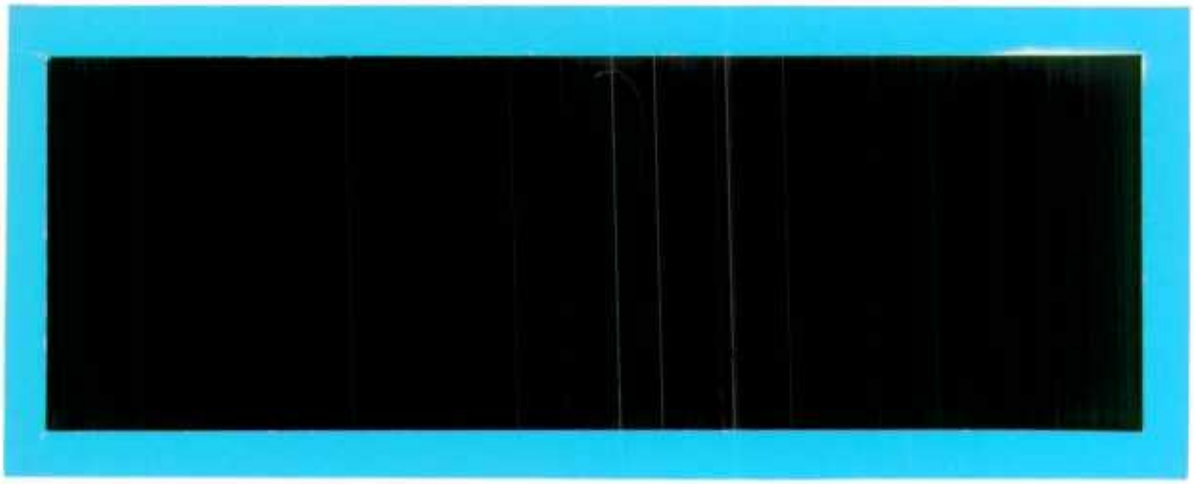
11-613F
no.88-24
c.2

atistique
nada



Methodology Branch

Social Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes sociales

Canada

WORKING PAPER NO. SSMD-88-024 F

METHODOLOGY BRANCH

c. 2

# MODÈLES DES ERREURS D'ÉCHANTILLONNAGE

SSMD-88-024 E

J. Dufour et G. Lemaître

STATISTICS CANADA STATISTIQUE CANADA
OCT 18 1994
LIBRARY
BIBLIOTHÈQUE

## ABSTRACT

This paper presents a number of possible generalized variance functions (GVF's) that can be used to model variances estimates of Labour Force Survey characteristics as functions of the estimates. The characteristics are classified into groups on the basis of the homogeneity of design effects in order to reduce the number of models. Different formulae are developed, depending on models and groups, to determine coefficients of variation of monthly estimates, month-to-month and year-to-year changes and averages of monthly estimates at the provincial level. Multiplicative factors needed to obtain these CV's at the subprovincial level are also developed.



## RÉSUMÉ

Ce document traite de différentes fonctions de variance généralisées (FVG) dans le but de modéliser les variances estimées pour un certain nombre de caractéristiques de l'enquête sur la population active en fonction de ces mêmes caractéristiques. Afin de restreindre le nombre de modèles possibles, les caractéristiques sont regroupées de façon à respecter l'homogénéité de l'effet de plan de sondage dans chacun des groupes. Différentes formules sont développées selon les modèles et les groupes retenus pour diverses caractéristiques afin de déterminer les coefficients de variation d'estimations mensuelles, de variations mensuelles ou annuelles, de moyennes trimestrielles ou annuelles au niveau provincial. On détermine certains facteurs multiplicatifs permettant également d'estimer cette gamme de coefficients de variation pour des estimations correspondant à des unités géographiques plus petites, c'est-à-dire au niveau infraprovincial.

Table des matières


1. Introduction 1

2. Modèles FVG 2

3. Regroupement des caractéristiques 6

3.1 Analyse visuelle 6
3.2 Comparaisons multiples 7

4. Analyse des modèles 8

5. Aspects de la construction des Tableaux de référence des Coefficients de variation 9

5.1 Tableau de référence des CV d'une variation mensuelle au niveau provincial et pour le Canada 9

5.2 Tableau de référence des CV d'une variation mensuelle au niveau provincial et pour le Canada 10

5.3 Tableau de référence des CV d'une moyenne de niveaux selon la province et pour le Canada 12

5.4 Tableau de référence des CV d'une variation de moyennes annuelles au niveau provincial et pour le Canada 16

5.5 Tableaux de référence des CV d'une estimation mensuelle, d'une moyenne de niveaux, d'une variation mensuelle, ou d'une variation de moyennes annuelles au niveau infraprovincial 18

6. Exemple d'utilisation des tableaux de référence des coefficients de variation 21

7. Discussion sur la qualité des tableaux de référence des coefficients de variation 22

8. Conclusion 24

Bibliographie 26

Annexe A 27

Annexe B 35

**Modèles des erreurs d'échantillonnage**

J. Dufour et G. Lemaître

## 1. Introduction

L'enquête sur la population active (EPA) est la plus vaste enquête-ménages menée par Statistique Canada. Les renseignements obtenus au cours de l'enquête permettent de publier une grande variété d'estimations. A chacune de ces estimations est associée une erreur d'échantillonnage qu'il est utile de connaître. Or le calcul des erreurs d'échantillonnage pour les enquêtes à plan complexe étant généralement coûteux; il n'est possible d'effectuer ce calcul que pour un nombre restreint de caractéristiques. Pour pallier à ce problème, des fonctions de variances généralisées (FVG) qui donnent une approximation de la variance d'une estimation quelconque comme fonction simple de l'estimation ont été développées. Ces fonctions sont obtenues généralement en modélisant les variances estimées pour un certain nombre de caractéristiques en fonction des estimations de ces mêmes caractéristiques.

Dans la littérature, on retrouve les travaux de Hansen, Hurwitz et Madow (1953), Kalton (1977), Hanson (1978), Ghangurde (1981) ainsi que Wolter (1985). Les travaux de ces auteurs traitent de fonctions de variances généralisées pour des estimations de niveaux ou de taux; les modèles retenus pouvant différer quelque peu selon l'auteur. Par exemple, le CPS (Current Population Survey) et le NHIS (National Health Interview Survey) utilisent un modèle FVG différent de celui utilisé par l'EPA (voir Ghangurde (1981)), ce dernier étant similaire au modèle employé au Bureau de la Statistique d'Australie. Une étude comparative de ces modèles est disponible dans Valliant (1987). Dans ce qui suit, on traite des modèles FVG permettant d'estimer l'erreur d'échantillonnage d'un niveau et on introduit des modèles FVG utiles pour estimer cette même erreur pour des variations mensuelles ou annuelles, des moyennes trimestrielles, semestrielles ou annuelles au niveau provincial. On discute également de facteurs multiplicatifs permettant de convertir l'estimation de ces erreurs obtenues au niveau provincial à un niveau infraprovincial.

La deuxième section de ce rapport décrit les modèles FVG testés. On retrouve à la troisième section l'explication des regroupements des

caractéristiques pour lesquelles on désire estimer l'erreur d'échantillonnage de façon à réduire le nombre de modèles. La quatrième section fait part d'une analyse critique des modèles suggérés, tandis qu'à la section 5 on retrouve l'explication de divers tableaux construits à l'aide du modèle retenu pour les utilisateurs des données de l'EPA. On retrouve à la section 6 un exemple d'utilisation des tableaux de référence des coefficients de variation et la section 7 discute de la qualité des tableaux. La section 8 sert en guise de conclusion.

## 2. Modèles FVG

L'EPA utilise un plan de sondage aréolaire à plusieurs degrés (voir Platek et Singh (1976)). Le pays est d'abord divisé en provinces et par la suite chacune d'elles sont divisées en régions économiques, c'est-à-dire en secteurs possédant des structures économiques semblables. Chaque région économique est formée d'unités autoreprésentatives (UAR), soit de villes dont la population se situe entre 10,000 et 25,000 personnes selon la province et d'unités non-autoreprésentatives (UNAR), c'est-à-dire les régions situées à l'extérieur des UAR. On utilise par la suite deux types de plan de sondage à plusieurs degrés qui diffèrent selon qu'il s'agit des UAR ou des UNAR. La dernière étape de l'échantillonnage dans les deux régions consiste à sélectionner des logements à l'intérieur des grappes.

Afin d'estimer la variance des totaux estimés, que l'on note $V(\hat{X})$, obtenus à l'aide d'un tel plan de sondage pour une caractéristique à caractère dichotomique quelconque, on suppose que les unités primaires d'échantillonnage (UPE) sont sélectionnées de façon indépendante et on fait appel à la méthode élaborée par Keyfitz (1957) adaptée aux exigences de l'EPA (voir Choudhry et Lee (1987)). La notion d'effet de plan de sondage introduite par Kish (1965) permet de relier cette variance basée sur un plan complexe à la variance basée sur un échantillon aléatoire simple de taille $n$ tiré d'une population finie de taille $P$ de la façon suivante:

$$V(\hat{X}) \simeq F\left(\frac{P}{n} - 1\right) \hat{X} \left(1 - \frac{\hat{X}}{P}\right),$$

où $F$: effet de plan de sondage,

P: taille de la population

$\hat{X}$: estimation du total $X$ d'une caractéristique à caractère dichotomique

et n: taille de l'échantillon.

Le coefficient de variation $CV(\hat{X})$ (voir Cochran (1977)) correspondant à une estimation $\hat{X}$ de la caractéristique $X$ s'exprime en fonction de l'écart-type et de la caractéristique elle-même,

$$CV(\hat{X}) = \frac{[Var(\hat{X})]^{\frac{1}{2}}}{X} \simeq \frac{[F(\frac{P}{n} - 1)\ \hat{X}\ (1 - \frac{\hat{X}}{P})]^{\frac{1}{2}}}{\hat{X}} \quad . \qquad (1)$$

Il est à remarquer qu'en présence de $F$, $P$ et $n$ connus, on peut exprimer le $CV(\hat{X})$ en fonction de $\hat{X}$ seulement. On désire déterminer un modèle FVG c'est-à-dire un modèle qui s'ajuste bien aux estimations $CV(\hat{X})$ et $\hat{X}$ dans le but de construire des tableaux de référence des coefficients de variation en fonction de diverses valeurs possibles de $\hat{X}$. On applique une transformation logarithmique à l'équation (1) et on obtient

$$\ln CV(\hat{X}) \simeq \frac{1}{2} \ln\left(F(\frac{P}{n} - 1)\right) - \frac{1}{2} \ln\left(\frac{\hat{X}}{1 - \frac{\hat{X}}{P}}\right). \qquad (2)$$

De l'équation (2), on envisage deux modèles postulant certaines hypothèses. Dans un premier temps, on considère le terme $F(P/n - 1)$ comme étant constant et on obtient le modèle

$$\ln \hat{CV}(\hat{X}) = a + b \ln\left(\frac{\hat{X}}{1 - \frac{\hat{X}}{P}}\right) + \varepsilon \qquad (3)$$

où $\varepsilon$ représente le terme d'erreur. Ce modèle nécessite toutefois une estimation de $P$, la taille de la population. Théoriquement, on reconnaît que la valeur de $b$ doit égaler -1/2. On s'attend donc à obtenir une estimation de $b$ au voisinage de -1/2.

