Techniques d'enquête, décembre 1993
Vol. 19, n° 2, pp. 159–174
Statistique Canada

Modélisation conjointe robuste de séries de données sur l'activité pour de petites régions

D. PFEFFERMANN et S.R. BLEUER[1]

RÉSUMÉ

Dans cet article, nous présentons les résultats de l'application d'un modèle d'espace d'états aux taux de chômage canadiens. Le modèle suppose une décomposition additive des valeurs de la population en une tendance, une composante saisonnière et une composante irrégulière, ainsi que des relations autorégressives distinctes pour les six séries d'erreurs de l'enquête correspondant aux six estimateurs de panel mensuels. Le modèle tient compte des effets des groupes de renouvellement et permet que changent, dans le temps, les variances liées au plan qui affectent les erreurs de l'enquête. L'ajustement du modèle est effectué au niveau de petites régions, mais il tient compte de corrélations entre les séries des composantes pour différentes régions. On obtient la robustesse des estimateurs produits par le modèle en imposant, à titre de contrainte, que les estimateurs globaux mensuels fondés sur le modèle visant un groupe de petites régions pour lequel la taille d'échantillon totale est suffisamment grande coïncident avec les estimateurs directs correspondants de l'enquête. La performance du modèle, dans le cas d'une application aux provinces de l'Atlantique, est évaluée au moyen de diverses statistiques de diagnostic et de graphiques de résidus, ainsi que par des comparaisons avec des estimateurs actuellement en usage.

MOTS CLÉS: Variance liée au plan; filtre de Kalman; enquête par panel; biais lié au renouvellement; modèle d'espace d'états.

1. INTRODUCTION

Un modèle de série chronologique appliqué à des données d'enquête est la combinaison de deux modèles distincts: le "modèle du recensement", qui décrit l'évolution dans le temps des valeurs de la population finie, et le modèle des erreurs de l'enquête, qui décrit les relations qui existent dans la série chronologique des erreurs des estimateurs de l'enquête. Il y a au moins quatre raisons principales qui incitent à modéliser les estimateurs bruts de l'enquête:

a) Les estimateurs des valeurs de la population qui résultent du processus de modélisation (estimateurs fondés sur le modèle) ont en général des variances moindres que les estimateurs de l'enquête, notamment dans les petites régions, où les tailles des échantillons sont peu élevées.

b) Le modèle que nous employons permet d'obtenir des estimateurs pour les effets saisonniers, ainsi que pour les variances de ces estimateurs, comme sous-produit du processus d'estimation.

c) Le modèle permet d'établir des prévisions des valeurs de la population, de la tendance et des composantes saisonnières à l'égard de périodes postérieures à la période d'observation de l'échantillon pour laquelle les estimateurs directs de l'enquête sont disponibles. De telles prévisions sont importantes pour l'évaluation de la performance du modèle, de même que pour l'élaboration de politiques.

d) Le modèle peut servir à détecter des points tournants dans le niveau des séries chronologiques et à en évaluer les conséquences. (Les travaux réalisés dans ce domaine seront exposés dans un article distinct.)

La méthodologie décrite dans le présent article intègre les méthodologies présentées dans Pfeffermann et Burck (1990) et Pfeffermann (1991), avec certaines modifications et extensions. Les principales caractéristiques du modèle sont les suivantes:

1) Le modèle décompose les valeurs de la population en trois composantes non observables: tendance, saisonnalité et termes irréguliers. Des prédicteurs lissés de ces composantes (et donc des valeurs de la population) fondés sur l'ensemble des données disponibles, ainsi que les erreurs-types des erreurs de prédiction, sont obtenus directement par l'application du filtre de Kalman. Les erreurs types sont modifiées pour tenir compte de la variation additionnelle causée par l'utilisation de valeurs estimées de paramètres.

2) Le modèle utilise comme données d'entrée les différents estimateurs de panel mensuels. L'utilisation des estimateurs propres à chaque panel offre deux avantages importants par rapport à l'emploi des estimateurs moyens:

[1] D. Pfeffermann, Department of Statistics, Hebrew University, Jerusalem 91905; S.R. Bleuer, Division des méthodes d'enquêtes sociales, Statistique Canada, Ottawa, Ontario, K1A 0T6.

i) elle facilite la détermination du modèle de série chronologique reflétant les erreurs de l'enquête, en permettant l'analyse des contrastes entre les estimateurs des différents panels, et ii) elle donne des estimateurs plus efficaces pour les paramètres du modèle, et donc de meilleurs prédicteurs des composantes non observables du modèle.

3) Le modèle tient compte des changements des variances des erreurs de l'enquête au fil du temps, ainsi que des effets possibles des groupes de renouvellement.

4) Le modèle peut être appliqué simultanément aux estimateurs des panels de petites régions distinctes. Dans ce cas, le modèle du recensement est étendu de manière à tenir compte des corrélations croisées entre les composantes non observables des valeurs de la population qui existent dans ces régions.

5) Une modification visant à assurer la robustesse des estimateurs des petites régions à l'égard d'éventuelles défaillances du modèle est incorporée aux équations du modèle. La modification consiste à contraindre les estimateurs fondés sur le modèle d'agrégats de valeurs de la population pour un groupe de petites régions dont la taille totale de l'échantillon est suffisamment grande à coïncider avec les estimateurs de l'enquête pour les agrégats correspondants. Ainsi, les variations soudaines du niveau de la série sont répercutées sans retard sur les estimateurs fondés sur le modèle.

Le modèle, ainsi que les modifications apportées pour en assurer la robustesse, sont décrits de façon plus détaillée à la section 2. Les résultats empiriques de l'application du modèle aux quatre provinces de l'Atlantique du Canada sont présentés à la section 3. La section 4 présente un bref sommaire ainsi que des suggestions quant à la direction que pourraient emprunter les recherches futures.

Avant de conclure la présente section, mentionnons qu'aux États-Unis, les estimations du chômage sont produites pour la plupart des États à l'aide de modèles de série chronologique ayant une structure semblable à celle du modèle de notre étude. Voir Tiller (1992) pour plus de détails. Une différence importante entre les deux, toutefois, est qu'aux États-Unis, le modèle postulé pour les valeurs de la population inclut également des variables explicatives, de sorte que la tendance et la composante saisonnière ne reflètent que la tendance et les variations saisonnières non prises en compte par les variables explicatives. Les modèles ajustés aux erreurs de l'enquête sont, comme dans notre cas, du type ARMMI, et ils tiennent compte eux aussi des changements des variances des erreurs de l'enquête. Les modèles sont par ailleurs différents en raison des modes très distincts de renouvellement de l'échantillon utilisés dans les deux pays. Une autre différence notable entre les deux modèles vient du fait qu'aux États-Unis, les modèles sont ajustés à chaque État séparément et que les données d'entrée comprennent seulement les estimations moyennes de l'enquête, c'est-à-dire une seule observation chaque mois. Les modèles, par conséquent, ne tiennent pas compte des biais liés aux groupes de renouvellement.

