htm>. (Read January 18, 2001).

Annexe I – Cotes de rendement des processeurs centraux

Catégorie de processeurs	Cotes d'unité centrale au 17-07-2000	Catégorie de processeurs	Cotes d'unité centrale au 17-07-2000
486DX4 / 100	118	Celeron 466	1373
Pentium / 75	186	Xeon II / 450 (512 Ko)	1406
Pentium / 90	220	Xeon II / 450 (1 Mo)	1420
Pentium / 100	245	Pentium III / 450	1500
Pentium / 120	270	Celeron 500	1502
Pentium / 133	298	Celeron 533	1628
Pentium / 150	306	Xeon III / 500 (512 Ko)	1649
Pentium / 166	339	Pentium III / 500	1650
Pentium / 200	377	Xeon III / 500 (1 024 Ko)	1667
Pentium MMX / 166	422	Pentium III / 533	1752
Pentium Pro 150	459	Pentium III / 550	1780
Pentium MMX / 200	478	Xeon III / 550 (512 Ko)	1827
Pentium Pro 180	516	Pentium III / 500E	1867
Pentium MMX / 233	531	Pentium III / 600	1930
Celeron 266	548	Pentium III / 533EB	1960
Pentium Pro 200	574	Pentium III / 600B	2036
Celeron 300	582	Pentium III / 600E	2110
Pentium II / 233	693	Pentium III / 600EB	2177
Celeron 300A	762	Xeon III / 667	2242
Pentium II / 266	784	Pentium III / 650	2270
Celeron 333	818	Pentium III / 667	2320
Pentium II / 300	857	Pentium III / 700	2420
Celeron 366	890	Xeon III / 733	2477
Pentium II / 333	940	Pentium III / 733	2510
Pentium II / 350	1000	Pentium III / 750	2540
Celeron 400	1011	Pentium III / 800	2690
Pentium II / 400	1130	Xeon III / 800	2714
Pentium II / 450	1240	Pentium III / 866	2890
Celeron 433	1248	Pentium III / 1 GHz	3280

Annexe II – Taux de déperdition d'échantillon sur 12 mois

Mois de référence	Taille d'échantillon au début	Nombre d'appariements après 12 mois	Taux de déperdition de l'échantillon initial
199603	104	6	94,2
199604	86	4	95,3
199605	123	4	96,7
199606	116	4	96,6
199607	116	4	96,6
199608	91	13	85,7
199609	80	7	91,3
199610	89	6	93,3
199611	84	4	95,2
199612	84	3	96,4
199701	82	4	95,1
199702	67	3	95,5
199703	71	3	95,8
199704	103	0	100,0
199705	105	4	96,2
199706	66	3	95,5
199707	102	7	93,1
199708	195	33	83,1
199709	212	75	64,6
199710	190	76	60,0
199711	197	56	71,6
199712	196	56	71,4
199801	204	61	70,1
199802	210	80	61,9
199803	249	97	61,0
199804	284	131	53,9
199805	290	42	85,5
199806	232	58	75,0
199807	246	73	70,3
199808	285	107	62,5
199809	338	166	50,9
199810	353	182	48,4
199811	337	205	39,2
199812	361	227	37,1
199901	404	2	99,5
199902	450	21	95,3
199903	570	66	88,4
199904	556	66	88,1
199905	493	98	80,1
199906	506	104	79,4
Moyenne =	223,2	54,0	75,8