Deuxième hypothèse postulée, si l'estimation $\hat{X}$ est petite par rapport à la population $P$ alors $\hat{X}/P$ peut être considéré comme étant nul et par le fait même le dénominateur du deuxième terme du membre de droite est environ égal à un, d'où

$$\ln \hat{CV}(\hat{X}) = a + b \ln(\hat{X}) + \varepsilon \quad . \tag{4}$$

On a également testé un modèle autorégressif d'ordre 1; soit

$$\ln \hat{CV}(\hat{X}_t) = a + b \ln\left[\frac{\hat{X}_t}{1 - \frac{\hat{X}_t}{P_t}}\right] + \varepsilon_t \tag{5}$$

où $\varepsilon_t = \rho\varepsilon_{t-1} + \mu_t$, $t$ représente le temps et $\mu_t$ est un bruit blanc, $\mu_t \sim N(0, \sigma^2)$. On a à nouveau posé l'hypothèse précédente à savoir $\hat{X}_t/P_t \to 0$ pour obtenir

$$\ln \hat{CV}(\hat{X}_t) = a + b\ln(\hat{X}_t) + \varepsilon_t \quad . \tag{6}$$

Les modèles (3) et (4) reposent sur l'hypothèse de l'indépendance des erreurs. Toutefois lors de l'évaluation des modèles, on utilise vingt-deux estimations mensuelles pour chacune des caractéristiques et de plus deux mois consécutifs possèdent 5/6 de l'échantillon en commun. Les modèles (5) et (6) servent dans un deuxième temps comme palliatif à ce problème d'autocorrélation si effectivement il est présent.

Quelques modifications ont dû être apportées aux modèles (3), (4) (5) et (6). En effet, les données utilisées pour cette étude proviennent de l'échantillon du mois de mars 1985, ce qui correspond au moment de l'implantation du nouveau plan de sondage, jusqu'au mois de décembre 1986. Or au mois de juin 1986, la taille de l'échantillon s'est vue réduite d'environ 10%. On a donc travaillé avec toutes ces données afin d'obtenir une meilleure estimation du paramètre b, mais on a porté une attention particulière au terme constant. En effet, $a = F(P/n - 1)$ dépend de la taille de l'échantillon. Ainsi, les modèles deviennent

$$\ln \hat{CV}(\hat{X}) = (a\,\delta_1 + a'\delta_2) + b \ln\left[\frac{\hat{X}}{1 - \frac{\hat{X}}{P}}\right] + \varepsilon \quad , \tag{7}$$

$$\ln \hat{CV}(\hat{X}) = (a\,\delta_1 + a'\delta_2) + b \ln \hat{X} + \varepsilon \ , \tag{8}$$

$$\ln \hat{CV}(\hat{X}_t) = (a\,\delta_1 + a'\delta_2) + b \ln\left[\frac{\hat{X}_t}{1 - \frac{\hat{X}_t}{P_t}}\right] + \varepsilon_t \tag{9}$$

et

$$\ln \hat{CV}(\hat{X}_t) = (a\,\delta_1 + a'\delta_2) + b \ln(\hat{X}_t) + \varepsilon_t \tag{10}$$

où

$$\delta_1 = \begin{cases} 1 & \text{si le mois } (t) \text{ correspond à une période se situant avant mai 1986 (inclusivement)} \\ 0 & \text{sinon} \end{cases}$$

$$\delta_2 = \begin{cases} 1 & \text{si le mois } (t) \text{ correspond à une période se situant apès mai 1986} \\ 0 & \text{sinon.} \end{cases}$$

Ces modèles FVG sont évalués à la section 4. Il est à noter que les modèles (7) et (8) ressemblent à quelques détails près aux modèles étudiés par Ghangurde (1981).

D'autres types de modèles ont été explorés ou sont présentement utilisés. Par exemple, le CPS utilise depuis 1947, un modèle de la forme

$$\hat{CV}(\hat{X}) = (\alpha + \beta/\hat{X})^{\frac{1}{2}} + \varepsilon$$

pour des estimations $\hat{X}$ de moyennes ou de caractéristiques à caractère dichotomique telle le nombre de personnes ayant un emploi (voir Hansen, Hurwitz et Madow (1953), Hanson (1978)). Le NHIS pour sa part utilise un tel modèle pour mesurer la fiabilité d'une estimation de moyennes, de caractéristiques à caractère dichotomique ou pour des variables quantitatives. Valliant (1987) établit une comparaison entre ces différents modèles. Wolter (1981) pour sa part énumère d'autres modèles possibles tels

$$\hat{CV}(X) = (\alpha + \beta/\hat{X} + \alpha/\hat{X}^2)^{\frac{1}{2}} + \varepsilon,$$

$$\hat{CV}(X) = (\alpha + \beta\hat{X})^{-\frac{1}{2}} + \varepsilon,$$

ou bien $$\hat{CV}(X) = (\alpha + \beta\hat{X} + \gamma\hat{X}^2)^{-\frac{1}{2}} + \varepsilon.$$

Toutefois aucune analyse pratique associée à ces modèles n'est discutée. Les modèles précités sont toutefois exclus de notre analyse.

## 3. Regroupement des caractéristiques

La variable utilisée lors du regroupement des caractéristiques consiste en l'effet de plan de sondage F. Étant donné l'influence de la répartition de l'échantillon et par le fait même de la réduction de la taille de l'échantillon sur l'effet de plan de sondage, on ne peut travailler avec toutes les données de mars 1985 à décembre 1986 afin de déterminer les regroupements, c'est pour cette raison qu'on se limite aux données de mars 1985 à mai 1986. On a envisagé deux méthodes de regroupement: une méthode visuelle et une méthode basée sur des comparaisons multiples.

### 3.1 Analyse visuelle

Pour chaque province et pour chacune des 65 caractéristiques retenues pour cette étude, on a calculé la moyenne des effets de plan de sondage à l'aide des données de mars 1985 à mai 1986. Les 65 caractéristiques se composent de la population active, le nombre d'employés, le nombre de chômeurs (et ce pour 15 groupes d'âge à l'exception des 65+ pour les chômeurs), le nombre d'employés selon la branche d'activité (10 classes possibles), le nombre d'employés selon la profession (9 classes possibles) et s'il s'agit d'un emploi à temps plein ou à temps partiel. On a classé les 65 effets de plan de sondage moyens par ordre croissant et on les a portés sur un graphique. On a essayé de déterminer des bornes afin de constituer des groupes. Ces bornes sont choisies de façon à ce que l'hypothèse de l'égalité de l'effet de plan soit satisfaite le plus possible dans les premiers groupes et que le dernier groupe contienne les caractéristiques pour

lesquelles l'égalité des effets de plan est moins respectée. De façon visuelle, il est très difficile de déterminer ces bornes. Cette méthode nous donne une idée des bornes plutôt que la borne elle-même. Voulant une procédure plus statistique on a opté pour les comparaisons multiples.

### 3.2 Comparaisons multiples

On a considéré deux différents tests statistiques afin de déterminer les regroupements. Il s'agit en fait des tests de Scheffé et de Tukey. De ces tests on a retenu celui de Tukey, soit celui qui donne le plus de regroupements. C'était de plus celui qui représentait le mieux ce que l'on souhaitait obtenir, c'est-à-dire que les regroupements suivaient un patron plus ou moins uniforme d'une province à l'autre quant au nombre de groupes et d'éléments dans chacun d'eux. A l'aide du test de Tukey, les provinces et le Canada ont été divisés en quatre groupes. Le dernier groupe étant constitué des effets de plan pour lesquels l'égalité de moyennes des effets de plan est plutôt rudimentaire. Le test de Tukey s'énonce comme suit: On désire confronter les hypothèses

$$H_0: \mu_i - \mu_j = \lambda = 0 \qquad i \leq j$$

et

$$H_1: \lambda \neq 0,$$

$\mu_i$ étant la moyenne des effets de plan de sondage pour la caractéristique i. Pour une province fixée, on définit $F_{ik}$ comme étant l'effet de plan de sondage pour la caractéristique i au mois k. On note $\delta = F_{i.} - F_{j.}$ où $F_{i.} = \sum_{k=1}^{15} F_{ik}/15$.

L'intervalle de confiance au niveau $\alpha$ pour $(\mu_i - \mu_j)$ est donné par

$$I = \delta \pm k_\alpha \left(\frac{V_R}{15}\right)^{\frac{1}{2}}$$

où

$$V_R = \sum_{i=1}^{r} \sum_{j=1}^{15} \frac{(F_{ij} - F_{i.})^2}{N - r}$$

et $k_\alpha$ qui dépend de $(\alpha, r, N - r)$ s'obtient dans une table de Tukey (voir Neter, Wasserman et Kutner (1985)). Les paramètres $\alpha$, r et N représentent respectivement le niveau de confiance, le nombre de caractéristiques (au nombre de 65) et le nombre total d'observations. Autrement dit, si $\mu_{(1)}, \mu_{(2)}, \ldots, \mu_{(65)}$

représentent les moyennes des effets de plan de sondage ordonnées alors on peut à l'aide du test de Tukey déterminer des regroupements. Par exemple, l'obtention de $\mu_{(1)} = \mu_{(k)}$ et $\mu_{(k+1)} = \mu_{(65)}$ indique l'existence de deux regroupements; les bornes du premier groupe étant $(\mu_{(1)}, \mu_{(k)})$ et $(\mu_{(k+1)}, \mu_{(65)})$ pour le second. On a ainsi formé quatre regroupements des 65 caractéristiques retenues pour chaque province et pour le Canada (Voir les tableaux I et II à l'annexe A à cet effet).

4. Analyse des modèles

Les résultats obtenus lors de l'évaluation des modèles (7) et (8) sont données à l'annexe A par l'intermédiaire des tableaux III et IV, soit l'estimation des constantes, du paramètre b ainsi que la mesure d'ajustement $R^2$. Il est à remarquer que les constantes a et $a'$ ont été divisées en deux vecteurs de dimension quatre soit $(a_1, a_2, a_3, a_4)$ et $(a'_1, a'_2, a'_3, a'_4)$ respectivement afin d'estimer une constante qui dépend de l'effet de plan de sondage dans chacun des quatre groupes.

Les deux modèles ne tenant pas compte du facteur d'autocorrélation affichent des valeurs de $R^2$ très élevées (Tableaux III et IV, Annexe A), ce qui indique une bonne qualité d'ajustement. Une analyse des résidus de ces modèles permet de constater une variance constante dans les erreurs. Toutefois, malgré la qualité de ces deux modèles, le modèle (7) semble le plus se rapporter à la théorie dans le sens que les estimations obtenues pour le paramètre b se situent au voisinage de - 0,5 (Tableau III, Annexe A), tandis que celles obtenues avec le modèle(8) se situent aux environs de - 0,6 (Tableau IV, Annexe A).

Pour ce qui est des modèles corrigés en fonction de l'effet d'autocorrélation, soit les modèles (9) et (10), les résultats s'apparentent bien avec ceux obtenus avec les modèles (7) et (8) dans seulement quelques occasions. Dans plusieurs provinces, les estimations obtenues pour le paramètre b surestiment la valeur théorique - 0,5. Ces modèles tenant compte de l'autocorrélation sont mis à l'écart pour cette étude.