2. UN MODÈLE D'ESPACE D'ÉTATS POUR LA SÉRIE CHRONOLOGIQUE DES DONNÉES SUR LE CHÔMAGE AU CANADA

2.1 L'enquête sur la population active du Canada

Les données canadiennes sur le chômage sont recueillies dans le cadre de l'enquête sur la population active (EPA) réalisée par Statistique Canada. L'EPA canadienne est une enquête par panel mensuelle avec renouvellement, dans laquelle chaque nouveau panel de ménages observé demeure dans l'échantillon pendant six mois consécutifs avant d'être remplacé par un autre panel de la même unité primaire d'échantillonnage (UPÉ) ou de la même strate. Les UPÉ sont délimitées géographiquement (pâtés de maisons ou centres urbains dans les régions urbaines, et groupes de secteurs de dénombrement dans les régions rurales). Les strates sont des groupes homogènes d'UPÉ délimités géographiquement (p. ex. secteurs de recensement, subdivisions de recensement et secteurs de dénombrement). Dans les régions urbaines (environ les deux tiers de l'échantillon), chaque UPÉ est représentée dans un seul panel. Dans les régions rurales, les UPÉ sont représentées dans tous les panels, mais les secteurs de dénombrement sont différents d'un panel à l'autre. Par conséquent, les estimateurs de panel distincts peuvent être présumés indépendants, propriété qui a été validée et utilisée dans d'autres études; voir par exemple Lee (1990). Pour une description récente du plan de l'EPA et de la construction des estimateurs directs de l'enquête, le lecteur est invité à consulter Singh et coll. (1990).

2.2 Le modèle du recensement

Dans les paragraphes qui suivent, nous examinons le cas d'une petite région unique, tandis que dans la section 2.4, nous aborderons la modélisation conjointe des estimations des panels pour un groupe de petites régions. Le modèle postulé pour les valeurs de la population est le modèle structurel de base (MSB), représenté par l'ensemble suivant d'équations:

$$Y_t = L_t + S_t + \epsilon_t; \quad L_t = L_{t-1} + R_{t-1} + \eta_{Lt};$$

$$R_t = R_{t-1} + \eta_{Rt}; \quad \sum_{j=0}^{11} S_{t+j} = \eta_{St}. \qquad (2.1)$$

Dans 2.1, Y_t est la valeur de la population (taux de chômage "vrai") au temps t, L_t est le niveau de la tendance, R_t est l'accroissement, S_t est l'effet saisonnier et ϵ_t est le terme irrégulier qu'on suppose être un bruit blanc avec moyenne zéro et variance σ_ϵ^2. Ainsi, la première équation de (2.1) suppose la décomposition classique d'une série chronologique en trois composantes: tendance, variations saisonnières et terme irrégulier. Une telle décomposition est inhérente aux méthodes courantes de désaisonnalisation des données; voir par exemple Dagum (1980).

Notons toutefois que dans le présent cas, la série $\{Y_t\}$ est elle-même non observable. Les séries $\{\eta_{Lt}\}$, $\{\eta_{Rt}\}$ et $\{\eta_{St}\}$ sont des perturbations de bruit blanc indépendantes de moyennes zéro et de variances σ_L^2, σ_R^2 et $\sigma_S^2 \times g(t)$ respectivement. Par conséquent, la deuxième et la troisième équation de (2.1) définissent une approximation locale d'une tendance linéaire, tandis que la dernière équation décrit l'évolution des effets saisonniers de telle manière que la somme de n'importe quel ensemble de 12 effets successifs fluctue autour de zéro. Notons que les variances des termes d'erreur η_{St} dépendent du temps. Les fonctions $g(t)$ sont définies à la fin de la section 3.1.

Les propriétés théoriques du MSB en comparaison d'autres modèles sont examinées dans Harrison et Stevens (1976), Harvey (1984) et Maravall (1985). Les résultats empiriques illustrant la performance du modèle sont présentés dans Harvey et Todd (1983), Morris et Pfeffermann (1984) et Pfeffermann (1991). Bien qu'il soit plus limité que la famille des modèles ARMMI, le MSB est maintenant considéré comme suffisamment flexible pour donner une approximation du comportement d'une grande variété de séries chronologiques.

2.3 Le modèle des erreurs de l'enquête

Le modèle représentant les erreurs de l'enquête a été déterminé initialement par une analyse distincte des séries de pseudo-erreurs $e_{t,p}^{(j)} = (y_t^{(j)} - \bar{y}_t)$, $t = 1, \ldots, N$, où $y_t^{(j)}$ est l'estimateur de Y_t d'après le j-ième panel, $j = 1, \ldots, 6$, (le panel observé pour le j-ième mois consécutif) et $\bar{y}_t = \sum_{j=1}^{6} y_t^{(j)}/6$ est l'estimateur moyen. Notons que $(y_t^{(j)} - \bar{y}_t) = (e_t^{(j)} - \sum_{j=1}^{6} e_t^{(j)}/6)$, où $e_t^{(j)} = (y_t^{(j)} - Y_t)$ sont les vraies erreurs de l'enquête. Ainsi, la caractéristique notable des contrastes $(y_t^{(j)} - \bar{y}_t)$ est qu'ils sont fonction uniquement des erreurs de l'enquête, peu importe le modèle représentant les valeurs de la population.

Il y a deux facteurs principaux à considérer dans le choix d'un modèle pour les erreurs de l'enquête:

a) Le modèle devrait tenir compte d'éventuels biais liés aux groupes de renouvellement ou, plus généralement, permettre des moyennes différentes pour les erreurs propres à différents panels.

b) Le modèle devrait permettre que les variances des erreurs de l'enquête changent au fil du temps.

Des biais liés aux groupes de renouvellement peuvent survenir lorsque des répondants ne donnent pas les mêmes informations à des interviews différentes; ces effets peuvent dépendre de la durée de participation à l'enquête des répondants, ou de la méthode de collecte des données (p. ex. par téléphone ou par interviews à domicile). (Dans l'EPA canadienne, les membres du premier panel sont interviewés à domicile, tandis que ceux des autres panels sont interviewés par téléphone.) Une autre cause possible d'écarts entre les moyennes des erreurs propres aux panels réside dans les différences entre les profils de non-réponse d'un panel à l'autre. Voir Pfeffermann (1991) pour une analyse plus approfondie du problème et des références concernant les études antérieures effectuées dans ce domaine.

