

Mesures alternatives de la durée moyenne du chômage

par Miles Corak* et Andrew Heisz**

No. 83

11F0019MPF No. 83

ISSN: 1200-5231

ISBN: 0-660-94635-1

24F Immeuble R.H. Coats, Ottawa, K1A 0T6
Analyse des entreprises et du marché du travail
* (613) 951-9047 Internet: coramil@statcan.ca

** (613) 951-3748

Télécopieur: 1 (613) 951-5403

Vous pouvez vous procurer une copie du document par internet: www.statcan.ca

Décembre 1995

Une première version a été présentée à la 23^e conférence générale de l'Association internationale de recherches sur le revenu et la fortune, qui a eu lieu à St. Andrew, au Nouveau-Brunswick. Nous remercions Denise Dorion, Andrew Harvey, Ian Macredie et Deborah Sunter de leurs précieux commentaires. Les auteurs assument l'entière responsabilité des idées exprimées dans ce document.

L'auteur assume seul la responsabilité des opinions dans le présent document qui ne représente pas nécessairement le point de vue de Statistique Canada.

Also available in English

RÉSUMÉ

L'article porte sur la robustesse d'une mesure de la durée complète moyenne du chômage au Canada par rapport à diverses hypothèses de calcul. Contrairement à la durée incomplète moyenne du chômage, qui est un indicateur cyclique retardé, cette statistique est un indicateur contemporain du cycle économique. On examine l'impact d'une hypothèse d'état stable plutôt que de l'hypothèse alternative, de même que l'incidence des diverses méthodes de correction du biais dans les réponses. On conclut qu'un estimateur d'état non stable apporterait un complément utile aux données sur la durée du chômage qui sont actuellement publiées par bien des organismes statistiques, plus particulièrement Statistique Canada.

Mots clés : Chômage, durée, dynamique, Canada.

Mesures alternatives de la durée moyenne du chômage

Bien qu'il représente assurément l'un des indicateurs économiques que l'on suit de plus près, le taux de chômage, pris isolément, donne un aperçu tout à fait incomplet de la situation du marché du travail. Un taux de chômage de 10 %, supposons, peut révéler que 10 % de la population active perd son emploi chaque mois et ne consacre que quelques semaines à la recherche d'un nouveau travail, ou encore que cette même tranche de 10 % de la population active reste sans emploi pendant toute l'année. Dans le premier cas, le marché du travail se caractérise par un «roulement» considérable, mais une période de chômage sans conséquences graves, tandis que dans le deuxième cas, il s'agit d'un marché stagnant où le chômage cause de graves difficultés. Les répercussions de ces deux scénarios sur le bien-être des travailleurs peuvent être très différentes. Pour bien comprendre la situation, il faut disposer d'un indicateur fiable de la durée moyenne d'une période de chômage.

La structure des enquêtes comme l'Enquête sur la population active (EPA) ou la *U.S. Current Population Survey* (CPS), au même titre que celle des enquêtes semblables menées dans d'autres pays, tient compte de la nature dynamique du marché du travail. Toutefois, les seules données officielles qui sont publiées sont des données groupées sur la durée déclarée du chômage, et la durée moyenne des périodes de chômage en cours (soit une durée incomplète). Les analystes ont pu se servir de ces données pour mettre au point des mesures de la durée moyenne d'une période *complète* de chômage, à l'aide de méthodes paramétriques et non paramétriques. Baker et Trivedi (1985) ont avancé que les méthodes non paramétriques, qui reposent sur les résultats de Kaplan et Meier (1958), étaient supérieures. Entre autres exemples, Kaitz (1970) s'est servi des données de la CPS et, plus récemment, aux États-Unis, Sider (1985) et Baker (1992a), ont fait de même. Au Canada, Corak (1993) se sert des données canadiennes.

Ces documents faisant autorité ont permis de pousser la réflexion sur la façon dont les données sur la durée du chômage sont publiées officiellement. L'article vise principalement à explorer certaines de ces réflexions et à recommander une méthode assez robuste pour permettre la publication de données officielles. Tout au long de l'article, nous avons utilisé des données canadiennes, mais bon nombre de nos conclusions et de nos recommandations peuvent s'appliquer à des données provenant d'autres pays. Nous comparons d'abord la durée incomplète moyenne du chômage à une mesure de la durée complète du chômage afin d'illustrer certains des biais que comporte la première mesure. Nous abordons ensuite le calcul de la seconde mesure. Plus particulièrement, nous examinons la robustesse de la durée complète moyenne du chômage selon qu'on adopte une hypothèse d'état stable ou d'état non stable, et selon la manière dont sont corrigés les biais dans les réponses. Nous constatons que la durée complète moyenne du chômage fondée sur une hypothèse d'état stable, même si elle est plus facile à calculer, risque de sous-évaluer considérablement la durée de la période lorsque le taux de chômage augmente (lorsqu'il y a un afflux de chômeurs) et de la surévaluer lorsque ce taux (et l'afflux de chômeurs) baisse. L'hypothèse d'état stable fausse aussi le caractère saisonnier de la statistique. Nous concluons que les avantages de l'estimateur d'état stable, en ce qui

a trait à la facilité de calcul, sont mineurs et ne compensent pas ces biais plutôt importants. Nous constatons aussi des pointes importantes dans la distribution par fréquences de la durée déclarées des périodes de chômage, pointes qui sont associées à une «préférence en matière de chiffres» des répondants de l'enquête. Toutefois, le lissage des données pour corriger cette erreur de mesure n'est essentiel que pour déterminer le niveau de la durée complète moyenne, mais non sa variabilité cyclique, et est important surtout pour les parties les plus denses de la distribution. Nous recommandons donc d'adopter une méthode assez robuste pour permettre la publication de données officielles et soutenons qu'elle est un complément utile à l'unique statistique sur la durée (incomplète) du chômage qui est actuellement publiée par bien des organismes.