Si on revient aux modèles de départ, soit le (7) et (8), notre choix lors de la construction des tableaux de référence s'arrête au modèle (8). Ce choix est

motivé par la simplicité de ce modèle, c'est-à-dire par le fait que le modèle ne requiert aucune estimation de $P$, ainsi que par les bons résultats obtenus.

## 5. Aspects de la construction des Tableaux de référence des Coefficients de variation

Il arrive très souvent d'être confronté à de nouvelles demandes des utilisateurs des données de l'EPA. On désire par exemple connaître le coefficient de variation (CV) d'une caractéristique quelconque ou encore le CV d'une variation d'une caractéristique pour deux mois consécutifs. Pour satisfaire à ces demandes, on a construit ce que l'on appelle des tableaux de référence (voir Dufour et Lemaître (1987b)). On retrouve également les instructions se rattachant à ces tableaux de référence des coefficients de variation se rapportant aux caractéristiques de l'EPA dans Dufour et Lemaître (1987a).

### 5.1 Tableau de référence des CV d'une estimation mensuelle au niveau provincial et pour le Canada

Pour chacune des provinces et pour le Canada, on a construit des tableaux de référence des coefficients de variation. Ces tableaux ont été créés suivant le modèle linéaire logarithmique (8). On a d'ailleurs utilisé les constantes $(a_1', a_2', a_3', a_4')$ dans l'évaluation du modèle étant donné qu'on voulait estimer les CV pour la période après mai 1986. Ce qui signifie que pour diverses valeurs de $\hat{X}$, une estimation de la caractéristique étudiée, on a estimé le $CV(\hat{X})$ à l'aide du modèle $\ln CV(\hat{X}) = a' + b\ln(\hat{X})$ en tenant compte du groupe auquel appartient la caractéristique.

Pour se servir de ces tableaux, on doit connaître à quel groupe appartient la caractéristique étudiée. En effet, les tableaux sont donnés par province et par groupe (déterminé selon le test de Tukey). Ainsi pour chaque caractéristique désirée, on cherche dans la liste de référence des groupes par province et pour le Canada, le groupe auquel appartient la caractéristique. Par la suite connaissant le groupe, la province et une estimation de la caractéristique on se réfère au tableau approprié pour lire la valeur du coefficient de variation.

## 5.2 Tableau de référence des CV d'une variation mensuelle au niveau provincial et pour le Canada

A l'aide de ces tableaux, on désire déterminer le coefficient de variation d'un changement pour une caractéristique selon deux mois consécutifs. Soit $X$ la caractéristique qui nous intéresse et $\hat{X}$ une estimation de celle-ci. On note par $\hat{X}_1$ une estimation de $X$ au mois $t=1$ et $\hat{X}_2$ celle du mois $t=2$. Le problème consiste à estimer le coefficient de variation du changement entre les deux mois, soit de $X_2-X_1$.

Un estimateur naturel de la différence $(X_2 - X_1)$ s'écrit $(\hat{X}_2 - \hat{X}_1)$. Il ne reste plus qu'à déterminer le $CV(\hat{X}_2 - \hat{X}_1)$. Or par définition

$$CV(\hat{X}_2 - \hat{X}_1) = \frac{[Var(\hat{X}_2 - \hat{X}_1)]^{\frac{1}{2}}}{\hat{X}_2 - \hat{X}_1}$$

$$= \frac{[Var(\hat{X}_1) + Var(\hat{X}_2) - 2Cov(\hat{X}_1, \hat{X}_2)]^{\frac{1}{2}}}{(\hat{X}_2 - \hat{X}_1)} .$$

On fait l'hypothèse suivante à savoir $CV(\hat{X}_1) = CV(\hat{X}_2) = C$. Par conséquent, $Var(\hat{X}_1) = C^2\hat{X}_1^2$ et $Var(\hat{X}_2) = C^2\hat{X}_2^2$. De même,

$$Cov(\hat{X}_1,\hat{X}_2) = Corr(\hat{X}_1,\hat{X}_2)[Var(\hat{X}_1)Var(\hat{X}_2)]^{\frac{1}{2}} = \rho C^2\hat{X}_1\hat{X}_2 .$$

Substituant ces valeurs dans l'équation du CV, on obtient

$$CV(\hat{X}_2-\hat{X}_1) = \frac{C[\hat{X}_1^2 + \hat{X}_2^2 - 2\rho\, \hat{X}_1\hat{X}_2]^{\frac{1}{2}}}{(\hat{X}_2-\hat{X}_1)}.$$

Soit $r = (\hat{X}_2 - \hat{X}_1)/\hat{X}_1$, l'écart relatif pour cette caractéristique. On peut écrire

$$CV(\hat{X}_2 - \hat{X}_1) = C[1 + 2(1 - \rho)\ (\frac{1}{r} + \frac{1}{r^2})]^{\frac{1}{2}} .$$

Connaissant le CV d'une estimation mensuelle de même niveau C, le coefficient de corrélation $\rho$ de cette caractéristique et le changement relatif r, on peut facilement déterminer le $CV(\hat{X}_2 - \hat{X}_1)$. En effet, connaissant $\rho$ et r, on peut lire dans un tableau de facteurs (voir Dufour et Lemaître (1987b)) la valeur de $[1 + 2(1 - \rho)\ (1/r + 1/r^2)]^{\frac{1}{2}}$ et multiplier cette valeur par C. La lecture du facteur multiplicatif dans le tableau approprié exige néanmoins la connaissance du coefficient de corrélation. Il suffit de se référer à la liste de référence des groupes et des coefficients de corrélation par province et pour le Canada dans le document cité plus haut pour obtenir une estimation de cette valeur. Il s'agit ensuite de déterminer r et le tour est joué.

Le coefficient de variation du changement étant déterminé, il reste maintenant à savoir si le changement entre les deux mois est significatif. Pour chaque niveau de confiance désiré, on peut déterminer la zone de rejet associé à la confrontation des hypothèses $H_0$: $\hat{X}_1 = \hat{X}_2$ et $H_1$: $\hat{X}_1 \neq \hat{X}_2$ à l'aide de la statistique de Student. Plus précisément, on rejette l'hypothèse nulle, au niveau $\alpha$, lorsque

$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\left[\mathrm{Var}\ (\hat{X}_1 - \hat{X}_2)\right]^{\frac{1}{2}}} \right| \geq t_{n-1;\ \frac{\alpha}{2}} \quad ,$$

où $t_{n-1;\frac{\alpha}{2}}$, est le percentile supérieur $\alpha$ de la loi de Student et n est la taille de l'échantillon.

Puisque l'on obtient une estimation du $CV(\hat{X}_1 - \hat{X}_2)$ dans un tableau de référence alors il est plus simple de dire que l'on rejette $H_0$: $\hat{X}_1 = \hat{X}_2$, au niveau $\alpha$ lorsque

$$CV(\hat{X}_1 - \hat{X}_2) \leq \frac{1}{t_{n-1;\ \frac{\alpha}{2}}}.$$

5.3 Tableau de référence des CV d'une moyenne de niveaux selon la province et pour le Canada

Le coefficient de variation d'une moyenne arithmétique de niveaux peut s'exprimer en fonction du CV d'une estimation mensuelle de même niveau et du coefficient de corrélation. En effet, si $\theta$ est le nombre de mois pour lequel on désire une moyenne on peut écrire

$$CV\left(\sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta}\right) = \frac{\left[Var\left(\sum_{i=1}^{\theta} \hat{X}_i\right)\right]^{\frac{1}{2}}}{\sum_{i=1}^{\theta} \hat{X}_i}$$

$$= \frac{\left[\sum_{i=1}^{\theta} Var(\hat{X}_i) + 2 \sum_{i<j} Cov(\hat{X}_i, \hat{X}_j)\right]^{\frac{1}{2}}}{\sum_{i=1}^{\theta} \hat{X}_i} .$$

On fait une première hypothèse à savoir $CV(\hat{X}_i) = C \ \forall i, \ i = 1, \ldots, \theta$ et par le fait même $Var(\hat{X}_i) = C^2 \hat{X}_i^2$ et $Cov(\hat{X}_i, \hat{X}_j) = \rho_{ij} C^2 \hat{X}_i \hat{X}_j$. Par conséquent,

$$CV\left(\sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta}\right) = \frac{\left[\sum_{i=1}^{\theta} C^2 \hat{X}_i^2 + 2 \sum_{i<j} \rho_{ij} C^2 \hat{X}_i \hat{X}_j\right]^{\frac{1}{2}}}{\sum_{i=1}^{\theta} \hat{X}_i} .$$

Comme deuxième hypothèse, on suppose qu'il est possible d'estimer $X_1, \ldots, X_\theta$ par une même quantité $\bar{X}$. On a donc

$$CV\left(\sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta}\right) = \frac{C[\theta \bar{X}^2 + 2 \sum_{i<j} \rho_{ij} \bar{X}^2]^{\frac{1}{2}}}{\theta \bar{X}} = C\left[1 - 2 \sum_{i<j} \frac{(1 - \rho_{ij})}{\theta^2}\right]^{\frac{1}{2}} .$$