Il survient au fil du temps des changements des variances des erreurs de l'enquête lorsque les variances sont fonction du niveau de la série. En fait, comme le montre la figure 1 de la section 3, les estimations des écarts-types des erreurs de l'enquête sont soumis à des effets saisonniers, selon un profil qui suit l'évolution saisonnière des valeurs de la population. Des changements des variances des erreurs de l'enquête peuvent aussi s'expliquer par des modifications du plan d'échantillonnage. Par exemple, en 1985-1986, la taille totale de l'échantillon de l'EPA canadienne a été ramenée de 55,000 à 48,000 ménages. Cette réduction s'est accompagnée d'autres modifications du plan. Voir Singh et coll. (1990) pour plus de détails.

L'application de méthodes simples d'estimation de modèles et de diagnostic aux pseudo-erreurs de l'enquête suggère un modèle autorégressif (AR) de 3e ordre pour les erreurs de l'enquête standardisées $\tilde{e}_t^{(j)} = (e_t^{(j)} - \beta_j)/SD(e_t^{(j)})$, soit:

$$\tilde{e}_t^{(j)} = \phi_{j1}\, \tilde{e}_{t-1}^{(j-1)} + \phi_{j2}\, \tilde{e}_{t-2}^{(j-2)} + \phi_{j3}\, \tilde{e}_{t-3}^{(j-3)} + u_t^{(j)},\ j = 1, \ldots, 6, \quad (2.2)$$

où les $\beta_j = E(e_t^{(j)})$ sont les biais liés aux groupes de renouvellement, les $SD(e_t^{(j)})$ sont les écarts-types liés au plan et les $u_t^{(j)}$ sont des termes de bruit blanc indépendants de moyenne zéro et de variances σ_j^2. On suppose que $\sum_{j=1}^{6} \beta_j = 0$, d'où il résulte que l'estimateur moyen de l'enquête, $\bar{y}_t$, est non biaisé. Voir Pfeffermann (1991) pour une analyse du besoin d'imposer une contrainte aux coefficients de biais. L'analyse subséquente de l'ajustement du modèle combiné défini par (2.1) et (2.2) (voir la section 2.4) permet de valider ce modèle, avec l'observation additionnelle selon laquelle les coefficients $(\phi_{j1}, \phi_{j2}, \phi_{j3})$ peuvent être supposés égaux pour $j = 4, 5, 6$. En outre, pour le premier panel, un modèle AR(1) donne déjà un bon ajustement, tandis que pour le deuxième et le troisième panel, un modèle AR(2) est approprié, bien que les coefficients soient différents. Ces relations sont valables pour chacune des quatre provinces de l'Atlantique.

L'un des arbitres ayant revu le présent article s'est demandé si le modèle AR(3) défini par (2.2) est suffisamment souple pour tenir compte des corrélations entre les estimations des panels à des décalages importants, qu'on croît être élevées en raison des "effets des UPÉ". Comme il a été mentionné à la section 2.1, les panels qui sortent de l'échantillon sont remplacés par des panels des mêmes UPÉ, et il faut habituellement plusieurs années avant qu'une UPÉ ne soit épuisée et remplacée par une UPÉ voisine. Lee (1990) présente deux ensembles de corrélations entre estimations de panels pour l'EPA canadienne. Le premier ensemble, dénoté par ρ_j, comprend les corrélations entre les estimations produites par le même panel, de sorte que j varie de 1 à 5. Le deuxième ensemble, dénoté par γ_j, comprend les corrélations entre les estimations produites par un panel et son prédécesseur, de sorte que j varie de 1 à 11. Les corrélations ρ sont généralement élevées, comme il faut s'y attendre, mais il convient de souligner qu'elles sont plus faibles pour les données du

chômage que pour celles de l'emploi, ce qui reflète la mobilité élevée de la population active en chômage. Les corrélations γ sont beaucoup plus faibles que les corrélations ρ, mais comme le signale l'auteur, le calcul de ces corrélations est beaucoup moins fiable et leur comportement est quelque peu erratique, affichant parfois une tendance à la hausse. Nous avons calculé les corrélations sérielles d'après les modèles (2.2), en remplaçant les coefficients ϕ par leurs valeurs estimées et nous avons observé en général un ajustement étroit avec les corrélations ρ à tous les décalages (de 1 à 5). Les corrélations à des décalages plus élevés sont différentes des corrélations γ correspondantes, mais, fait intéressant, elles sont la plupart du temps plus élevées et diminuent toujours à mesure que j augmente.

Une autre question ayant trait au modèle (2.2) qu'ont soulevée les arbitres avait trait à la possibilité d'appliquer la transformation logarithmique aux données brutes de façon à stabiliser les variances des erreurs de l'enquête, plutôt que de modéliser les erreurs standardisées. Deux raisons principales incitent à ne pas recourir à la transformation logarithmique dans notre cas. En premier lieu, l'utilisation de cette transformation entraînerait une décomposition multiplicative des taux de chômage de la population, ce qui va à l'encontre de la pratique courante qui consiste à supposer une décomposition additive. À Statistique Canada, les taux de chômage des deux plus grandes provinces sur les quatre examinées dans notre étude sont désaisonnalisés sur la base d'une décomposition additive. Aux États-Unis, les modèles ajustés aux séries du chômage des États reposent eux aussi sur une décomposition additive. Voir Tiller (1992). La deuxième raison est que les changements des variances des erreurs de l'enquête peuvent résulter de modifications du plan d'échantillonnage et, en particulier, de changements des tailles des échantillons. De tels changements produisent des mouvements discrets des variances qui ne peuvent être traités efficacement par la transformation logarithmique. Comme l'a signalé par ailleurs l'un des arbitres, le fait de transformer les données a l'inconvénient de produire une non-linéarité dans l'agrégation des estimations sur les divers panels et/ou petites régions.