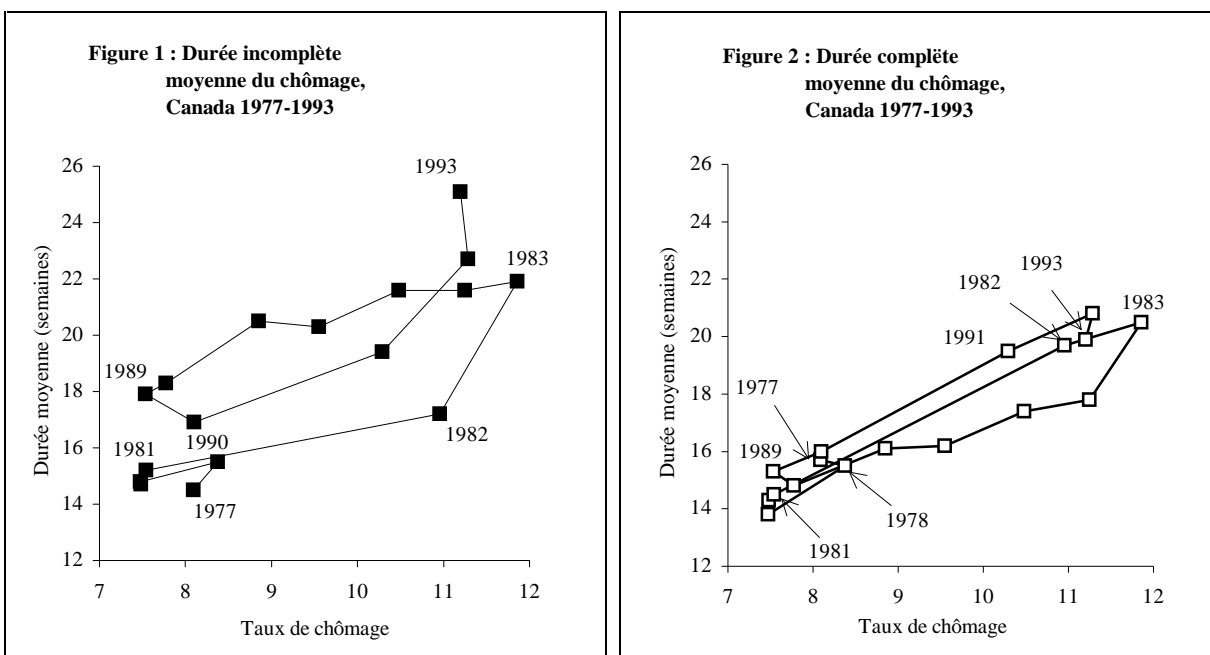
Aperçu

Les données sur la durée moyenne du chômage officiellement publiées par Statistique Canada sont calculées à partir d'un échantillon de personnes actuellement en chômage. L'EPA saisit non pas la durée complète d'une période de chômage, mais uniquement sa durée jusqu'à la semaine de référence. Il se peut que la période de chômage se poursuive encore pendant des semaines ou prenne fin dès le lendemain de l'enquête. La durée moyenne du chômage correspond au quotient de la somme de toutes les périodes de chômage en cours par le nombre de chômeurs¹. Ainsi, la statistique officielle mesure la durée *incomplète* moyenne pour les personnes *actuellement* en chômage.

La statistique que nous avons calculée et examinons dans l'article est la durée complète moyenne prévue du chômage d'une cohorte d'individus dont la période de chômage débute en même temps. Cet indicateur, qui mesure la période complète de chômage, repose sur l'hypothèse que les conditions économiques observées au moment où les personnes en question perdent leur emploi se maintiendront pendant toute la durée de leur période de chômage. C'est ce que nous allons désormais appeler tout simplement «la durée complète moyenne du chômage».

La durée incomplète moyenne du chômage constitue une mesure biaisée de la durée, et ce, pour deux raisons. Elle est entachée d'un biais qui a trait à la durée et l'autre, à l'échantillonnage. Salant (1977) expose clairement et étudie ces deux biais. Le premier est dû au fait que, bien évidemment, seules les périodes de chômage en cours sont échantillonnées, d'où une sous-estimation de la durée complète. Quant au biais d'échantillonnage, il résulte du fait que la probabilité qu'un chômeur soit choisi pour

¹ L'EPA classe essentiellement les répondants comme chômeurs s'ils sont sans emploi et se cherchent du travail. Les personnes mises à pied ne sont pas tenues de remplir la condition de la recherche d'emploi pour être considérées comme des chômeurs. Dans ce cas, la durée du chômage est égale au nombre de semaines écoulées depuis le début de la mise à pied. En outre, on peut être considéré comme chômeur dans l'EPA si l'on a trouvé du travail et qu'on s'attend à commencer à travailler dans quatre semaines ou moins. La durée du chômage est comptabilisée pour les personnes qui ont un emploi devant commencer à une date future seulement si elles se trouvent à chercher du travail pendant la semaine de référence. Voir Statistique Canada (1992).



l'enquête est proportionnelle à la durée de sa période de chômage. Les personnes en chômage pendant de courtes périodes seront sous-représentées en conséquence. Cela implique que la durée complète du chômage sera surestimée. Salant propose un théorème selon lequel la durée incomplète moyenne sera supérieure à la durée complète moyenne si les taux de sortie diminuent à mesure que la période de chômage s'allonge, de sorte que le biais d'échantillonnage compensera, et au-delà, le biais de durée.

C'est ce qui se produit pour la plupart des données que nous examinons. Au cours de la période 1977-1993, la durée incomplète moyenne s'établit à 18,7 semaines, et la durée complète moyenne, à 16,9 semaines.

Cette différence entre les deux statistiques est bien connue. Il faut toutefois mentionner une autre différence, celle de leur profil cyclique. Le lien entre chaque statistique et le taux de chômage applicable à l'ensemble du Canada est illustré aux figures 1 et 2. La durée incomplète moyenne se présente comme une large boucle qui évolue en sens inverse des aiguilles d'une montre. Ce phénomène est dû en grande partie aux changements qui surviennent dans la composition des chômeurs pendant le cycle économique, de sorte que la durée incomplète moyenne est un indicateur cyclique *retardé*. Au début d'une récession, un afflux massif de chômeurs fait que les personnes qui viennent juste de perdre leur emploi ont une pondération supérieure. Même si leur période de chômage peut être logique au bout du compte, seule la durée du chômage jusqu'au moment de l'enquête est retenue dans la durée moyenne de la période de chômage. Ainsi, de 1981 à 1982, alors que s'amorçait la récession, le taux de chômage a grimpé de 3,5 points, tandis que la durée moyenne du chômage selon l'EPA n'a augmenté que d'une semaine environ. De

même, le taux de chômage a progressé au début de la dernière récession en 1990, alors que la durée moyenne a de fait diminué de 1989 à 1990².