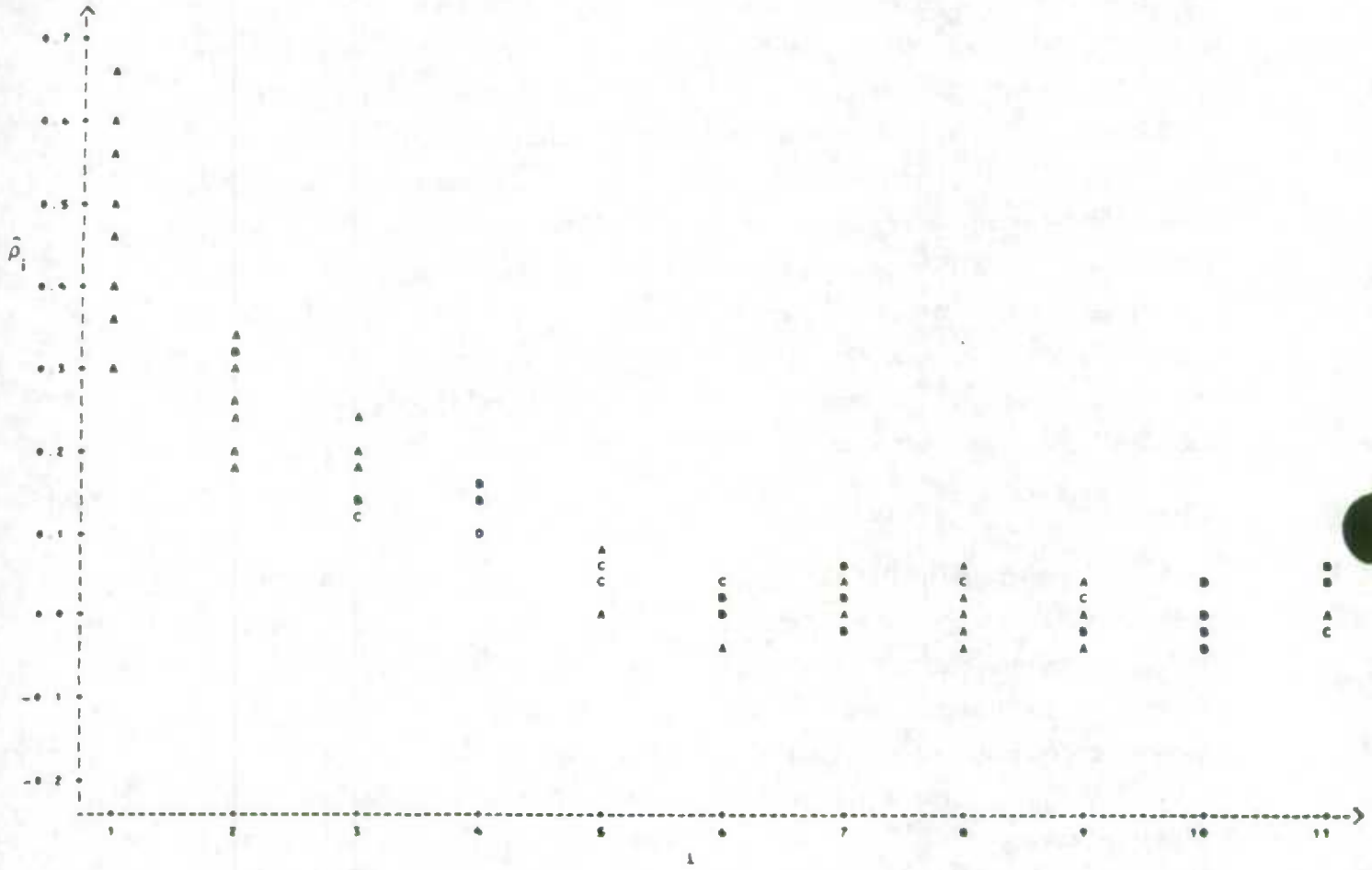
Comme dernière hypothèse, on prétend que $\rho_{ij} = \rho_{kl}$ lorsque $j-i = l-k$ et l'on note $\rho_i = Corr\ (X_t, X_{t+i})$. De cette dernière hypothèse découle l'équation générale du CV d'une moyenne de niveaux, soit

$$CV\left(\sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta}\right) = C\left[1-2\sum_{i=1}^{\theta-1} \frac{(\theta-i)}{\theta^2}(1-\rho_i)\right]^{\frac{1}{2}}. \qquad (11)$$

Afin d'évaluer cette équation, on doit connaître une estimation de $\rho_1, \ldots, \rho_{\theta-1}$. Pour des questions de temps et de ressources, on ne peut déterminer toutes ces valeurs pour chacune des caractéristiques de l'EPA retenues pour l'étude. Toutefois, les valeurs de $\hat{\rho}_1$ c'est-à-dire les estimations du coefficient de corrélation pour deux mois consécutifs pour les caractéristiques d'intérêt étant connues, on peut modéliser les $\hat{\rho}_i$, $i>1$, en fonction des $\hat{\rho}_1$. Afin de construire un modèle, on se fixe une échelle de départ constituée de treize valeurs de $\hat{\rho}_1$ soit 0,30 ; 0,35 ; 0,40 ; 0,45 ; 0,50 ; 0,55 ; 0,60 ; 0,65 ; 0,70 ; 0,75 ; 0,80 ; 0,85 ; 0,90. On détermine diverses caractéristiques ayant ces valeurs de $\hat{\rho}_1$ et pour celles-ci on évalue à l'aide de l'enquête les valeurs de $\hat{\rho}_2 \ldots, \hat{\rho}_{\theta-1}$. Pour chacune des treize caractéristiques choisies et pour chaque valeur de $i$, on calcule cinq estimations des $\rho_i$, $i>1$. On fait l'hypothèse que deux caractéristiques qui ont la même valeur de $\hat{\rho}_1$, se comportent de façon similaire dans le temps. Ce qui signifie que si on s'intéresse à une moyenne annuelle, on doit déterminer $\hat{\rho}_2, \ldots, \hat{\rho}_{11}$.

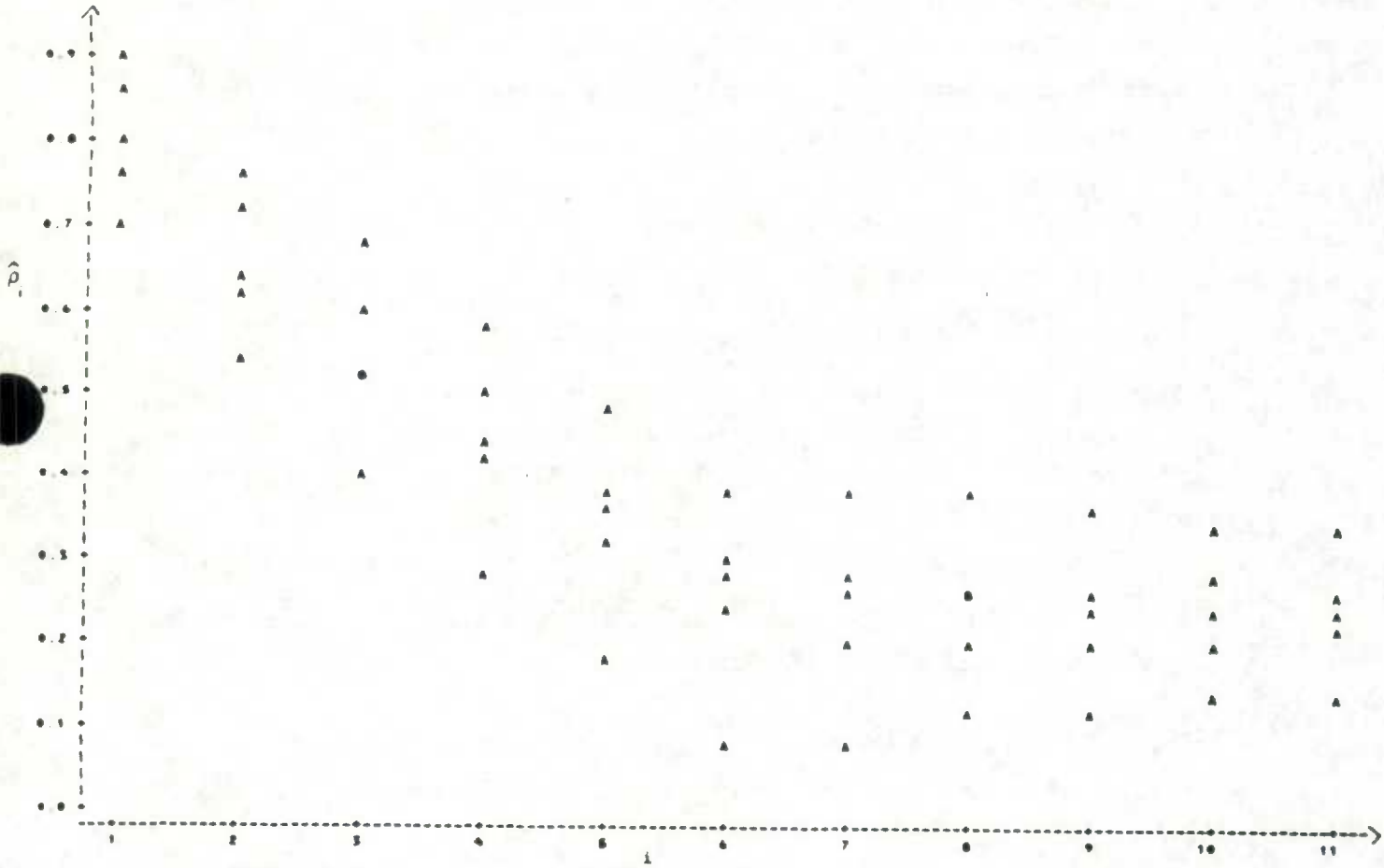
Une analyse visuelle des $\hat{\rho}_1, \ldots, \hat{\rho}_{11}$ permet de constater qu'il est préférable de modéliser selon deux cas. En effet, lorsque $\hat{\rho}_1 \leq 0,65$, les valeurs de $\hat{\rho}_i$ décroient plus rapidement afin de se localiser dans un couloir autour de l'axe $\hat{\rho}_i=0$; tandis que lorsque $\hat{\rho}_1>0,65$ la décroissance est plus lente. De plus, on doit apporter une attention particulière à la valeur de $i$. En effet lorsque $i<6$, c'est-à-dire lorsque les échantillons ne sont pas mutuellement exclusifs puisqu'un même ménage demeure six mois dans l'échantillon, on observe un comportement différent de lorsque $i \geq 6$. Les graphiques 1 et 2 démontrent ces constatations où on retrouve en abscisse les valeurs de $i=1, \ldots, 11$ et en ordonnée la valeur moyenne des $\hat{\rho}_i$ pour chaque valeur initiale $\hat{\rho}_1$. On remarque également d'après ces graphiques que les observations se comportent de façon exponentielle, c'est pourquoi on a envisagé ce type de modèle. On trouve comme modèle pour $i>1$

**GRAPHIQUE 1 :** $\hat{\rho}_i$ estimées selon les valeurs de départ $\hat{\rho}_1 =$ 0,30 - 0,35 - 0,40 - 0,45 - 0,50 - 0,55 - 0,60 - 0,65 en fonction de i=1,...,11 , le nombre de mois d'écart




A: 1 obs., B: 2 obs., etc.

**GRAPHIQUE 2 :** $\hat{\rho}_i$ estimées selon les valeurs de départ $\hat{\rho}_1 = 0,70$ - 0,75 - 0,80 - 0,85 - 0,90 - 0,95 en fonction de i=1,...,11, le nombre de mois d'écart



$$\hat{\rho}_i = \begin{cases} \gamma_1(0{,}640\ \hat{\rho}_1^{0{,}894i} + 0{,}058) - \gamma_2(0{,}518\ \hat{\rho}_1^{0{,}892i} - 0{,}017) & \text{lorsque } \hat{\rho}_1 \leq 0{,}65 \\ \gamma_1(0{,}868\ \hat{\rho}_1^{0{,}757i}) + \gamma_2(0{,}457\ \hat{\rho}_1^{0{,}791i} + 0{,}122) & \text{lorsque } \hat{\rho}_1 > 0{,}65 \end{cases}$$

et $\gamma_1 = \begin{cases} 1 \text{ si } i<6 \\ 0 \text{ sinon} \end{cases}$ , $\gamma_2 = \begin{cases} 1 \text{ si } i \geq 6 \\ \text{sinon} \end{cases}$ .

Les deux modèles retenus, c'est-à-dire lorsque $\hat{\rho}_1 \leq 0{,}65$ et $\hat{\rho}_1 > 0{,}65$ affichent des valeurs de $R^2$ élevées, 0,75 et 0,86 respectivement. De plus, une analyse des résidus permet de déceler une variance constante dans les erreurs; celles-ci étant par ailleurs normales.

Des tableaux sont disponibles pour évaluer le terme

$$[1 - 2 \sum_{i=1}^{\theta-1} \frac{(\theta-1)}{\theta^2} (1 - \rho_i)]^{\frac{1}{2}}$$ (équation (11)) lorsque $\theta$ = 3, 6, 12,

c'est-à-dire moyenne trimestrielle, semestrielle et annuelle, connaissant une estimation de $\rho_1$ qui peut être lue dans une liste de référence des groupes et des coefficients de corrélation des caractéristiques de l'EPA (voir Dufour et Lemaître (1987b)).