Le modèle défini en (2.2) tient compte des deux facteurs à considérer énoncés plus haut. L'application réelle du modèle, toutefois, exige deux modifications:

1. Pour les trois premiers panels, il n'y a pas assez de données antérieures pour permettre l'ajustement d'un modèle AR(3). Par exemple, l'erreur de l'enquête $e_t^{(1)}$ correspond au panel qui fait partie de l'enquête pour la première fois. Pour surmonter ce problème, nous remplaçons les erreurs de l'enquête manquantes par les erreurs de l'enquête correspondant aux panels précédemment choisis dans les mêmes UPÉ ou strates. Par exemple, le modèle AR(2) ajusté à $\tilde{e}_t^{(2)}$ est

$$\tilde{e}_t^{(2)} = \phi_{21}\,\tilde{e}_{t-1}^{(1)} + \phi_{22}\,\tilde{e}_{t-2}^{(6)} + u_t^{(2)}. \qquad (2.3)$$

Notons que le panel observé pour la deuxième fois au mois t remplace au temps $(t-1)$ le panel observé pour la sixième fois au mois $(t-2)$, de sorte que les deux panels représentent les mêmes UPÉ ou strates. L'utilisation de valeurs substituts pour les erreurs de l'enquête dans le cas des trois premiers panels peut expliquer les modèles différents obtenus pour ces panels, par rapport au modèle obtenu pour les trois autres panels.

2. Les écarts-types vrais des erreurs de l'enquête sont inconnus, tandis que les estimations de l'enquête des écarts-types sont elles-mêmes l'objet d'erreurs d'échantillonnage. Pour surmonter ce problème, nous utilisons des valeurs lissées des écarts-types estimés, obtenues par l'ajustement de la relation

$$(\widetilde{SD})_t = \hat{\gamma}(\widehat{SD})_{t-1} + \hat{\gamma}_0\, t + \sum_{i=1}^{12} \hat{\gamma}_i D_{it}, \qquad (2.4)$$

où les coefficients γ sont estimés par les moindres carrés ordinaires. La notation $(\widehat{SD})_t$ définit l'estimation brute, non lissée, de l'écart-type lié au plan de l'estimateur moyen de l'enquête, $\bar{y}_t$, au mois t et les $\{D_{it}\}$ sont des variables fictives tenant compte des effets saisonniers mensuels de telle sorte que $D_{it} = 1$ quand $t = 12k + i$, $k = 0, 1, \ldots, i = 1, \ldots, 12$ et $D_{it} = 0$ dans les autres cas. Les écarts-types lissés des erreurs propres aux panels sont donnés par $\widetilde{SD}(e_t^{(j)}) = \sqrt{6}(\widetilde{SD})_t$. Ces dernières estimations sont utilisées comme substituts des écarts-types vrais, qui sont inconnus.

2.4 Représentation sous forme d'espace d'états et estimation du modèle représentant les estimateurs de l'enquête

Il découle de (2.1) que les estimateurs des panels peuvent être modélisés par

$$y_t^{(j)} = L_t + S_t + \epsilon_t + e_t^{(j)}, \quad j = 1, \ldots, 6, \qquad (2.5)$$

où

$$L_t = L_{t-1} + R_{t-1} + \eta_{Lt}; \quad R_t = R_{t-1} + \eta_{R1};$$

$$\sum_{j=0}^{11} S_{t+j} = \eta_{St}, \qquad (2.6)$$

avec $\{\epsilon_t\}$, $\{\eta_{Lt}\}$, $\{\eta_{Rt}\}$ et $\{\eta_{St}\}$ définis comme en (2.1). Les modèles distincts définis par (2.5), (2.6) et (2.2) peuvent être fondus en une représentation compacte d'espace d'états avec $\underset{\sim}{y}_t' = (y_t^{(1)}, \ldots, y_t^{(6)})$ comme données d'entrée, semblable à la représentation décrite dans Pfeffermann (1991). Suivant cette représentation, les erreurs de l'enquête (et, dans la présente étude, les termes irréguliers du recensement également) sont inclus dans le vecteur d'états, de façon qu'il n'y ait pas de termes résiduels dans l'équation des observations définie par (2.5). Contrairement à la situation décrite dans Pfeffermann (1991), toutefois, la matrice de transition et la matrice de

variance-covariance (V-C) des termes d'erreur des états ne sont pas fixes dans le temps, car elles dépendent des variances liées au plan des erreurs de l'enquête qui, comme il a été expliqué à la section 2.3, changent avec le temps.

La représentation du modèle sous forme d'espace d'états nous permet de mettre à jour, de lisser ou de prédire les vecteurs d'états et, par conséquent, la tendance, la composante saisonnière et la valeur de la population à n'importe quel mois t, au moyen du filtre de Kalman. Désignons par $\underset{\sim}{\alpha}_t$ le vecteur d'états correspondant au mois t. Le vecteur d'états comprend le niveau de la tendance, l'accroissement et les effets saisonniers, les biais liés aux groupes de renouvellement et les erreurs de l'enquête. Voir Pfeffermann (1991) pour plus de détails. Par "mise à jour", nous entendons l'estimation de $\underset{\sim}{\alpha}_t$, au mois t, d'après toutes les données existant jusqu'au mois t inclusivement. Par "lissage", nous entendons l'estimation de $\underset{\sim}{\alpha}_t$ d'après toutes les données disponibles pour l'ensemble des mois qui précèdent et qui suivent le mois t. Le lissage est nécessaire pour améliorer des estimations passées, par exemple lorsqu'on estime les effets saisonniers ou qu'on estime des variations des valeurs de la population ou de la tendance. Quant à la "prédiction" des vecteurs d'états relatifs à des mois postérieurs à la période d'observation de l'échantillon, elle est importante pour l'élaboration de politiques. Les prédictions portant sur la période d'observation de l'échantillon permettent d'évaluer la performance du modèle, p. ex. en comparant les estimations de panel découlant de la prédiction de vecteurs d'états avec les estimations réelles. Voir la section 3 pour plus de détails. La théorie des modèles d'espace d'états et du filtre de Kalman est élaborée dans plusieurs publications; voir Pfeffermann (1991) pour les équations de filtrage et de lissage, avec mention de références. Notons que les équations de filtrage et de lissage fournissent non seulement les trois ensembles d'estimateurs pour n'importe quel mois donné t, mais aussi les matrices V-C des erreurs d'estimation correspondantes.

L'application réelle du filtre de Kalman exige l'estimation des paramètres inconnus du modèle et l'initialisation du filtre, c'est-à-dire l'estimation du vecteur d'états initial $\underset{\sim}{\alpha}_0$ et de la matrice V-C correspondante des erreurs d'estimation. Pour une petite région unique, les paramètres inconnus du modèle sont les quatre variances des termes d'erreur du modèle du recensement (2.1), ainsi que les huit coefficients d'autorégression et les six variances résiduelles des modèles des erreurs propres aux panels (2.2). (Les moyennes des groupes de renouvellement sont incluses dans les vecteurs d'états à titre de coefficients fixes, invariables dans le temps.) Afin de réduire le nombre de paramètres inconnus dans le modèle d'espace d'états combiné, nous supposons que $\sigma_j^2 = \sigma^2 \times \tilde{\sigma}_j^2, j = 1, \ldots, 6$, où les $\{\sigma_j^2\}$ sont les variances résiduelles dans (2.2) et les $\tilde{\sigma}_j^2$ sont les estimations des variances résiduelles obtenues par l'ajustement des équations d'autorégression aux pseudo-erreurs de l'enquête, $e_{t,p}^{(j)}$, définies à la section 2.3. Cette hypothèse réduit le nombre de paramètres inconnus de 18 à 13. (Les estimations $\tilde{\sigma}_j^2$ sont très voisines pour $j = 4, 5, 6$ et ont été supposées égales.)