On observe le phénomène inverse en période de reprise et d'expansion : l'afflux des chômeurs ralentit et l'effectif des chômeurs compte un plus fort pourcentage de personnes se trouvant dans une période de chômage plutôt longue qui a débuté pendant la récession et qui témoigne de l'état de l'économie pendant cette période. Ainsi, lorsque l'économie s'est redressée en 1983, le taux de chômage a augmenté de moins d'un point, mais la durée moyenne s'est accrue d'environ 5 semaines. De 1983 à 1985, l'économie était en pleine expansion et le taux de chômage a reculé de 1,5 point, mais la durée incomplète moyenne du chômage n'a presque pas changé. De la même manière, entre 1992 et 1993, le taux de chômage a légèrement fléchi, mais la durée moyenne du chômage a augmenté de presque 2,5 semaines.

À l'opposé, le profil cyclique de la durée moyenne de la période complète de chômage des nouveaux sans-emploi est stable tout au long de la période. Les données forment une boucle, mais son mouvement dans le sens des aiguilles d'une montre est à peine perceptible. De plus, les points de retournement de cette statistique correspondent aux sommets et aux creux des conditions du marché du travail. La durée moyenne d'une période complète de chômage atteint son point culminant au même moment que le taux de chômage, diminue pendant la période de reprise et d'expansion, puis remonte dès le début d'une récession. La courbe observée pendant la récession des années 90 semble correspondre grosso modo à celle dégagée pendant la récession de 1981-1982. La situation de ceux qui ont perdu leur emploi au début de la décennie n'a pas été bien différente, en moyenne, de celle qu'ont connue les personnes qui sont tombées en chômage dix ans plus tôt. Au cours de 1983, la durée moyenne de la période complète de chômage était de 20,5 semaines, contre 20,8 semaines en 1992³.

Méthodologie

Pour calculer la durée complète moyenne, nous nous sommes inspirés des travaux de Sider (1985), Baker (1992a) et Corak (1993), qui emploient une méthode de cohorte fictive. Contrairement à de nombreuses méthodes de calcul utilisées auparavant, celle-ci ne repose pas sur l'hypothèse d'un état stable.

Soit $S(x,t)$ la probabilité conditionnelle qu'un individu reste chômeur jusqu'au mois x si l'on suppose qu'il était au chômage pendant $x-1$ mois. $S(x,t)$, que nous appelons taux de

² La récession de 1981-1982 a commencé en juillet 1981 et s'est terminée en novembre 1982. Les points de retournement de la récession de 1990 n'ont pas été établis, mais il est probable qu'elle a commencé en avril 1990 et a pris fin au milieu de 1992.

³ Corak (1993) fait remarquer que, malgré cette similitude, le chômage a été plus « polarisé » au cours de la dernière récession : les chômeurs de courte durée se sont relativement mieux tirés d'affaires, mais les chômeurs de longue durée ont vécu une situation plus difficile.

persistance, est égal à un moins le taux de probabilité. On peut l'estimer à partir d'un échantillon d'individus comme $N(x,t)/N(x-1,t-1)$, où $N(x,t)$ représente le nombre d'individus au chômage pendant au moins x mois, mais moins de $x+1$ mois à la période t . En d'autres termes, la probabilité de subsister jusqu'à x^e mois de chômage, compte tenu d'une période de chômage de $x-1$ mois, est simplement le rapport entre le nombre d'individus qui déclarent une période de chômage de x mois pendant la période t et le nombre de personnes qui ont déclaré être au chômage pendant $x-1$ mois au cours du mois précédent.

À partir des données de l'EPA sur le nombre déclaré de semaines de chômage, nous calculons les taux de persistance sur des intervalles de plus en plus grands : un mois, deux mois, trois mois, de quatre à six mois, de sept mois à un an, plus d'un an. La fourchette doit être plus large quand la durée s'allonge à cause des limites de la taille de l'échantillon. Plus précisément, les taux de persistance sont calculés comme les ratios du nombre d'individus dans chacune des catégories suivantes:

5 - 8	semaines dans le mois t	à	< 5	semaines dans le mois t-1
9-12	semaines dans le mois t	à	5-8	semaines dans le mois t-1
13-16	semaines dans le mois t	à	9-12	semaines dans le mois t-1
27-39	semaines dans le mois t	à	13-26	semaines dans le mois t-3
53-78	semaines dans le mois t	à	27-52	semaines dans le mois t-6
99 +	semaines dans le mois t	à	53-98	semaines dans le mois t-12.

La limite de codage supérieure des données de l'EPA est 99 semaines. Les quatrième, cinquième et sixième taux de persistance sont convertis en équivalents mensuels par élévation à la puissance 1/3, 1/6 et 1/12 respectivement. Cela suppose que les taux de persistance mensuels sont constants à l'intérieur de chaque intervalle. Ces taux mensuels servent au calcul de la durée complète moyenne du chômage pour un groupe de personnes qui ont commencé leur période de chômage au temps t :

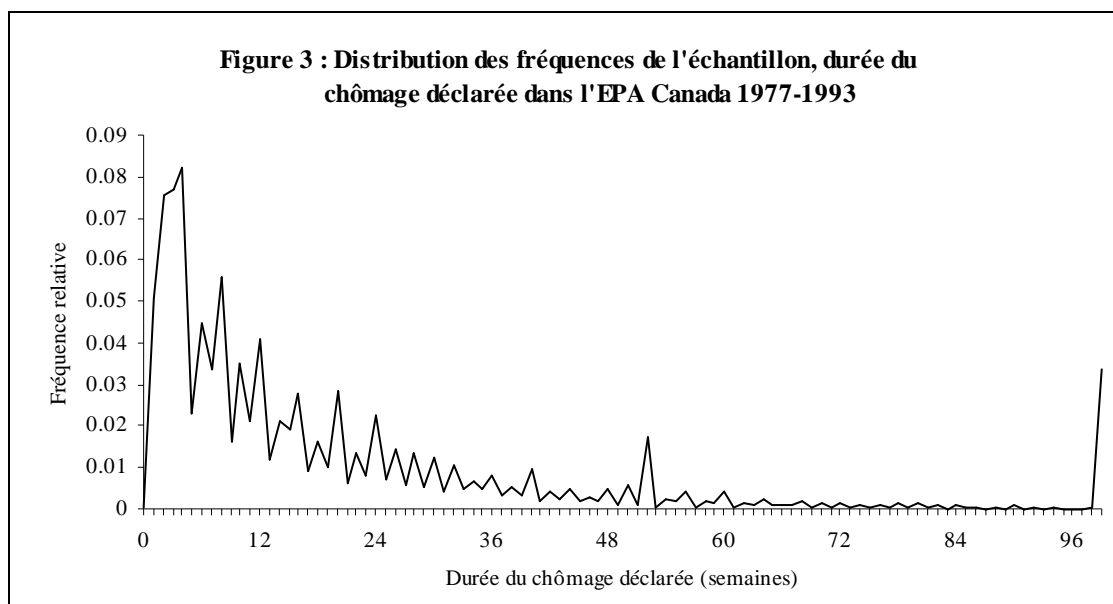
$$AvgDur(t) = \sum_{x=1}^n \prod_{i=1}^x S(i,t) \quad (1)$$

Lorsque $n = 25$ dans nos données. Le premier terme de la somme est un. Chaque terme est une estimation d'un point de la fonction de survie. La sommation est la version temporelle discrète du résultat suivant lequel, en temps continu, la durée moyenne est donnée par l'intégrale de la fonction de survie⁴.