## 5.4 Tableau de référence des CV d'une variation de moyennes annuelles au niveau provincial et pour le Canada

Le CV d'une variation de moyennes annuelles peut s'écrire en terme du CV d'une estimation mensuelle de même niveau et du coefficient de corrélation. En effet,

$$\underline{CV} = CV\left( \sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta} - \sum_{i=\theta+1}^{2\theta} \frac{\hat{X}_i}{\theta} \right) = \left[ Var \left( \sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta} - \sum_{i=\theta+1}^{2\theta} \frac{\hat{X}_i}{\theta} \right) \right]^{1/2} \bullet$$

$$\left[ \sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta} - \sum_{i=\theta+1}^{2\theta} \frac{\hat{X}_i}{\theta} \right]^{-1}$$

$$= \frac{\left[\sum_{i=1}^{2\theta} Var(\hat{X}_i) + 2 \sum_{i<j\leq\theta} Cov(\hat{X}_i,\hat{X}_j) + 2 \sum_{\theta<i<j\leq 2\theta} Cov(\hat{X}_i,\hat{X}_j) - 2 \sum_{\substack{1\leq i\leq\theta \\ \theta<j\leq 2\theta}} Cov(\hat{X}_i, \hat{X}_j)\right]^{\frac{1}{2}}}{\left(\sum_{i=1}^{\theta} \hat{X}_i - \sum_{i=\theta+1}^{2\theta} \hat{X}_i\right)}.$$

La première hypothèse postulée consiste à considérer $CV(\hat{X}_i) = C$, $i=1, \ldots, 2\theta$, et par le fait même $Cov(\hat{X}_i, \hat{X}_j) = \rho_{ij} C^2 \hat{X}_i \hat{X}_j$. On obtient alors

$$\underline{CV} = \frac{\left[\sum_{i=1}^{2\theta} C^2 \hat{X}_i^{\,2} + 2 \sum_{i<j\leq\theta} C^2\rho_{ij}\hat{X}_i\hat{X}_j + 2 \sum_{\theta<i<j\leq 2\theta} C^2\rho_{ij} \hat{X}_i\hat{X}_j - 2 \sum_{\substack{1\leq i\leq\theta \\ \theta<j\leq 2\theta}} C^2\rho_{ij}\hat{X}_i\hat{X}_j\right]^{\frac{1}{2}}}{\left(\sum_{i=1}^{\theta} \hat{X}_i - \sum_{i=\theta+1}^{2\theta} \hat{X}_i\right)}$$

$$= C\left[1 + \frac{\sum_{\substack{1\leq i\leq\theta \\ \theta<j\leq 2\theta}} 2(1-\rho_{ij})\hat{X}_i\hat{X}_j - 2 \sum_{i<j\leq\theta} (1-\rho_{ij})\hat{X}_i\hat{X}_j - 2 \sum_{\theta<i<j\leq 2\theta} (1-\rho_{ij})\hat{X}_i\hat{X}_j}{\left(\sum_{i=1}^{\theta} \hat{X}_i - \sum_{i=\theta+1}^{2\theta} \hat{X}_i\right)^2}\right]^{\frac{1}{2}}.$$

Comme deuxième hypothèse, on suppose qu'il est possible d'estimer $X_1, \ldots, X_\theta$ par une même quantité soit $\bar{X}$ et $X_{\theta+1}, \ldots, X_{2\theta}$ par $\bar{Y}$. On a donc

$$\underline{CV} = C\left[1 + \frac{2 \sum_{\substack{1\leq i\leq\theta \\ \theta<j\leq 2\theta}} (1-\rho_{ij})\bar{X}\bar{Y} - 2 \sum_{\theta<i<j\leq 2\theta} (1-\rho_{ij})\bar{Y}^2 - 2 \sum_{i<j\leq\theta} (1-\rho_{ij})\bar{X}^2}{\theta^2(\bar{X}-\bar{Y})^2}\right]^{\frac{1}{2}}.$$

Comme troisième hypothèse, on pose $\rho_{ij} = \rho_{kl}$ lorsque $j-i = l-k$ et on note $\rho_i = CORR(X_t, X_{t+i})$. De cette dernière hypothèse et par quelques manipulations algébriques, on peut écrire

$$\underline{CV} = C \cdot \left[1-2 \sum_{i=1}^{\theta-1} \frac{(\theta-i)(1-\rho_i)}{\theta^2} + \left[\frac{2\,\bar{X}\bar{Y}}{\theta^2(\bar{X}-\bar{Y})^2}\left(\sum_{i=1}^{\theta-1} (3i-2\theta)(1-\rho_i) + \sum_{i=1}^{\theta} i(1-\rho_{2\theta-i})\right)\right]\right]^{\frac{1}{2}}.$$

Si on pose $r = \frac{\tilde{Y}-\tilde{X}}{\tilde{X}}$, l'écart relatif entre $\tilde{X}$ et $\tilde{Y}$ et $\theta=12$ puisqu'on s'intéresse à une variation de moyennes annuelles, la formule devient

$$\underline{CV} = C \cdot \left[ 1 - \sum_{i=1}^{11} \frac{2(12-i)(1-\rho_i)}{144} + \left[ \sum_{i=1}^{12} \frac{2i(1-\rho_{24-i})}{144} + \sum_{i=1}^{11} \frac{2(3i-24)(1-\rho_i)}{144} \right] \left(\frac{1}{r^2} + \frac{1}{r}\right) \right]^{\frac{1}{2}} .$$

Afin d'évaluer cette équation, on doit connaître une estimation de $\rho_1, \ldots, \rho_{23}$. Pour ce faire, on utilise le modèle explicité à la section 5.3. Ainsi connaissant une estimation de $\rho_1$, disponible dans la liste de référence des groupes et des coefficients de corrélation des caractéristiques de l'EPA, et l'écart relatif entre les deux estimations annuelles il est possible de lire dans le tableau approprié la valeur du terme entre crochets (voir Dufour et Lemaître (1987b)).

Le CV de la variation de moyennes annuelles étant déterminé, on peut se référer au tableau du niveau de confiance associé au changement afin d'établir si ce changement est significatif ou non.

## 5.5 Tableaux de référence des CV d'une estimation mensuelle, d'une moyenne de niveaux, d'une variation mensuelle, ou d'une variation de moyennes annuelles au niveau infraprovincial

Il n'est pas nécessaire de construire tous ces tableaux de référence puisque tous ces types de coefficients de variation au niveau infraprovincial peuvent s'exprimer en fonction de leur vis-à-vis au niveau provincial et d'un facteur de correction formulant certaines hypothèses. Par exemple, pour déterminer le CV d'une estimation mensuelle au niveau infraprovincial, il s'agit tout simplement d'établir une correspondance entre le CV au niveau provincial et celui au niveau infraprovincial. Soit $\hat{X}$ une estimation d'une caractéristique $X$ quelconque. Au niveau provincial, on peut écrire

$$CV(\hat{X}) \simeq [F(\frac{p}{n} - 1)]^{\frac{1}{2}} \hat{X}^{-\frac{1}{2}} \qquad (11)$$

où $F$ : effet de plan de sondage au niveau provincial

$n$ : taille de l'échantillon dans la province

$P$ : population de la province

$CV(\hat{X})$ : coefficient de variation au niveau de la province.

Pour cette même caractéristique $X$ quelconque, on a également

$$CV'(\hat{X}') \simeq [F'(\frac{P'}{n'} - 1)]^{\frac{1}{2}} \hat{X}'^{-\frac{1}{2}} \tag{12}$$

où $CV'(\hat{X}')$, $P'$, $n'$, $F'$ ont les mêmes définitions que $CV(\hat{X})$ $F$, $P$ et $n$ à l'exception que " ' " signifie au niveau infraprovincial plutôt que provincial. Le rapport des équations (12) et (13) pour une même valeur de $\hat{X}$ et $\hat{X}'$, soit $\bar{X}$, donne

$$\frac{CV'(\bar{X}')}{CV(\bar{X})} \simeq \left[\frac{(\frac{P'}{n'} - 1)}{(\frac{P}{n} - 1)} (\frac{F'}{F})\right]^{\frac{1}{2}}.$$

On pose $fc = \left[\frac{(\frac{P'}{n'})}{(\frac{P}{n})} (\frac{F'}{F})\right]^{\frac{1}{2}}$ (puisque $\frac{P'}{n'} >> 1$ et $\frac{P}{n} >> 1$) et par le fait même $CV'(\bar{X}') \simeq f_c \, CV(\bar{X})$.

Il est à noter que $F' \neq F$ étant donné l'effet de grappe. En pratique $\hat{X}'$ diffère de $\hat{X}$ toutefois le facteur de correction $f_c$ reste valide pour faire l'approximation de ce CV. En fait, il s'agit de choisir des caractéristiques de mêmes valeurs et du même groupe.

Il est maintenant facile de déterminer un coefficient de variation pour une estimation mensuelle au niveau infraprovincial. Des tableaux donnant la valeur de $f_c$ sont disponibles selon qu'il est question d'une région métropolitaine de recensement ou d'une région économique (voir Dufour et Lemaître (1987b)).

Le CV d'une variation mensuelle au niveau provincial, pour une caractéristique quelconque $X$, peut s'écrire

$$CV(\hat{X}_2 - \hat{X}_1) = CV(\hat{X})[2(1-\rho)(\frac{1}{r} + \frac{1}{r^2})]^{\frac{1}{2}} \tag{13}$$

où

$CV(\hat{X})$: CV de l'estimation mensuelle au niveau provincial

$\rho$ : coefficient de corrélation au niveau provincial

$r$ : écart relatif entre les estimations $\hat{X}_1$ et $\hat{X}_2$.