Si l'on suppose que les termes d'erreur dans les modèles du recensement et des erreurs de l'enquête ont une distribution normale, les paramètres inconnus du modèle peuvent être estimés par une maximisation de la vraisemblance. Voir Pfeffermann et Burck (1991) pour une brève description de la façon d'utiliser l'algorithme de maximisation de la méthode de notation et pour des détails sur l'initialisation du filtre. Cet article inclut des références à des analyses plus approfondies.

2.5 Rajustements pour tenir compte de l'utilisation de valeurs estimées des paramètres

Une fois que les paramètres inconnus du modèle ont été estimés, les équations du filtre de Kalman peuvent être appliquées, les valeurs réelles des paramètres étant remplacées par les estimations. Comme il a été indiqué à la section 2.4, le filtre de Kalman produit non seulement des estimations des vecteurs d'états, mais aussi des matrices V-C des erreurs d'estimation correspondantes. L'utilisation de ces matrices V-C peut toutefois poser un problème du fait que ces dernières ne tiennent pas compte de la variation additionnelle attribuable au fait que les paramètres sont estimés, ce qui peut engendrer une sous-estimation des variances vraies.

En notation mathématique, supposons que $\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}})$ soit l'estimateur de $\underset{\sim}{\alpha}_t$, au mois t, d'après toutes les données disponibles jusqu'à un mois donné n, où $\hat{\underset{\sim}{\lambda}}$ représente les estimateurs des paramètres inconnus du modèle. L'erreur d'estimation peut être ainsi décomposée:

$$[\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}}) - \underset{\sim}{\alpha}_t] = [\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}) - \underset{\sim}{\alpha}_t] + [\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}}) - \hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda})]. \quad (2.7)$$

Il s'agit de la somme de l'erreur qu'on obtiendrait si $\underset{\sim}{\lambda}$ était connu et de l'erreur attribuable à l'estimation de $\underset{\sim}{\lambda}$. Les deux termes du côté droit de (2.7) sont non corrélés. Un moyen simple de vérifier cette propriété consiste à observer que $\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}) = E(\underset{\sim}{\alpha}_t \mid Y, \underset{\sim}{\lambda})$, où Y représente l'ensemble des données disponibles. Puisque la condition porte sur Y et $\underset{\sim}{\lambda}$, $[\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}}) - \hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda})]$, est non stochastique, tandis que $E\{[\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}) - \underset{\sim}{\alpha}_t] \mid Y, \underset{\sim}{\lambda}\} = \underset{\sim}{0}$. Il s'ensuit, d'après (2.7), que

$$\begin{aligned} Q_t &= E\{[\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}}) - \underset{\sim}{\alpha}_t][\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}}) - \underset{\sim}{\alpha}_t]'\} \\ &= E\{[\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}) - \underset{\sim}{\alpha}_t][\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}) - \underset{\sim}{\alpha}_t]'\} \\ &\quad + E\{[\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}}) - \hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda})][\hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}}) - \hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda})]'\} \\ &= A_t + B_t. \end{aligned} \quad (2.8)$$

Pour estimer A_t et B_t, nous faisons porter la condition sur Y et nous suivons la méthode proposée par Hamilton (1986). Selon cette méthode, des réalisations $\underset{\sim}{\lambda}_{(k)}$, $k = 1, \ldots, K$ sont produites à partir de la distribution postérieure normale asymptotique de $\underset{\sim}{\lambda}$, c.-à-d. à partir d'une distribution $N(\hat{\underset{\sim}{\lambda}}, \hat{\Lambda})$ où $\hat{\underset{\sim}{\lambda}}$ est l'estimateur du maximum de vraisemblance de $\underset{\sim}{\lambda}$ et $\hat{\Lambda}$ est la matrice V-C asymptotique de $\hat{\underset{\sim}{\lambda}}$. ($\hat{\underset{\sim}{\lambda}}$ et $\hat{\Lambda}$ sont tous deux obtenus par la méthode de notation.) Le filtre de Kalman est alors appliqué avec

chacune de ces réalisations, ce qui donne des estimations $\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}_{(k)})$ avec matrices V-C $P_t(\underset{\sim}{\lambda}_{(k)})$. Les matrices A_t et B_t sont ainsi estimées:

$$\hat{A}_t = \frac{1}{k} \sum_{k=1}^{K} P_t(\underset{\sim}{\lambda}_{(k)});$$

$$\hat{B}_t = \frac{1}{k} \sum_{k=1}^{K} [\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}_{(k)}) - \hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}})][\hat{\underset{\sim}{\alpha}}_t(\underset{\sim}{\lambda}_{(k)}) - \hat{\underset{\sim}{\alpha}}_t(\hat{\underset{\sim}{\lambda}})]'. \tag{2.9}$$

Ansley et Kohn (1986) proposent un estimateur pour B_t fondé sur une approximation de série de Taylor de premier ordre. L'utilisation de leur estimateur exige moins de calculs, mais la méthode proposée par Hamilton est un peu plus flexible sur le plan des hypothèses posées, et elle donne une meilleure idée de la mesure dans laquelle les résultats produits par le filtre de Kalman sont sensibles aux erreurs des estimateurs des paramètres.

2.6 Modélisation conjointe pour un groupe de petites régions

Le modèle examiné jusqu'ici concerne une seule région. Quand les tailles d'échantillon des diverses régions sont peu élevées, on peut souvent obtenir des estimateurs plus efficaces en modélisant en outre les relations transversales entre les valeurs de la population de plusieurs régions. De toute évidence, l'accroissement d'efficacité résultant d'une telle modélisation conjointe dépend de la taille des échantillons des petites régions et de la similitude, entre les régions, du comportement des valeurs de la population dans le temps.