⁴ Si τ dénote la durée complète d'une période de chômage et si $f(\tau)$ représente la fonction de densité correspondante, $\mathcal{F}(\tau) = \int_0^{\tau} f(u)du$ est la fonction de distribution cumulative et $\lambda(\tau) = 1 - \mathcal{F}(\tau)$ est la fonction de survie. La durée moyenne du chômage est $\int_0^{\infty} \tau f(\tau)d\tau$. L'intégration de cette équation par parties donne $\int_0^{\infty} \lambda(\tau)d\tau$. Voir Baker et Trivedi (1985) pour plus de détails.

Ce calcul ne repose pas sur l'hypothèse d'un état stable, mais si une telle hypothèse était retenue, le calcul de la statistique s'en trouverait simplifié sur le plan de la manipulation des données. Dans un état stable, le taux auquel les personnes tombent en chômage et les taux de persistance sont constants, de sorte que l'afflux de chômeurs compense le nombre de personnes qui se trouvent du travail. $N(x-1,t)$ serait alors égal à $N(x-1,t-1)$ et $S(x,t)$ serait simplifié en $N(x,t)/N(x-1,t)$; la probabilité de subsister jusqu'au x^e mois de chômage est le ratio du nombre d'individus qui déclarent une période de chômage de x mois pendant la période t au nombre de personnes qui ont déclaré être au chômage pendant $x-1$ mois au cours de la même période. Il est donc possible de calculer la durée complète moyenne du chômage en se servant des données d'un seul mois. Une fois encore, il faut insister sur le fait que le calcul de la persistance dans l'équation (1) repose sur l'hypothèse que les conditions observées dans le passé récent se maintiendront.

On demande aux répondants de l'EPA de déclarer en semaines la durée de leur période de chômage. Nous nous servons des résultats mensuels de 1977 à 1993. La distribution des fréquences pour la totalité de l'échantillon présente des pointes importantes aux intervalles de 2 et, surtout, de 4 semaines. On note aussi des pointes marquées à 52 et à 99 semaines (voir la figure 3). Lorsqu'ils déclarent la durée de leur période de chômage, les répondants semblent préférer les nombres pairs aux nombres impairs et les mois complets aux fractions de mois. Cette préférence en matière de chiffres a aussi été remarquée dans les données provenant de la CPS. Sider (1985) propose de lisser les données avant de calculer la durée moyenne, et Baker (1992b), qui utilise aussi les données de la CPS, examine les conséquences de diverses hypothèses de lissage. Comme on utilise de larges intervalles pour calculer la durée moyenne, il ne faut lisser que les semaines qui se trouvent à la limite de deux intervalles. Une partie des individus déclarant une période dont la durée coïncide avec cette limite doit être réaffectée à l'intervalle suivant. Sider réaffecte 50 % et Baker (1992b) 30, 40 et 50 % à mesure que l'intervalle s'allonge. Corak (1993) applique l'algorithme de Baker aux données de l'EPA. Baker et Trivedi (1985) notent que, même s'il est préférable d'opter pour les intervalles les plus courts possibles pour calculer la durée moyenne, il y a peut-être un compromis à faire : plus les intervalles sont courts, plus la préférence en matière de chiffres devient évidente et, par conséquent, plus les résultats deviennent dépendants des hypothèses (arbitraires) de lissage retenues. Autrement dit, les intervalles plus longs peuvent rendre la statistique moins efficace, mais ils peuvent aussi réduire la distorsion causée par l'erreur de mesure.



La pointe relevée à 99 semaines représente à la fois l'effet de la préférence en matière de chiffres et la troncation de la distribution imputable à la limite de codage supérieure imposée par les administrateurs de l'enquête. Il faut donc avancer une hypothèse sur la façon de réaffecter cette pointe aux intervalles adjacents. Sider (1985), Baker (1992a, 1992b) et Corak (1993) fondent leur méthode de lissage sur l'hypothèse que l'on retrouve la moitié des répondants à 99 semaines à cause d'un biais dans les réponses.

Dans la littérature existante, les coefficients de lissage ont été choisis arbitrairement. Toutefois, la structure de l'EPA permet d'analyser de plus près la nature de ce biais dans les réponses. Elle prévoit un mécanisme de renouvellement selon lequel on interroge des répondants pendant six mois consécutifs avant de passer à d'autres. Paul (1986) apparie les dossiers de réponses individuelles relatives aux mois déclarés dans des mois d'enquête adjacents pour examiner le biais dans les semaines de chômage déclarées. Elle constate que les chômeurs qui sont considérés comme en recherche d'emploi (soit environ 89 % des chômeurs répondants) déclarent, mois après mois, la même durée de chômage 67,8 % du temps (pendant la période de 1979 à 1982). Paul encode les réponses sur la durée du chômage dans des intervalles de quatre semaines et définit une réponse cohérente si le dossier apparié révèle une augmentation d'un intervalle d'un mois à l'autre. Le changement de durée d'un mois à l'autre s'établissait en moyenne à 2,9 semaines pour les répondants qui ne déclaraient pas toujours la même durée, soit moins que les quatre semaines prévues. Ainsi, en moyenne (pour environ 30 % de l'échantillon), on note que les répondants ont tendance à déclarer une période de chômage inférieure à la réalité. C'est aussi le cas des personnes qui perdaient leur emploi à cause d'une mise en disponibilité temporaire. D'après ces constatations, on peut conclure sommairement que le coefficient approprié pour lisser les données de l'EPA peut être de l'ordre de 30 % et que la redistribution doit se faire dans le sens d'un allongement des périodes de chômage.