Cette même quantité au niveau infraprovincial devient

$$CV'(\hat{X}_2'-\hat{X}_1') = CV'(\hat{X}')[2(1-\rho')(\frac{1}{r} + \frac{1}{r^2})]^{\frac{1}{2}} \tag{14}$$

où le symbole " ' " signifie infraprovincial. Le rapport des équations (14) et (13) pour une même valeur de $\hat{X}$ et $\hat{X}'$ permet d'écrire

$$\frac{CV'(\hat{X}_2'-\hat{X}_1')}{CV(\hat{X}_2-\hat{X}_1)} = \left[ \frac{(\frac{P'}{n'} - 1)}{(\frac{P}{n} - 1)} (\frac{F'}{F}) \right]^{\frac{1}{2}} (\frac{1-\rho'}{1-\rho})^{\frac{1}{2}}$$

$$= f_c(\frac{1-\rho'}{1-\rho})^{\frac{1}{2}}.$$

On fait l'hypothèse que le coefficient de corrélation au niveau infraprovincial peut être estimé par celui correspondant au niveau provincial, et on obtient

$$CV'(\hat{X}_2'-\hat{X}_1') = CV(\hat{X}_2-\hat{X}_1)\ f_c.$$

En ce qui concerne le CV d'une moyenne trimestrielle, semestrielle ou annuelle au niveau infraprovincial, on a

$$CV'(\sum_{i=1}^{\theta} \frac{\hat{X}_i'}{\theta}) = CV'(\hat{X}')[1 - 2 \sum_{i=1}^{\theta-1} \frac{(\theta-i)}{\theta^2} (1-\rho_i')]^{\frac{1}{2}}$$

$$\simeq f_c\ CV(\hat{X})\ [1 - 2 \sum_{i=1}^{\theta-1} \frac{(\theta-i)}{\theta^2} (1-\rho_i)]^{\frac{1}{2}}$$

$$= f_c\ CV(\sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta}).$$

De même, pour le CV d'une variation de moyennes annuelles au niveau infraprovincial,

$$CV'(\sum_{i=1}^{\theta} \frac{\hat{X}_i'}{\theta} - \sum_{i=\theta+1}^{2\theta} \frac{\hat{X}_i'}{\theta}) = CV'(\hat{X}')\{1 - \sum_{i=1}^{11} \frac{2(12-i)(1-\rho_i')}{144} + [\sum_{i=1}^{12} \frac{2i(1-\rho_{24-i}')}{144}$$

$$+ \sum_{i=1}^{11} \frac{2(3i-24)(1-\rho_i')}{144}] (\frac{1}{r} + \frac{1}{r^2})\}^{\frac{1}{2}}$$

$$= f_c \, CV(\hat{X})\{1 - \sum_{i=1}^{11} \frac{2(12-i)(1-\rho_i)}{144} + [(\sum_{i=1}^{12} \frac{2i(1-\rho_{24-i})}{144}$$

$$+ \sum_{i=1}^{11} \frac{2(3i-24)(1-\rho_i)}{144}) \, (\frac{1}{r} + \frac{1}{r^2})\}^{\frac{1}{2}}$$

$$= f_c \, CV(\sum_{i=1}^{\theta} \frac{\hat{X}_i}{\theta} - \sum_{i=\theta+1}^{2\theta} \frac{\hat{X}_i}{\theta}).$$

## 6. Exemple d'utilisation des tableaux de référence des coefficients de variation

Dans l'exemple qui suit, on désire déterminer le coefficient de variation associé au niveau d'emploi pour le Canada pour les 15 ans et plus et ce pour un mois quelconque. Dans un premier effort, on doit connaître à quel groupe appartient la caractéristique à l'étude. On retrouve à l'Annexe B un extrait de la liste des groupes associés à certaines caractéristiques de l'EPA permettant de déterminer le groupe approprié. On cherche dans cette liste de référence des groupes la caractéristique emploi pour "Tous" et on lit le groupe correspondant, soit le numéro 3.

Supposons maintenant que l'estimation du niveau d'emploi qui nous intéresse soit de 11,641,000. Connaissant le numéro de groupe (3) et une estimation de la caractéristique (11,641,000), on se réfère au tableau de référence des coefficients de variation au Canada (voir le Tableau I à l'Annexe B) afin de lire le coefficient de variation du niveau. Dans l'exemple, il correspond à 0,3%.

Il est important de se souvenir, lors de l'utilisation des tableaux de référence, que les CV obtenus sont des approximations et non pas des valeurs exactes. En présence d'une situation où la valeur ne peut être lue directement dans un tableau, il est prudent d'être conservateur c'est-à-dire choisir le plus grand CV des deux valeurs possibles.

## 7. Discussion sur la qualité des tableaux de référence des coefficients de variation

La qualité des tableaux de référence des coefficients de variation construits à l'aide de modèles linéaires logarithmiques repose essentiellement sur deux facteurs: le regroupement des caractéristiques et le modèle lui-même.

En ce qui concerne la formation des groupes, on se souviendra qu'ils ont été déterminés en fonction de l'effet de plan de sondage et qu'étant donné la réduction de la taille de l'échantillon au mois de juin 1986, seules les données des mois de mars 1985 à mai 1986 ont été utilisées. Puisque la répartition de l'échantillon n'est pas uniforme à l'intérieur des strates pour chacune des provinces alors il est probable que la réduction de la taille de l'échantillon affecte l'effet de plan de sondage. Afin d'évaluer la stabilité des groupes, c'est-à-dire vérifier si l'effet de plan de sondage moyen obtenu à l'aide des données de juin 1986 à décembre 1987 appartient au même groupe que son équivalent basé sur les données de mars 1985 à mai 1986, on s'est limité à sept caractéristiques. Il s'agit des caractéristiques: population active, nombre d'employés, nombre de chômeurs, nombre d'employés dans le secteur agricole, nombre d'employés dans le secteur non-agricole, emploi à temps plein et à temps partiel pour les dix provinces et pour le groupe 15 ans et plus.

Le Tableau V de l'Annexe A présente les résultats obtenus. On remarque que sur les 70 caractéristiques (7 caractéristiques x 10 provinces), 20 caractéristiques ont changé de groupes. Toutes les provinces et toutes les caractéristiques ont subi au moins un changement. Tous les changements enregistrés correspondent à un saut d'une seule classe. Le saut d'une classe n'est toutefois pas trop alarmant, les différences entre les estimations des CV d'une classe à l'autre n'étant pas si énormes. De plus, 6 des 20 changements peuvent s'expliquer facilement puisque la moyenne des effets de plan de sondage basée sur les données de mars 1985 à mai 1986 pour ces 6 caractéristiques correspondaient exactement à la borne du groupe en question. Des 14 changements restants, 3 se sont effectués en direction d'un groupe inférieur ce qui signifie qu'en utilisant une classe supérieure dans l'estimation du CV on adopte une attitude conservatrice. Il reste donc 11 changements pour les 70 caractéristiques étudiées auxquels on ne peut associer une explication évidente. Il faudra donc faire preuve de prudence

lors de l'utilisation des tableaux de référence et se rappeler que ces tableaux donnent une estimation du coefficient de variation. On devrait également envisager de mettre à jour la répartition des caractéristiques selon les différents groupes étant donné la disponibilité de plusieurs mois de données depuis la réduction de la taille de l'échantillon.

Afin d'éprouver la qualité des estimations des coefficients de variation déterminées à l'aide d'un modèle linéaire logarithmique, on s'est limité à une seule catégorie, soit population active, pour le Canada et selon 15 groupes d'âge (tous, 15-24, 15-19, 20-24, 25+, 25-34, 35-44, 45-54, 55-64, 25-44, 45+, 45-64, 65+, 25-54, 55+). A l'aide des données de l'année 1986 et de 1987, on a estimé séparément les coefficients de variation à l'aide du modèle et on a déterminé pour chacune des périodes l'erreur quadratique moyenne associée à chacune des 15 caractéristiques précitées soit

$$EQM = E[(\hat{C}V(\hat{X}) - CV(\hat{X}))^2] .$$

Les valeurs des EQM trouvées sont présentées au Tableau VI de l'Annexe A. Une comparaison des EQM pour chacune des caractéristiques basées sur 1986 et 1987 permet de constater que l'EQM basée sur les données de 1987 est supérieure à l'EQM obtenue à l'aide des données de 1986 dans seulement trois occasions. Puisque le modèle a été déterminé à l'aide des données de mars 1985 à décembre 1986, et que les estimations obtenues pour l'année 1987 semblent précises, il n'y a pas d'objection évidente au fait de continuer à utiliser un tel modèle.

La présentation des erreurs d'échantillonnage dans les publications de l'EPA se fait sous forme d'indicateurs alphabétiques pour la majorité des caractéristiques. Ces indicateurs correspondent aux intervalles des coefficients de variation en pourcentage; par exemple A (0,0 - 0,5%), B (0,6 - 1%), C (1,1 - 2,5%), D (2,6 - 5%), etc. Une association des symboles littérals aux 15 caractéristiques étudiées pour l'année 1987 pour les estimations et les vraies valeurs des coefficients de variation démontre que les indicateurs alphabétiques diffèrent très rarement.

Suite à ces constatations face à la qualité des estimations obtenues à l'aide d'un modèle linéaire logarithmique, il paraît une sage décision de continuer à

utiliser un tel modèle. Il faudra éventuellement songer à tester à nouveau la qualité du modèle, par exemple sur les données de 1988, et ce sur une plus grande échelle et prendre une décision à savoir si on doit estimer à nouveau les paramètres du modèle ou tout simplement envisager un autre type de modèle FVG.

## 8. Conclusion

Il existe diverses fonctions de variances généralisées (FVG) permettant d'obtenir une approximation de la variance d'une estimation quelconque comme fonction simple de l'estimation. Pour cette étude, on s'est toutefois limité à des modèles linéaires logarithmiques. Deux types de modèles ont été analysés: modèle linéaire logarithmique et modèle linéaire logarithmique autorégressif. Le deuxième type de modèle n'a cependant pas donné les résultats escomptés. Les résultats obtenus nous ont amené à omettre de l'étude ce genre de modèle. Pour ce qui est du type de modèle linéaire logarithmique, deux modèles ont été envisagés. Tous deux affichent des résultats assez intéressants. Notre choix s'est toutefois arrêté sur le plus simple des deux modèles, soit celui qui n'exige aucune information auxiliaire telle la taille de la population.

La détermination d'un modèle FVG permet de construire différents tableaux de référence des coefficients de variation de caractéristiques quelconques utiles aux utilisateurs des données de l'EPA. Il est également possible de dériver une gamme de tableaux de référence tels d'une variation ou d'une moyenne à l'aide du CV d'un niveau. Etant donné le nombre élevé de caractéristiques et la diversité des demandes associées aux coefficients de variation, il n'a pas été possible de travailler avec chacune des caractéristiques à l'étude. On a donc envisager de regrouper les différentes caractéristiques en plusieurs groupes tout en respectant l'homogénéité des effects de plan de sondage à l'intérieur de chacun d'eux.

La qualité des tableaux de référence repose donc sur ces deux facteurs: efficacité du modèle et qualité de la méthode de regroupement. Une analyse de ces deux aspects ne nous a toutefois pas permis de déceler de façon évidente des anomalies au niveau de ces deux facteurs. Il faudra cependant être prudent lors de l'utilisation des tableaux de référence puisque ces derniers donnent une estimation du coefficient de variation et non la valeur réelle. Il serait également

sage de mettre à jour la répartition des caractéristiques entre les différents groupes selon la disponibilité d'autres mois de données et de vérifier si le modèle colle toujours à la réalité. Il serait par ailleurs intéressant, dans un avenir approché, d'effectuer une étude comparative entre les modèles suggérés par d'autres auteurs tels Kalton (1977), Hansen, Hurwitz et Madow (1953), Hanson (1978) et Wolter (1985) et celui utilisé lors de la construction des tableaux de référence soutenu dans ce rapport.

## Remerciements

Les auteurs tiennent à remercier J. Armstrong pour les précieux commentaires apportés lors du déroulement et de la rédaction de cette étude, ainsi que P. Lavallée et J. Gambino dont les commentaires ont permis d'améliorer le contenu de cet article. Ils remercient également P. Pariseau et J. Clarke pour leur excellent travail de dactylographie.

## Bibliographie

Choudhry, G.H., et Lee, H. (1987), "Estimation de la variance pour l'enquête sur la population active", Techniques d'enquête, Vol: 13, No: 2, 1987, pages: 157-173.