Les erreurs de l'enquête étant indépendantes entre les régions, toute modélisation conjointe des estimateurs de l'enquête s'applique seulement au modèle du recensement. Pour modéliser les taux de chômage dans les quatre provinces de l'Atlantique, nous adoptons la démarche de Pfeffermann et Burck (1990) et permettons des corrélations contemporaines non nulles entre les termes d'erreur correspondants des modèles du recensement valables dans ces provinces. Par conséquent, si $\underset{\sim}{y}'_{ta} = (\epsilon_t^{(a)}, \eta_{Lt}^{(a)}, \eta_{Rt}^{(a)}, \eta_{St}^{(a)})$ dénote le vecteur des termes d'erreur au temps t associé au modèle du recensement valable dans la région a, on suppose que $C_{a,b} = E(\underset{\sim}{y}_{ta}\,\underset{\sim}{y}'_{tb})$ est diagonale, mais comporte peut-être des covariances non nulles sur la diagonale principale. La conséquence réelle de cette hypothèse est que si, par exemple, il y a une augmentation sensible du niveau de la tendance dans une province, on peut s'attendre à ce que des augmentations semblables surviennent dans d'autres provinces.

Le modèle conjoint résultant, valable pour les quatre provinces (ou plus généralement pour un groupes de régions), peut encore être présenté sous forme d'un espace d'états; voir les équations (2.7) et (2.8) dans Pfeffermann et Burck (1990). L'ajustement de ce modèle pose toutefois un problème important, du fait que l'estimation conjointe de tous les paramètres inconnus exige trop de l'ordinateur en temps de calcul et en espace de stockage. (Le programme informatique rédigé pour l'application de la méthode de notation utilise des dérivées du premier ordre numériques, de sorte que le calcul de chaque dérivée exige un balayage distinct de la totalité des données. Chaque balayage nécessite le calcul des équations du filtre de Kalman pour chaque mois inclus dans la période d'observation de l'échantillon.)

Pour solutionner ce problème, nous avons d'abord ajusté les modèles définis par (2.5), (2.6) et (2.2) séparément pour chacune des provinces. Nous avons également supposé des corrélations égales entre les termes d'erreur correspondants des modèles du recensement des différentes provinces, c'est-à-dire

$$\phi_{a,b} = C_{a,a}^{-1/2}\, C_{a,b}\, C_{b,b}^{-1/2} = \phi \qquad 1 \le a,b \le 4, \tag{2.10}$$

où $C_{a,a} = E(\underset{\sim}{y}_{ta}\,\underset{\sim}{y}'_{ta})$. Les quatre corrélations maximisant la vraisemblance du modèle conjoint ont été déterminées par une méthode de recherche par quadrillage, les autres paramètres du modèle étant maintenus fixes à leur valeur préalablement estimée.

L'hypothèse des corrélations égales réduit de beaucoup le nombre de paramètres inconnus. Elle se justifie aussi par le petit nombre de régions examinées dans cette étude, qui permet de supposer qu'aucune autre structure préexistante régissant ces corrélations ne pourrait être détectée de façon sûre. Plus concrètement, une simple ventilation de la population active par industrie (tableau 1 de la section 3) montre des fréquences relatives très semblables dans les quatre provinces, ce qui laisse croire à un degré élevé d'homogénéité de leurs économies.

2.7 Modifications pour se prémunir contre les défaillances du modèle

L'utilisation d'un modèle pour la production de statistiques officielles soulève la question de la protection contre d'éventuelles défaillances du modèle. Comme il est indiqué ci-après, il n'est pas possible de tester le modèle chaque fois que de nouvelles données deviennent disponibles. C'est pourquoi il faut doter le modèle d'un mécanisme intrinsèque visant à assurer la robustesse des estimateurs lorsque le modèle ne tient plus.

Pour la modélisation des séries sur la population active dans de petites régions, nous avons utilisé la modification proposée par Pfeffermann et Burck (1990). En vertu de cette modification, les estimations des vecteurs d'états mis à jour à n'importe quel temps t sont soumises à la contrainte suivante:

$$\sum_{a=1}^{A} w_{ta}\,\hat{Y}_{ta} = \sum_{a=1}^{A} w_{ta}\,\bar{y}_{ta} \qquad t = 1, 2, \ldots, \tag{2.11}$$

où $\hat{Y}_{ta}$ est l'estimateur fondé sur le modèle de la valeur de la population Y_{ta} dans la région a, $\bar{y}_{ta} = 1/6 \sum_{j=1}^{6} y_{ta}^{(j)}$ est l'estimateur de l'enquête correspondant et $w_{ta} = M_{ta}/M_t$ est la taille relative de la population active dans cette région, c'est-à-dire que $M_t = \sum_{a=1}^{4} M_{ta}$ et $\sum_{a=1}^{4} w_{ta} = 1$.

Notons que $\sum_{a=1}^{A} w_{ta} \hat{Y}_{ta}$ et $\sum_{a=1}^{A} w_{ta} \bar{y}_t$ sont respectivement l'estimateur fondé sur le modèle et l'estimateur direct de l'enquête de la valeur totale de la population pour le groupe de régions examiné. La condition 2.11 peut aussi être écrite sous la forme $\sum_{a=1}^{A} w_{ta} \bar{e}_{ta} = 0$ où $\bar{e}_{ta} = \sum_{j=1}^{6} e_{ta}^{(j)}/6$ est l'erreur de l'enquête moyenne pour la région a. Pfeffermann et Burck (1990) montrent comment modifier les équations du filtre de Kalman de façon qu'elles produisent l'estimateur des vecteurs d'états assujetti à cette contrainte et sa matrice V-C appropriée selon le modèle (sans la contrainte), pour chaque mois t.

La justification de cette modification est simple. On suppose que la taille totale de l'échantillon de l'ensemble des régions est suffisamment grande et, par conséquent, que les estimateurs globaux de l'enquête sont fiables. Cette hypothèse, en fait, dicte le niveau d'agrégation requis; voir ci-dessous. En obligeant les estimateurs globaux fondés sur le modèle à coïncider avec les estimateurs globaux de l'enquête, l'analyste s'assure que toute variation réelle des valeurs de la population reflétée dans les estimateurs de l'enquête sera aussi reflétée dans les estimateurs fondés sur le modèle. Signalons qu'en l'absence d'une telle contrainte imposée aux estimateurs, s'il survenait par exemple des variations soudaines du niveau de la série, celles-ci ne se répercuteraient sur les estimateurs fondés sur le modèle que plusieurs mois plus tard, car ces estimateurs dépendent non seulement des données courantes, mais aussi des données passées. Par contre, si aucune variation prononcée ne survient, on peut s'attendre à ce que les estimateurs fondés sur le modèle se conforment approximativement aux contraintes, même si ces dernières ne sont pas imposées explicitement. Par conséquent, au cours de périodes normales, les estimateurs assujettis à la contrainte devraient être presque aussi performants que les estimateurs exempts de contrainte.