Résultats

Le tableau 1 donne un aperçu de nos résultats. La valeur moyenne de la statistique de durée entre 1977 et 1993 est présentée selon l'hypothèse retenue, c'est-à-dire l'état stable ou l'état non stable, et le degré de lissage des données. Le coefficient de lissage est appliqué à toutes les données situées aux limites des intervalles de durée, sauf celles à 99 semaines, qui ne sont pas lissées du tout. Sur la foi des résultats de Paul, nous considérons qu'un coefficient de 30 % est approprié. Comme les réponses à 99 semaines reflètent à la fois un biais de réponse et la limite supérieure de codage, le lissage de ces données est examiné à part. Pour la période échantillonnale, l'estimation basée sur l'hypothèse d'état stable est d'environ 0,4 à 0,6 semaine inférieure à l'estimation fondée sur l'hypothèse d'état non stable, alors que le degré de lissage peut entraîner une différence de plus de trois semaines. Dans le reste du document, nous examinons plus en détail ces deux dimensions.

Tableau 1 : Durée complète moyenne du chômage, état stable et état non stable (en semaines)

Coefficient de pondération utilisé pour corriger le biais dans les réponses	État non stable	État stable
0,0	15,2	14,8
0,2	16,3	15,8
0,3	16,9	16,4
0,4	17,6	17,0
0,5	18,3	17,7

Les données sont des moyennes de 1977 à 1993.

A) *État stable et état non stable*

Les différences annuelles entre les estimateurs d'état stable et d'état non stable (fondées sur un coefficient de 0,3 pour lisser les données) sont décrites au tableau 2. La durée moyenne du chômage pour toute la période avec l'estimateur d'état non stable s'établit à 16,9 semaines et l'erreur-type à 2,28. L'estimateur d'état stable donne une moyenne globale de 16,4 semaines et une erreur-type de 1,97 seulement, parce que l'effet cyclique est alors amorti. Les changements qui surviennent dans la fréquence du chômage en période de récession et d'expansion font que, dans l'ensemble des périodes de chômage en cours qui entrent dans le calcul fondé sur l'hypothèse d'état stable, les périodes de courte durée sont sur- ou sous-représentées. L'estimation d'état stable est donc plus faible que l'estimation d'état non stable au début des récessions (alors que la fréquence augmente) et plus élevée pendant les périodes de reprise et d'expansion (lorsque la fréquence baisse). Le plus grand écart entre ces mesures survient en 1982, au plus fort de la récession, lorsque l'estimation d'état stable est de 3,1 semaines inférieure à l'estimation d'état non stable. On note aussi des écarts importants entre les deux mesures pendant la récession du début des années 90, l'estimation d'état stable étant de 1,9 semaine inférieure à l'estimation d'état non stable. Pendant la période de reprise et d'expansion allant du milieu

à la fin des années 1980, l'estimation d'état stable est plus élevée que celle de l'état non stable, mais de tout au plus 0,4 semaine.

Une comparaison de l'élasticité de chaque mesure par rapport au taux de chômage donne un aperçu immédiat des différences de variation cyclique. Ces élasticités sont calculées à l'aide d'une régression par les moindres carrés du logarithme (naturel) de la durée moyenne du chômage sur le logarithme du taux de chômage (national désaisonnalisé) faisant appel à 204 observations mensuelles couvrant la période de janvier 1977 à décembre 1983. (Les régressions comportent aussi une constante, des indicateurs saisonniers et un terme de tendance.) L'élasticité de l'estimation d'état non stable vis-à-vis du taux de chômage est de 0,71, alors que celle de l'état stable s'établit à seulement 0,57⁵. Ces estimations appuient clairement l'observation selon laquelle l'estimation d'état stable suit un profil cyclique atténué.

L'hypothèse d'état stable influe non seulement sur le profil cyclique de l'estimation, mais aussi sur son caractère saisonnier. Nous examinons cet aspect au moyen d'une série de régressions par les moindres carrés, en nous servant de données mensuelles pour l'estimateur d'état non stable, l'estimateur d'état stable et la différence entre les deux (état non stable moins état stable), la variable dépendante. Les variables explicatives sont le taux de chômage (non désaisonnalisé), une tendance temporelle et douze variables d'indicateurs mensuels. Les valeurs prévues pour chaque mois de l'année sont présentées au tableau 3, le taux de chômage étant maintenu à sa valeur moyenne de 9,3 % pour la période et la tendance temporelle à zéro. Pour bien des mois, l'écart est assez marqué, variant de 6,1 semaines en janvier à -2,6 semaines en avril. Les écarts sont négligeables pour quelques mois seulement, par exemple février, juin et décembre.

⁵ Les erreurs-types associées à ces estimations sont de 0,033 et de 0,026 respectivement.

Tableau 2 : Durée complète moyenne du chômage, état stable et état non stable : 1977-1993

Année	État non stable (semaines)	État stable (semaines)	État non stable moins état stable (semaines)
1977	15,7	14,8	0,9
1978	15,5	15,2	0,3
1979	13,8	14,2	-0,4
1980	14,3	14,0	0,3
1981	14,5	13,8	0,7
1982	19,7	16,6	3,1
1983	20,5	19,9	0,6
1984	17,8	18,0	-0,2
1985	17,4	17,8	-0,4
1986	16,2	16,5	-0,3
1987	16,1	16,4	-0,3
1988	14,8	15,2	-0,4
1989	15,3	15,0	0,3
1990	16,0	14,9	1,1
1991	19,5	17,6	1,9
1992	20,8	18,9	1,9
1993	19,9	20,2	-0,3
Moyenne 1977-1993	16,9 (2,28)	16,4 (1,97)	0,5

Les données du tableau sont des moyennes annuelles de données mensuelles, lissées à l'aide d'un coefficient de 0,3.

() indique l'erreur-type échantillonnage.

Tableau 3 : Variations saisonnières prévues de la durée complète moyenne du chômage, état non stable et état stable

Mois	État non stable (semaines)	État stable (semaines)	État non stable moins état stable (semaines)
Janvier	20,6	14,5	6,1
Février	18,4	18,7	-0,3
Mars	19,5	17,1	2,4
Avril	14,9	17,5	-2,6
Mai	15,2	13,7	1,5
Juin	15,0	14,4	0,6
Juillet	15,6	14,2	1,4
Août	15,5	17,5	-2,0
Septembre	12,4	13,6	-1,2
Octobre	15,6	14,8	0,8
Novembre	16,3	13,7	2,6
Décembre	17,0	17,5	-0,5

Les données du tableau sont des estimations établies par régression, à l'aide des moindres carrés, de chaque variable dépendante sur les indicateurs mensuels, le taux de chômage et un terme de tendance, à partir de données mensuelles de 1977 à 1993.