Cochran, W.G. (1977). Sampling techniques (3e édition). New York: John Wiley and Sons.

Dufour, J., et Lemaître, G. (1987a). "L'utilisation des tableaux de référence des coefficients de variation se rapportant aux caractéristiques de l'EPA", Statistique Canada, document non-publié.

Dufour, J., et Lemaître, G. (1987b). "Tableaux de référence des coefficients de variation se rapportant aux caractéristiques de l'Enquête sur la population active", Statistique Canada, document non-publié.

Ghangurde, P.D. (1981), "Modèles des erreurs d'échantillonnage", Techniques d'enquête, Vol: 7, No: 2, 1981, pages: 187-203.

Hansen, M.H., Hurwitz, W.N., et Madow, W.G. (1953). Sample Survey Methods and Theory, 2 Volumes. New York: John Wiley and Sons.

Hanson, R.H. (1978). The Current Population Survey — Design and Methodology, Technical Paper 40, U.S. Bureau of the Census, Washington, D.C. 20233.

Kalton, G. (1977). "Practical methods for Estimating Survey Sampling Errors", Bulletin of the International Statistical Institute, 47, pages: 495-514.

Keyfitz, N. (1957). "Estimates of sampling variance where two units are selected from each stratum", Journal of the American Statistical Association, 52, 1957, pages: 503-510.

Kish, L. (1965). Survey Sampling. New York: John Wiley and Sons.

Neter, J., Wasserman, W. et Kutner, M.H. (1985). Applied Linear Statistical Models, Irwin, Illinois, 2e édition.

Platek, R., et Singh, M.P. (1976). Méthodologie de l'Enquête sur la population active, no: 71-526 au catalogue, hors-série.

Valliant, R. (1987). "Generalized Variance Functions in Stratified Two-Stage Sampling", Journal of the American Statistical Association, Juin 1987, Vol: 82, No: 398, pages: 499-509.

Wolter, K.M. (1985). Introduction to Variance Estimation, New York: Springer-Verlag.

Annexe A

## TABLEAU I

### Valeurs Limites des Effets de Plan de Sondage Pour les Groupes I, II, III, IV

| PROVINCES | GROUPE I | GROUPE II | GROUPE III | GROUPE IV |
|---|---|---|---|---|
| Terre-Neuve | ≤0,9960 | (0,9960 - 1,5906] | (1,5906 - 2,0633] | >2,0633 |
| Ile Prince-Édouard | ≤0,5410 | (0,5410 - 0,7571] | (0,7571 - 1,2278] | >1,2278 |
| Nouvelle-Écosse | ≤0,6725 | (0,6725 - 1,0907] | (1,0907 - 1,4465] | >1,4465 |
| Nouveau-Brunswick | ≤0,7762 | (0,7762 - 1,2492] | (1,2492 - 1,8547] | >1,8547 |
| Québec | ≤0,7991 | (0,7991 - 1,3788] | (1,3788 - 1,7282] | >1,7282 |
| Ontario | <0,4043 | (0,4043 - 0,6846] | (0,6846 - 1,2575] | >1,2575 |
| Manitoba | ≤0,4576 | (0,4576 - 0,7499] | (0,7499 - 1,2085] | >1,2085 |
| Saskatchewan | ≤0,5759 | (0,5759 - 1,1306] | (1,1306 - 1,5887] | >1,5887 |
| Alberta | ≤0,6702 | (0,6702 - 1,0837] | (1,0837 - 1,6020] | >1,6020 |
| Colombie-Britannique | ≤0,5205 | (0,5205 - 1,0385] | (1,0385 - 1,2706] | >1,2706 |
| CANADA | ≤0,6484 | (0,6484 - 1,0586] | (1,0586 - 1,6963] | >1,6963 |

## TABLEAU II

### Nombre de Caractéristiques dans les Groupes I, II, III et IV

| PROVINCES | GROUPE I | GROUPE II | GROUPE III | GROUPE IV |
|---|---|---|---|---|
| Terre-Neuve | 20 | 27 | 12 | 6 |
| Ile Prince-Édouard | 16 | 9 | 21 | 19 |
| Nouvelle-Écosse | 21 | 18 | 19 | 7 |
| Nouveau-Brunswick | 20 | 18 | 18 | 9 |
| Québec | 23 | 24 | 14 | 4 |
| Ontario | 13 | 13 | 24 | 15 |
| Manitoba | 15 | 11 | 13 | 26 |
| Saskatchewan | 17 | 16 | 20 | 12 |
| Alberta | 19 | 9 | 19 | 18 |
| Colombie-Britannique | 14 | 15 | 15 | 21 |
| CANADA | 12 | 13 | 17 | 23 |

## TABLEAU III

Coefficients de Régression et $R^2$ Associés au Modèle Logarithmique

$$\ln \widehat{CV}(\hat{X}) = (a\,\delta_1 + a'\delta_2) + b \ln \left( \frac{\hat{X}}{1 - \hat{X}/p} \right) + \varepsilon$$

| PROVINCES | $a_1$ | $a_1'$ | $a_2$ | $a_2'$ | $a_3$ | $a_3'$ | $a_4$ | $a_4'$ | b | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Terre-Neuve | 6,45 | 6,47 | 6,77 | 6,82 | 6,96 | 6,99 | 7,22 | 7,30 | -0,50 | 0,96 |
| Ile Prince-Édouard | 6,05 | 5,99 | 6,31 | 6,20 | 6,53 | 6,51 | 6,73 | 6,71 | -0,52 | 0,97 |
| Nouvelle-Écosse | 6,48 | 6,56 | 6,84 | 6,95 | 6,97 | 7,09 | 7,19 | 7,41 | -0,51 | 0,97 |
| Nouveau-Brunswick | 6,42 | 6,46 | 6,69 | 6,77 | 6,85 | 6,89 | 7,06 | 7,00 | -0,50 | 0,97 |
| Québec | 7,22 | 7,36 | 7,62 | 7,76 | 7,73 | 7,85 | 8,07 | 8,07 | -0,50 | 0,98 |
| Ontario | 7,11 | 7,16 | 7,43 | 7,52 | 7,76 | 7,87 | 7,96 | 8,05 | -0,52 | 0,99 |
| Manitoba | 6,30 | 6,39 | 6,60 | 6,69 | 6,92 | 7,04 | 7,05 | 7,13 | -0,50 | 0,99 |
| Saskatchewan | 6,57 | 6,63 | 6,99 | 7,08 | 7,17 | 7,23 | 7,55 | 7,56 | -0,53 | 0,96 |
| Alberta | 6,68 | 6,77 | 6,98 | 7,10 | 7,24 | 7,37 | 7,44 | 7,54 | -0,51 | 0,97 |
| Colombie-Britannique | 6,77 | 6,86 | 7,11 | 7,24 | 7,36 | 7,50 | 7,48 | 7,60 | -0,50 | 0,98 |
| CANADA | 7,10 | 7,18 | 7,38 | 7,49 | 7,66 | 7,78 | 7,82 | 7,92 | -0,52 | 0,99 |

## TABLEAU IV

Coefficients de Régression et $R^2$ Associés au Modèle Logarithmique

$$\ln \widehat{CV}\ (\hat{X}) = (a\,\delta_1 + a'\delta_2) + b \ln\ (\hat{X}) + \varepsilon$$

| PROVINCES | $a_1$ | $a_1'$ | $a_2$ | $a_2'$ | $a_3$ | $a_3'$ | $a_4$ | $a_4'$ | b | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Terre-Neuve | 6,96 | 6,97 | 7,24 | 7,29 | 7,42 | 7,44 | 7,73 | 7,82 | -0,56 | 0,96 |
| Ile Prince-Édouard | 6,83 | 6,78 | 7,07 | 6,96 | 7,19 | 7,17 | 7,48 | 7,46 | -0,62 | 0,96 |
| Nouvelle-Écosse | 7,21 | 7,29 | 7,47 | 7,57 | 7,69 | 7,80 | 7,91 | 8,13 | -0,58 | 0,97 |
| Nouveau-Brunswick | 7,10 | 7,14 | 7,31 | 7,39 | 7,46 | 7,50 | 7,69 | 7,63 | -0,57 | 0,96 |
| Québec | 8,23 | 8,37 | 8,55 | 8,68 | 8,71 | 8,83 | 9,02 | 9,03 | -0,59 | 0,98 |
| Ontario | 8,36 | 8,42 | 8,67 | 8,76 | 8,87 | 8,97 | 9,18 | 9,27 | -0,62 | 0,98 |
| Manitoba | 7,20 | 7,29 | 7,46 | 7,55 | 7,66 | 7,77 | 7,90 | 7,97 | -0,58 | 0,98 |
| Saskatchewan | 7,32 | 7,37 | 7,67 | 7,75 | 7,89 | 7,95 | 8,22 | 8,22 | -0,60 | 0,95 |
| Alberta | 7,89 | 7,99 | 8,13 | 8,25 | 8,32 | 8,46 | 8,60 | 8,71 | -0,62 | 0,96 |
| Colombie-Britannique | 7,78 | 7,87 | 8,06 | 8,19 | 8,28 | 8,41 | 8,45 | 8,57 | -0,59 | 0,99 |
| CANADA | 8,50 | 8,59 | 8,76 | 8,87 | 8,89 | 9,00 | 9,17 | 9,28 | -0,62 | 0,99 |

Tableau V. Détermination du groupe d'appartenance selon les données de mars 1985 à mai 1986 et de juin 1986 à décembre 1987 pour différentes caractéristiques de l'EPA, par province et pour les 15 ans et plus

| | | Provinces | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Caractéristique | Période(1) | T.-N. | I.P.E. | N.-E. | N.-B. | Qué. | Ont. | Man. | Sask. | Alta | C.-B. |
| Population active | a | 3 | 3 | 2 | 4 | 2 | 3 | 3 | 4 | 3 | 2 |
| | b | 3 | 3 | 2 | 3* | 2 | 3 | 3 | 3* | 3 | 2 |
| Nombre d'employés | a | 3 | 4 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 3 |
| | b | 3 | 3* | 3* | 3 | 3 | 3 | 3 | 4 | 3* | 4* |
| Nombre de chômeurs | a | 4 | 4 | 3 | 4 | 3 | 3 | 4 | 3 | 4 | 4 |
| | b | 4 | 4 | 4* | 4 | 3 | 4* | 4 | 3 | 4 | 4 |
| Nb employés secteur agricole | a | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| | b | 4* | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Nb employés secteur non-agricole | a | 3 | 4 | 2 | 3 | 2 | 3 | 4 | 4 | 4 | 3 |
| | b | 3 | 4 | 3* | 3 | 2 | 3 | 4 | 4 | 4 | 3 |
| Employés à temps plein | a | 3 | 3 | 2 | 2 | 2 | 3 | 3 | 4 | 3 | 2 |
| | b | 2* | 3 | 3* | 2 | 2 | 3 | 3 | 3* | 3 | 3* |
| Employés à temps partiel | a | 2 | 3 | 3 | 2 | 2 | 3 | 3 | 4 | 3 | 3 |
| | b | 3* | 3 | 3 | 3* | 3* | 3 | 4* | 4 | 4* | 4* |

(1) a : basé sur les données de mars 1985 à mai 1986
b : basé sur les données de juin 1986 à décembre 1987