L'hypothèse selon laquelle la taille totale de l'échantillon de l'ensemble des régions est élevée et donc que l'estimateur global de l'enquête est suffisamment proche de la valeur correspondante de la population est critique. Elle garantit (probabilité élevée) que la modification interviendra uniquement s'il survient des changements réels dans les valeurs de la population, et non par suite de vastes erreurs d'échantillonnage. Il convient de signaler, comme l'a d'ailleurs mentionné l'un des arbitres, que dans l'application de la méthode aux provinces de l'Atlantique décrite à la section 3, l'estimateur global ne se fonde que sur quatre provinces, de sorte que son erreur-type est d'environ 50% des erreurs-types des estimateurs de l'enquête pour les provinces, selon les tailles des échantillons des provinces. (Les estimateurs de l'enquête pour les provinces sont conditionnellement indépendants, les conditions étant les valeurs de la population pour les provinces correspondantes.) Par conséquent, si les contraintes doivent être utilisées en pratique, l'agrégation devrait être faite sur un plus grand nombre de provinces ou sur d'autres petites régions.

Les deux autres méthodes suivantes ont été suggérées pour aborder le problème de la robustesse:

i) Procéder à une détection des valeurs aberrantes de la série chronologique, comme il es proposé par exemple dans Chang, Tiao et Chen (1988).

ii) Modéliser les séries chronologiques des proportions $\{\hat{\pi}_{ta} = \bar{y}_{ta}/\sum_{a=1}^{A} \bar{y}_{ta},\ a = 1, \ldots, (A-1)\}$, si ces séries chronologiques affichent un profil plus lisse que les séries $\{\bar{y}_{ta}\}$.

La détection des valeurs aberrantes est un important aspect de toute activité de modélisation, mais il reste à trouver la façon de modifier les estimations des valeurs de la population une fois que des observations (estimations de l'enquête) sont détectées à titre de valeurs aberrantes. Notons à ce propos que notre intérêt porte principalement sur les estimations courantes, c'est-à-dire les estimations les plus récentes disponibles. Dans Chang, Tiao et Chen (1988), la détection des valeurs aberrantes a pour objet de *retrancher* leur effet des observations, de façon à permettre de mieux comprendre la structure sous-jacente de la série et d'améliorer l'estimation des paramètres du modèle. Mais si la cause d'une valeur aberrante est un mouvement réel du niveau des valeurs de la population, ce mouvement ne doit pas être retranché, mais plutôt pris en compte dans les estimateurs fondés sur le modèle. Harrison et Stevens (1976) proposent de tenir compte de tels changements en modifiant la distribution antérieure des vecteurs d'états, p. ex. en accroissant les variances des erreurs des vecteurs d'états de façon à permettre des variations plus rapides des estimateurs des vecteurs d'états. Un exemple de cette façon de procéder est présenté dans Morris et Pfeffermann (1984). Notre méthode consistant à imposer la contrainte selon laquelle les estimateurs fondés sur le modèle doivent coïncider avec les estimateurs globaux de l'enquête offre un processus plus automatique, qui n'exige pas la modification de données antérieures.

La deuxième méthode suggérée pour aborder le problème de la robustesse est attrayante, car on peut s'attendre à ce que les variations brusques des valeurs de la population s'annulent dans les rapports $\hat{\pi}_{ta}$. Le principal inconvénient du recours à cette méthode est que le modèle valable pour les rapports "vrais" π_{ta} est naturellement très différent du modèle représentant les valeurs de la population Y_t tel qu'il est défini en (2.1) et que, notamment, il n'offre plus d'estimations de la tendance et des effets saisonniers, qui sont parmi les éléments les plus importants de notre approche, comme il est mentionné dans l'introduction. Il n'est pas évident, non plus, de trouver comment extraire les estimations des valeurs de la population Y_t du modèle valable pour les rapports $\hat{\pi}_{ta}$ sans faire des hypothèses additionnelles, comme par exemple notre hypothèse selon laquelle l'estimateur global de l'enquête est suffisamment voisin de la valeur correspondante de la population.

L'utilisation de contraintes de la forme (2.11) a été envisagée précédemment par Battese, Harter et Fuller (1988), ainsi que par Pfeffermann et Barnard (1991), pour l'analyse d'enquêtes transversales. Pfeffermann et Burck (1990) présentent des résultats empiriques illustrant la bonne performance des estimateurs modifiés au cours de périodes anormales. Voir aussi la section 3.

3. AJUSTEMENT DU MODÈLE AUX PROVINCES DE L'ATLANTIQUE: RÉSULTATS EMPIRIQUES

Le modèle défini par (2.2), (2.5), (2.6) et (2.10) a été ajusté aux estimateurs de panel mensuels pour les quatre provinces de l'Atlantique, en deux étapes. À la première étape, le modèle défini par (2.2), (2.5) et (2.6) a été ajusté à chacune des provinces séparément. À la deuxième étape, les corrélations définissant la matrice ϕ de (2.10) ont été estimées au moyen d'une recherche par quadrillage (voir la section 2.6). Les estimateurs obtenus sont $\text{Diag}(\phi) = (0.5, 0.25, 0.80, 0.0)$. Les données ayant servi à l'estimation du modèle portent sur les années 1982 à 1988. Les données de 1989 ont servi à faire le diagnostic du modèle, en comparant les résultats obtenus à l'intérieur et à l'extérieur de la période d'observation de l'échantillon.

3.1 Analyse préliminaire

Le tableau 1 présente une ventilation de la population active des quatre provinces, par industrie. Les chiffres du tableau reflètent la situation qui existait en mars 1991. Les tailles (attendues) des échantillons de l'EPA sont également indiquées. Comme on peut le voir, les répartitions en pourcentage dans les quatre provinces sont très voisines, ce qui justifie l'hypothèse de corrélations égales entre les termes d'erreur des modèles du recensement pour les différentes provinces. La similitude des répartitions permet également d'espérer une meilleure efficacité des estimateurs fondés sur le modèle, dans le cas du modèle conjoint, comparativement aux estimateurs qui ne tiennent pas compte des corrélations transversales entre les valeurs de la population pour les différentes provinces.