B) Corrections de la préférence en matière de chiffres

L'information présentée dans le tableau 4 montre l'effet des diverses hypothèses de lissage sur les estimations d'état non stable de la durée complète moyenne. Dans tous les cas, nous pouvons présumer qu'il n'est pas nécessaire de lisser les données à 99 semaines. Comme on peut le voir au tableau 1, en utilisant dans l'algorithme de lissage des coefficients variant de 0 à 0,5, on augmente la durée moyenne des périodes de chômage d'environ 3 semaines, la faisant passer de 15,2 à 18,3 semaines. Cet effet est relativement constant dans le temps. L'élasticité de chaque mesure vis-à-vis du taux de chômage confirme que le caractère cyclique de la statistique n'est guère touché. Elle varie de 0,69 pour un facteur de pondération de 0,0 à 0,72 lorsque le facteur est de 0,5. En changeant d'hypothèse de lissage, on obtient des résultats semblables pour l'estimation d'état stable : on augmente la moyenne globale d'environ 3 semaines, la faisant passer de 4,8 semaines sans lissage, à 17,7 semaines si le coefficient est de 0,5. De la même manière, l'élasticité - taux de chômage de l'estimation d'état stable est peu touchée et varie de 0,56 lorsque le coefficient est de 0,0 à 0,58 lorsqu'il est de 0,5. Par ailleurs, la variation saisonnière de la statistique est légèrement touchée les modifications du coefficient utilisé pour lisser les données. S'il passe de 0,0 à 0,5, la durée prévue passe de 1,9 semaine en août à 4,3 semaines en janvier⁶. La variation saisonnière associée au passage du coefficient de lissage de 0,0 à 0,5 environ $\pm 1,2$ semaine autour de la moyenne annuelle.

Le choix du coefficient de lissage des données est surtout critique pour les semaines qui représentent la partie la plus dense de la distribution. En particulier, la valeur de l'estimation obtenue est influencée d'abord par le facteur retenu pour réaffecter le nombre de répondants qui déclarent quatre semaines de chômage. À la deuxième colonne du tableau 5, on peut voir les résultats de la réaffectation, à l'intervalle suivant, de 30 % des répondants qui déclarent quatre semaines de chômage, sans autre réaffectation. Ces données devraient être comparées aux résultats associés à l'application d'un coefficient de 0,3 pour toutes les semaines de transition du tableau 4 et repris à la première colonne du tableau 5. La durée moyenne globale varie de 0,3 semaine seulement. L'écart maximal entre les deux estimations, qui s'établit à 0,5 semaine, survient en 1982 et en 1993. Le lissage des données à ce point particulier change la durée moyenne globale de 1,4 semaine, la faisant passer de 15,2 à 16,6 semaines (voir la première colonne du tableau 4). Ainsi, la durée estimée dépend en grande partie du coefficient retenu pour la quatrième semaine, mais ne varie pas beaucoup en fonction des autres semaines de transition associées à des périodes plus longues. Par contre, la variation cyclique de l'estimation n'est pas du tout touchée.

⁶ Les variations saisonnières prévues pour les coefficients de lissage des données de 0,0 et de 0,5 appliquées à l'estimateur d'état non stable sont les suivantes :

	Janv.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
0,0	18,2	17,1	17,2	13,4	13,4	13,4	13,6	14,4	10,6	14,3	14,7	15,6
0,5	22,5	19,3	21,2	16,1	16,6	16,2	17,1	16,3	13,3	16,5	17,4	17,9

Tableau 4 : Durée complète moyenne du chômage selon l'année et la valeur du coefficient de lissage

Année	Coefficients de correction du biais de réponse				
	0,0	0,2	0,3	0,4	0,5
1977	14,1	15,1	15,7	16,4	17,0
1978	13,9	14,9	15,5	16,1	16,7
1979	12,5	13,3	13,8	14,2	14,8
1980	12,9	13,9	14,3	14,9	15,4
1981	13,2	14,1	14,5	15,0	15,6
1982	17,7	19,0	19,7	20,5	21,3
1983	18,5	19,8	20,5	21,3	22,2
1984	15,9	17,1	17,8	18,5	19,2
1985	15,6	16,7	17,4	18,1	18,8
1986	14,6	15,6	16,2	16,8	17,5
1987	14,5	15,5	16,1	16,7	17,4
1988	13,3	14,3	14,8	15,3	15,9
1989	13,7	14,8	15,3	16,0	16,6
1990	14,4	15,4	16,0	16,6	17,2
1991	17,6	18,8	19,5	20,3	21,1
1992	18,5	20,0	20,8	21,7	22,6
1993	17,7	19,1	19,9	20,8	21,7
Moyenne 1977-1993	15,2	16,3	16,9	17,6	18,3

L'effet du lissage de la pointe observée à 99 semaines et plus (limite de codage supérieure) est aussi mis en évidence au tableau 5. Les données de la troisième colonne montrent l'effet du lissage de 50 % des répondants de cette catégorie à l'intervalle précédent. C'est la procédure que Corak (1993) a suivie. Lorsque l'on compare ces données à celles de la première colonne, la durée moyenne est de 0,7 semaine plus courte et peut varier entre 1,4 semaine et 0,2 semaine seulement. Le caractère cyclique de l'estimation est légèrement touché, ce qui ramène l'élasticité vis-à-vis du taux de chômage de 0,7 à 0,65. Comme on peut le voir à la figure 3, ce point représente largement 3 % des données de l'échantillon, de sorte que tout rajustement aura une grande influence sur l'estimation globale. Pour décider du degré de lissage de cet intervalle, il faut déterminer quelles proportions des réponses dépendent respectivement de la troncation de la distribution et du biais de réponse. Si le biais a une influence négligeable, il n'est pas nécessaire de lisser les données. Puisque le lissage des réponses au-delà de la quatrième semaine a peu d'effet sur la statistique de durée, nous nous attendons à ce que la correction du biais de réponse à 99 semaines ait aussi peu d'effet. Même si le biais de réponse était aussi important à cet intervalle qu'à 52 semaines, le lissage n'aurait pas d'effet important sur l'estimation finale malgré tout.