* : signifie un changement de groupe

Tableau VI. Erreur quadratique moyenne associée au coefficient de variation pour la population active au Canada selon différents groupes d'âge pour l'année 1986 et 1987 ($x10^{-4}$ puisque le CV est donné en %)

| | Période | |
|---|---|---|
| Groupes d'âge | année 1986 | année 1987 |
| Tous | 0,0017 | 0,0014 |
| 15-24 | 0,0104 | 0,0202 |
| 15-19 | 0,0261 | 0,0194 |
| 20-24 | 0,0118 | 0,0091 |
| 25+ | 0,0010 | 0,0006 |
| 25-34 | 0,0052 | 0,0034 |
| 35-44 | 0,0131 | 0,0085 |
| 45-54 | 0,0054 | 0,0027 |
| 55-64 | 0,0149 | 0,0006 |
| 25-44 | 0,0005 | 0,0004 |
| 45+ | 0,0020 | 0,0007 |
| 45-64 | 0,0099 | 0,0114 |
| 65+ | 0,1534 | 0,1242 |
| 25-54 | 0,0001 | 0,0001 |
| 55+ | 0,0132 | 0,0263 |

Annexe B

## Liste de référence des groupes et des coefficients de corrélation pour le Canada

**Canada**

| Caractéristique | Groupe | Coefficient de corrélation |
|---|---|---|
| Population active pour les deux sexes selon les groupes d'âge: | | |
| tous | 3 | 0,74 |
| 15-24 | 1 | 0,61 |
| 15-19 | 2 | 0,59 |
| 20-24 | 1 | 0,62 |
| 25+ | 2 | 0,77 |
| 25-34 | 1 | 0,71 |
| 35-44 | 1 | 0,71 |
| 45-54 | 1 | 0,74 |
| 55-64 | 2 | 0,77 |
| 25-44 | 1 | 0,74 |
| 45+ | 2 | 0,78 |
| 45-64 | 1 | 0,77 |
| 65+ | 3 | 0,79 |
| 25-54 | 1 | 0,75 |
| 55+ | 2 | 0,78 |
| Emploi pour les deux sexes selon les groupes d'âge: | | |
| Tous | 3 | 0,76 |
| 15-24 | 2 | 0,65 |
| 15-19 | 2 | 0,63 |
| 20-24 | 1 | 0,65 |
| 25+ | 3 | 0,79 |
| 25-34 | 1 | 0,74 |
| 35-44 | 1 | 0,74 |
| 45-54 | 1 | 0,76 |
| 55-64 | 2 | 0,78 |
| 25-44 | 2 | 0,76 |
| 45+ | 2 | 0,79 |
| 45-64 | 2 | 0,78 |
| 65+ | 3 | 0,80 |
| 25-54 | 2 | 0,77 |
| 55+ | 2 | 0,79 |

TABLEA

TABLEAU DES C CIENTS DE VARIATION (EPA)

CANADA

| ESTIMATION EN MILLIERS | COEFFICIENT DE VARIATION % GROUPE 1 | GROUPE 2 | GROUPE 3 | GROUPE 4 | ESTIMATION EN MILLIERS | COEFFICIENT DE VARIATION % GROUPE 1 | GROUPE 2 | GROUPE 3 | GROUPE 4 |
|---|---|---|---|---|---|---|---|---|---|
| 2.0 | 48.4 | — | — | — | 350.0 | 2.0 | 2.6 | 3.0 | 3.9 |
| 3.0 | 37.6 | 49.8 | — | — | 400.0 | 1.8 | 2.4 | 2.7 | 3.6 |
| 4.0 | 31.5 | 41.6 | 47.4 | — | 450.0 | 1.7 | 2.2 | 2.5 | 3.4 |
| 5.0 | 27.4 | 36.3 | 41.3 | — | 500.0 | 1.6 | 2.1 | 2.4 | 3.2 |
| 6.0 | 24.5 | 32.4 | 36.9 | 48.9 | 550.0 | 1.5 | 2.0 | 2.2 | 3.0 |
| 7.0 | 22.3 | 29.4 | 33.5 | 44.4 | 600.0 | 1.4 | 1.9 | 2.1 | 2.8 |
| 8.0 | 20.5 | 27.1 | 30.9 | 40.9 | 650.0 | 1.3 | 1.8 | 2.0 | 2.7 |
| 9.0 | 19.1 | 25.2 | 28.7 | 38.0 | 700.0 | 1.3 | 1.7 | 1.9 | 2.6 |
| 10.0 | 17.9 | 23.6 | 26.9 | 35.6 | 750.0 | 1.2 | 1.6 | 1.9 | 2.5 |
| 11.0 | 16.8 | 22.3 | 25.3 | 33.6 | 800.0 | 1.2 | 1.6 | 1.8 | 2.4 |
| 12.0 | 15.9 | 21.1 | 24.0 | 31.8 | 850.0 | 1.1 | 1.5 | 1.7 | 2.3 |
| 13.0 | 15.2 | 20.1 | 22.9 | 30.3 | 900.0 | 1.1 | 1.5 | 1.7 | 2.2 |
| 14.0 | 14.5 | 19.2 | 21.8 | 28.9 | 950.0 | 1.1 | 1.4 | 1.6 | 2.1 |
| 15.0 | 13.9 | 18.4 | 20.9 | 27.7 | 1000.0 | 1.0 | 1.4 | 1.6 | 2.1 |
| 16.0 | 13.3 | 17.6 | 20.1 | 26.6 | 1500.0 | 0.8 | 1.1 | 1.2 | 1.6 |
| 17.0 | 12.9 | 17.0 | 19.4 | 25.6 | 2000.0 | 0.7 | 0.9 | 1.0 | 1.3 |
| 18.0 | 12.4 | 16.4 | 18.7 | 24.8 | 2500.0 | 0.6 | 0.8 | 0.9 | 1.2 |
| 19.0 | 12.0 | 15.9 | 18.1 | 23.9 | 3000.0 | 0.5 | 0.7 | 0.8 | 1.0 |
| 20.0 | 11.6 | 15.4 | 17.5 | 23.2 | 3500.0 | 0.5 | 0.6 | 0.7 | 0.9 |
| 21.0 | 11.3 | 14.9 | 17.0 | 22.5 | 4000.0 | 0.4 | 0.6 | 0.7 | 0.9 |
| 22.0 | 11.0 | 14.5 | 16.5 | 21.9 | 4500.0 | 0.4 | 0.5 | 0.6 | 0.8 |
| 23.0 | 10.7 | 14.1 | 16.1 | 21.3 | 5000.0 | 0.4 | 0.5 | 0.6 | 0.8 |
| 24.0 | 10.4 | 13.7 | 15.6 | 20.7 | 5500.0 | 0.4 | 0.5 | 0.5 | 0.7 |
| 25.0 | 10.1 | 13.4 | 15.2 | 20.2 | 6000.0 | 0.3 | 0.4 | 0.5 | 0.7 |
| 30.0 | 9.0 | 12.0 | 13.6 | 18.0 | 6500.0 | 0.3 | 0.4 | 0.5 | 0.6 |
| 35.0 | 8.2 | 10.9 | 12.4 | 16.4 | 7000.0 | 0.3 | 0.4 | 0.5 | 0.6 |
| 40.0 | 7.6 | 10.0 | 11.4 | 15.1 | 7500.0 | 0.3 | 0.4 | 0.4 | 0.6 |
| 45.0 | 7.0 | 9.3 | 10.6 | 14.0 | 8000.0 | 0.3 | 0.4 | 0.4 | 0.6 |
| 50.0 | 6.6 | 8.7 | 9.9 | 13.1 | 8500.0 | 0.3 | 0.4 | 0.4 | 0.5 |
| 55.0 | 6.2 | 8.2 | 9.4 | 12.4 | 9000.0 | 0.3 | 0.3 | 0.4 | 0.5 |
| 60.0 | 5.9 | 7.8 | 8.9 | 11.7 | 9500.0 | 0.3 | 0.3 | 0.4 | 0.5 |
| 65.0 | 5.6 | 7.4 | 8.4 | 11.2 | 10000.0 | 0.2 | 0.3 | 0.4 | 0.5 |
| 70.0 | 5.3 | 7.1 | 8.1 | 10.7 | 10500.0 | 0.2 | 0.3 | 0.4 | 0.5 |
| 75.0 | 5.1 | 6.8 | 7.7 | 10.2 | 11000.0 | 0.2 | 0.3 | 0.4 | 0.5 |
| 80.0 | 4.9 | 6.5 | 7.4 | 9.8 | 11500.0 | 0.2 | 0.3 | 0.3 | 0.5 |
| 85.0 | 4.7 | 6.3 | 7.1 | 9.5 | 12000.0 | 0.2 | 0.3 | 0.3 | 0.4 |
| 90.0 | 4.6 | 6.1 | 6.9 | 9.1 | 12500.0 | 0.2 | 0.3 | 0.3 | 0.4 |
| 95.0 | 4.4 | 5.9 | 6.7 | 8.8 | 13000.0 | 0.2 | 0.3 | 0.3 | 0.4 |
| 100.0 | 4.3 | 5.7 | 6.5 | 8.6 | 13500.0 | 0.2 | 0.3 | 0.3 | 0.4 |
| 110.0 | 4.0 | 5.3 | 6.1 | 8.1 | 14000.0 | 0.2 | 0.3 | 0.3 | 0.4 |
| 120.0 | 3.8 | 5.1 | 5.8 | 7.6 | 14500.0 | 0.2 | 0.3 | 0.3 | 0.4 |
| 130.0 | 3.6 | 4.8 | 5.5 | 7.3 | 15000.0 | 0.2 | 0.3 | 0.3 | 0.4 |
| 140.0 | 3.5 | 4.6 | 5.2 | 6.9 | 15500.0 | 0.2 | 0.2 | 0.3 | 0.4 |
| 150.0 | 3.3 | 4.4 | 5.0 | 6.7 | 16000.0 | 0.2 | 0.2 | 0.3 | 0.4 |
| 160.0 | 3.2 | 4.2 | 4.8 | 6.4 | 16500.0 | 0.2 | 0.2 | 0.3 | 0.4 |
| 170.0 | 3.1 | 4.1 | 4.6 | 6.2 | 17000.0 | 0.2 | 0.2 | 0.3 | 0.4 |
| 180.0 | 3.0 | 3.9 | 4.5 | 5.9 | 17500.0 | 0.2 | 0.2 | 0.3 | 0.3 |
| 190.0 | 2.9 | 3.8 | 4.3 | 5.7 | 18000.0 | 0.2 | 0.2 | 0.3 | 0.3 |
| 200.0 | 2.8 | 3.7 | 4.2 | 5.6 | 18500.0 | 0.2 | 0.2 | 0.3 | 0.3 |
| 225.0 | 2.6 | 3.4 | 3.9 | 5.2 | 19000.0 | 0.2 | 0.2 | 0.3 | 0.3 |
| 250.0 | 2.4 | 3.2 | 3.7 | 4.9 | 19500.0 | 0.2 | 0.2 | 0.2 | 0.3 |
| 300.0 | 2.2 | 2.9 | 3.3 | 4.3 | 20000.0 | 0.2 | 0.2 | 0.2 | 0.3 |

Ca 60S

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA
1010175214