Tableau 5 : Durée complète moyenne du chômage, selon l'année et la méthode de lissage

Année	Coefficients de correction du biais de réponse			Grand intervalle
	0,3	0,3, 0,0*	0,3, 0,5 ⁺	
1977	15,7	15,4	15,4	17,7
1978	15,5	15,2	15,2	17,8
1979	13,8	13,6	13,6	15,8
1980	14,3	14,1	14,0	16,4
1981	14,5	14,3	14,2	17,0
1982	19,7	19,2	18,7	23,1
1983	20,5	20,2	19,4	24,1
1984	17,8	17,5	17,1	21,7
1985	17,4	17,1	16,6	20,9
1986	16,2	15,9	15,5	19,7
1987	16,1	15,8	15,3	19,8
1988	14,8	14,6	14,3	18,1
1989	15,3	15,0	14,7	18,5
1990	16,0	15,7	15,3	19,2
1991	19,5	19,3	18,3	23,4
1992	20,8	20,4	19,4	25,4
1993	19,9	19,4	18,7	24,1
Moyenne 1977-1993	16,9	16,6	16,2	20,2

* indique un coefficient de 0,3 pour le premier intervalle et de 0 pour tous les autres.

+ indique un coefficient de 0,3 pour tous les intervalles et de 0,5 pour la catégorie des 99 semaines et plus.

On entend par grand intervalle le regroupement des trois premiers mois en un seul intervalle et l'utilisation d'un coefficient de 0,3.

Enfin, le tableau 5 présente aussi les résultats de l'utilisation d'intervalles plus longs pour calculer les taux de persistance du chômage. Les trois premiers mois sont regroupés dans un même intervalle. Ainsi, la variation des résultats imputable au choix du coefficient de lissage se trouvera atténuée. On applique les hypothèses de lissage standard pour les autres intervalles (soit 0,3 pour toutes les limites des intervalles et 0,0 pour l'intervalle de 99 semaines et plus). Même si l'estimation qui en résulte ne présente pas de variation cyclique par rapport à l'estimation standard, la durée est nettement plus longue, passant de 16,9 à 20,2 semaines. En fait, cette méthode a pour effet d'accroître le niveau à tel point que, pour bien des années, il est supérieur à la durée incomplète moyenne et, par conséquent, contredit le théorème proposé par Salant dont il a été question précédemment. L'ampleur de cette augmentation souligne l'importance d'utiliser les intervalles les plus courts possibles, au moins pour la portion de gauche de la distribution. Puisque les taux de persistance ont tendance à augmenter selon la durée du chômage, le regroupement des trois premiers mois surestime le taux du premier mois et donne lieu à une estimation de la durée moyenne plus longue. Dans nos données, la perte d'information associée à l'élargissement des intervalles est plus importante que toute amélioration susceptible de découler d'une atténuation du biais dans les réponses. Toutefois, il y a lieu de signaler que Baker et Trivedi (1985) sont arrivés à la conclusion opposée avec des données relatives à l'Australie.

Mise en garde

L'estimateur d'état non stable repose sur l'hypothèse selon laquelle la situation économique actuelle se maintiendra à l'avenir. En particulier, on présume que les taux de persistance, calculés d'après l'expérience que les chômeurs ont eue récemment du marché du travail (c'est-à-dire jusqu'à il y a un an), se maintiendront pendant toute la période de chômage. Cette hypothèse n'est pas aussi restrictive que celle de l'état stable, mais elle peut néanmoins porter à croire que la statistique ne sera pas parfaitement exacte. Si le marché du travail se détériore, les taux de persistance devraient augmenter avec le temps et faire en sorte que la durée complète moyenne soit inférieure à la moyenne véritable, alors que s'il s'améliore, les taux de persistance devraient décroître et faire en sorte que la valeur estimée soit supérieure à la moyenne véritable.

Pour évaluer l'ampleur de l'erreur inhérente à cette hypothèse, nous devons calculer la durée moyenne du chômage d'après l'expérience réelle d'une cohorte de chômeurs. Nous calculons cette statistique en augmentant les taux de persistance de référence de la manière suivante :

$$TrueAvgDur(t) = \sum_{x=1}^n \prod_{i=1}^x S(i, t+i) \quad (2)$$

Contrairement à l'équation (1), que l'on pourrait qualifier d'estimateur «rétrospectif», celle-ci décrit un estimateur «prospectif». Elle peut être calculée à partir des taux de persistance utilisés précédemment, mais en suivant l'expérience d'une cohorte dans le temps. Comme il faut utiliser les intervalles futurs $n-1$, il est impossible d'effectuer le calcul pour les plus récentes périodes. Les statistiques qui en résultent revêtent un intérêt strictement historique.

Les moyennes annuelles sont présentées au tableau 6 (en fonction d'un coefficient de lissage de 0,3 pour tous les intervalles et de 0,0 pour le dernier). L'estimation comparable d'état non stable (prospective) tirée du tableau 2 est également fournie à titre de référence. La durée complète moyenne «véritable» suit de près le cycle économique, grimpant en flèche en 1981 pour atteindre un sommet de 19,1 semaines en 1982 et un creux de 14,8 semaines au cours de 1987. À l'opposé, les points de retournement de l'estimateur d'état non stable se produisent un an plus tard, le sommet de 20,5 semaines étant obtenu en 1983 et le creux de 14,8 semaines en 1988. Tel que prévu, l'estimateur rétrospectif sous-estime la durée moyenne pendant la récession et la surestime pendant la reprise. Les différences entre les deux statistiques peuvent être importantes aux points de retournement du cycle économique. Au cours de 1981, l'estimation rétrospective est inférieure de 3,0 semaines à l'estimation prospective, mais elle lui est supérieure de 2,5 semaines en 1983. De la même manière, au début de la dernière récession en 1990, l'estimation rétrospective était inférieure de 3,4 semaines à l'estimation prospective.

Le tableau 7 présente des résultats semblables à ceux du tableau 3 sur la variation saisonnière des statistiques. À certains points de l'année, surtout entre janvier et avril mais aussi à l'automne, la différence entre les deux estimations peut être assez importante.

L'estimation rétrospective dépasse l'autre de 3 semaines en janvier et de 5,5 semaines en mars, mais elle lui est inférieure de 3 semaines en septembre. Ces différences ressemblent à celles qui sont associées au cycle économique, à savoir que les moins bonnes conditions qui règnent en hiver et la conjoncture plus favorable de l'été dureront plus longtemps que ce n'est réellement le cas.

Tableau 6 : Durée complète moyenne du chômage, estimations prospective et rétrospective, 1977-1991

Année	Prospective (semaines)	Rétrospective (semaines)	Différence
1977	15,2	15,7	0,5
1978	14,8	15,5	0,7
1979	14,0	13,8	-0,2
1980	14,3	14,3	0,0
1981	17,5	14,5	-3,0
1982	19,1	19,7	0,6
1983	18,0	20,5	2,5
1984	16,8	17,8	1,0
1985	15,8	17,4	1,6
1986	15,2	16,2	1,0
1987	14,8	16,1	1,3
1988	14,9	14,8	-0,1
1989	16,6	15,3	-1,3
1990	19,4	16,0	-3,4
1991	21,1	19,5	-1,6
1977-1991	16,5	16,9	0,4

Les données du tableau sont des moyennes annuelles calculées à partir de données mensuelles, auxquelles on a appliqué un coefficient de lissage de 0,3.

Tableau 7 : Variations saisonnières prévues de la durée complète moyenne du chômage, estimations prospective et rétrospective

Mois	Prospective (semaines)	Rétrospective (semaines)	Différence
Janvier	17,6	20,6	3,0
Février	17,3	18,4	1,1
Mars	14,0	19,5	5,5
Avril	13,7	14,9	1,2
Mai	15,1	15,2	0,1
Juin	15,4	15,0	-0,4
Juillet	15,9	15,6	-0,3
Août	14,8	15,5	0,7
Septembre	15,4	12,4	-3,0
Octobre	17,3	15,6	-1,7
Novembre	16,9	16,3	-0,6
Décembre	16,7	17,0	0,3

Les données du tableau ont été calculées par régression de chaque variable dépendante selon la méthode des moindres carrés sur des indicateurs mensuels, le taux de chômage et un terme de tendance, à l'aide de données mensuelles de 1977 à 1993 pour l'estimation rétrospective et de 1977 à 1991 pour l'estimation prospective.

Conclusions

Dans cet article, nous examinons divers aspects du calcul de la durée complète moyenne du chômage. Nous obtenons trois résultats principaux. Tout d'abord, la durée complète moyenne du chômage est un indicateur plus fiable des conditions du marché du travail que la durée incomplète moyenne. Elle donne un aperçu de la période que les nouveaux chômeurs peuvent s'attendre à consacrer à la recherche d'un emploi et des fluctuations qui correspondent aux points de retournement du cycle économique. La durée incomplète moyenne, par contre, est un indicateur cyclique retardé. Ensuite, l'estimateur d'état non stable de la durée complète moyenne est préféré à l'estimateur d'état stable. Le premier est nettement plus court que le second au début des récessions et affiche un profil saisonnier différent. L'hypothèse d'état stable facilite certes les calculs, mais cet avantage n'est pas si considérable. Enfin, les hypothèses retenues pour lisser les données afin de corriger les erreurs dans les réponses sur la durée du chômage influent beaucoup sur la statistique de durée et, dans une moindre mesure, sur ses variations saisonnières, mais non sur son profil cyclique. Le lissage est surtout important pour la partie la plus dense de la distribution, généralement à la quatrième semaine, mais aussi pour toutes les limites de codage imposées par les administrateurs de l'enquête.

Ces résultats nous amènent à conclure que la statistique qui devrait être privilégiée pour diffusion officielle, du moins par Statistique Canada, est l'estimation d'état non stable utilisant un coefficient de lissage de 0,3. Ce coefficient doit absolument être appliqué à la quatrième semaine, mais son influence est négligeable pour les autres points. Ainsi, nous suggérons de ne pas lisser les données pour corriger l'erreur de mesure à la limite de codage supérieure de 99 semaines et plus. Même si l'estimateur d'état non stable est supérieur à l'estimateur d'état stable, il suppose la stabilité de la situation économique actuelle (représentée par les conditions observées dans un passé récent) tout au long de la période de chômage. L'examen de cette hypothèse porte à croire que l'estimateur d'état non stable surestimerait la durée du chômage en début de reprise et le sous-estimerait en début de récession. Nous notons un biais semblable associé aux variations saisonnières. Il faut donc faire preuve d'une certaine prudence dans l'interprétation de la statistique.

La prudence est également de mise lorsqu'on veut étendre ces recommandations aux données d'autres pays. L'hypothèse d'état non stable est nettement préférable à l'autre, mais le choix du coefficient de lissage est moins clair. D'après la littérature existante, ce choix semble être arbitraire. Pour nos données, un coefficient de 0,3 est justifiable, compte tenu de l'analyse des tendances relevées dans les réponses des participants à l'enquête, lesquelles peuvent varier d'un pays à l'autre. Cet aspect et le traitement de la limite de codage supérieure devraient faire l'objet d'un examen plus approfondi avant qu'on puisse généraliser les conclusions présentées ici.

Bibliographie

BAKER, G.M. et P.K. Trivedi (1985), «Estimation of Unemployment Duration from Grouped Data: A Comparative Study», *Journal of Labor Economics*, vol. 3, pp. 153-174.

BAKER, Michael (1992a), «Unemployment Duration: Compositional Effects and Cyclical Variability», *American Economic Review*, vol. 82, pp. 313-21.

_____ (1992b), «Digit Preference in CPS Unemployment Data», *Economics Letters*, vol. 39, pp. 117-21.

CORAK, Miles (1993), «La durée du chômage en période prospérité et de récession», *l'Observateur économique canadien*, publication n° 11-010 au catalogue de Statistique Canada. (Septembre), pp. 4.1-4.20.

KAITZ, H.B. (1970), «Analysing the Lengths of Spells of Unemployment», *Monthly Labor Review*, vol. 93, pp. 11-20.

KAPLAN, E.L. et P. Meier (1958), «Nonparametric Estimation from Incomplete Observations», *Journal of the American Statistical Association*, vol. 53, pp. 457-81.

PAUL, Betty Clayton (1986), *A Compendium of Studies on Data Quality of LFS Data on the Duration of Unemployment*, Statistique Canada, rapports du personnel n° 86-2 de la Division de l'analyse du marché du travail et des ménages.

SALANT, Stephen (1977), «Search Theory and Duration Data: A Theory of Sorts», *Quarterly Journal of Economics*, vol. 91, pp. 39-57.

SIDER, Hal (1985), «Unemployment Duration and Incidence: 1968-82», *American Economic Review*, vol. 75, pp. 461-72.

STATISTIQUE CANADA (1992), *Guide d'utilisation des données de l'Enquête sur la population active*, publication n° 71-528 au catalogue.