Deux autres importants facteurs à considérer, mentionnés à la section 2.3, sont les suivants: le modèle devrait tenir compte d'effets possibles liés aux groupes de renouvellement, ainsi que des changements des variances des erreurs de l'enquête avec le temps. Pour obtenir des estimations initiales des effets liés aux groupes de renouvellement, nous avons fait la moyenne des pseudo-erreurs de l'enquête, $e_{t,p}^{(j)} = (y_t^{(j)} - \bar{y}_t)$, $j = 1, \ldots, 6$, sur l'ensemble des mois de la période visée par l'échantillon. Nous avons ensuite divisé les moyennes par les estimations classiques des erreurs-types. (Les erreurs $e_{t,p}^{(j)}$ sont corrélées dans le temps, mais les corrélations sont faibles, car sauf pour les décalages 6, 12, *etc.*, les données de n'importe quel panel donné se rapportent à des UPÉ différentes dans les régions urbaines et à des secteurs de dénombrement différents dans les régions rurales. Voir la section 2.1.) Notons qu'en l'absence d'effets liés aux groupes de renouvellement, $E(e_{t,p}^{(j)}) = 0$ pour tous les j et t, peu importe le modèle postulé pour les valeurs de la population.

Cette analyse préliminaire (indépendante du modèle) donne des résultats semblables aux résultats obtenus avec le modèle complet, présentés au tableau 2 de la section 3.3.

Examinons maintenant les variances des erreurs de l'enquête.

La figure 1 présente le graphique des effets saisonniers des estimateurs globaux de l'enquête dans les quatre provinces, ainsi que des effets saisonniers des erreurs-types de ces estimations (multipliés par 100). Comme précédemment, désignons par w_{ta} la taille relative de la population active dans la province a au temps t. L'estimateur global de l'enquête est défini par $y_t^* = \sum_{a=1}^{4} w_{ta} \bar{y}_{ta}$ (équation 2.11). L'erreur-type de y_t^* est $(SD^*)_t = [\sum_{a=1}^{4} w_{ta}^2 (\widehat{SD})_{ta}^2]^{1/2}$. Les effets saisonniers ont été estimés par l'application du mode additif de la procédure X-11, afin qu'ils ne soient liés à aucun modèle particulier. Nous avons choisi le modèle additif puisque nous supposons une décomposition additive pour les estimateurs de l'enquête. (Comme le montre la figure 4, les effets saisonniers des estimateurs globaux de l'enquête produits par la procédure X-11 sont très voisins des effets saisonniers obtenus avec le modèle.)

La figure 1 montre que les erreurs-types sont assujetties à des variations saisonnières, affichant un profil saisonnier qui est très voisin de celui des estimateurs de l'enquête et donc des valeurs correspondantes de la population.

Tableau 1

Population active par industrie dans les provinces de l'atlantique, mars 1991

	Nouvelle-Écosse		Nouveau-Brunswick		Terre-Neuve		Île-du-Prince-Édouard	
Taille de l'échantillon	4,409		3,843		2,970		1,421	
	Milliers	%	Milliers	%	Milliers	%	Milliers	%
Agriculture	7	1.7	7	2.3	0.5	0.2	6.0	9.8
Autres industries primaires	18	4.4	13	4.2	18.0	7.7	4.0	6.6
Secteur manufacturier	44	10.7	37	11.9	23.0	9.9	6.0	9.8
Construction	24	5.9	21	6.8	18.0	7.7	4.0	6.6
Transp. et communication	35	8.6	30	9.6	20.0	8.6	5.0	8.3
Commerce	81	19.8	61	19.6	41.0	17.6	10.0	16.4
Finances	20	4.9	12	3.9	6.0	2.6	0.5	0.8
Services	143	35.0	107	34.4	83.0	35.6	19.0	31.1
Administration publique	36	8.8	22	7.0	23.0	9.9	6.0	9.8
Non classé	1	0.2	1	0.3	0.5	0.2	0.5	0.8
Total	409	100.0	311	100.0	233.0	100.0	61.0	100.0

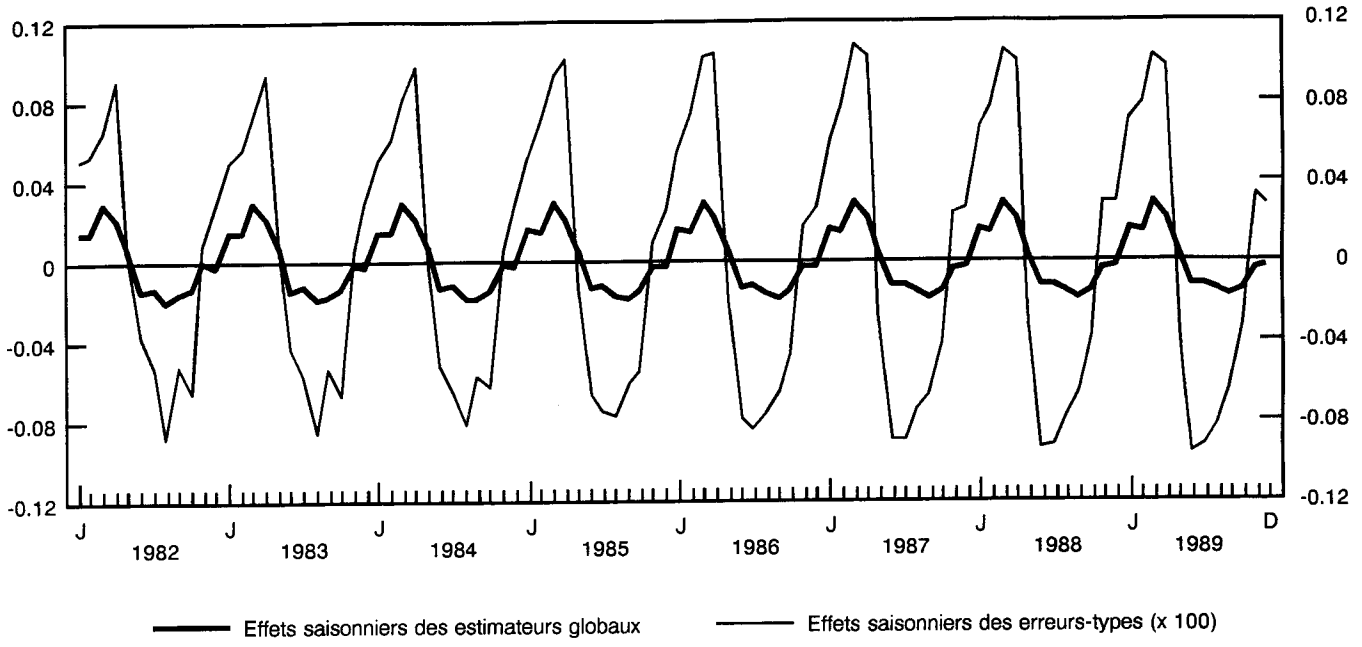

Figure 1. Effets saisonniers des estimateurs globaux de l'enquête et des erreurs-types des estimateurs globaux de l'enquête (× 100)

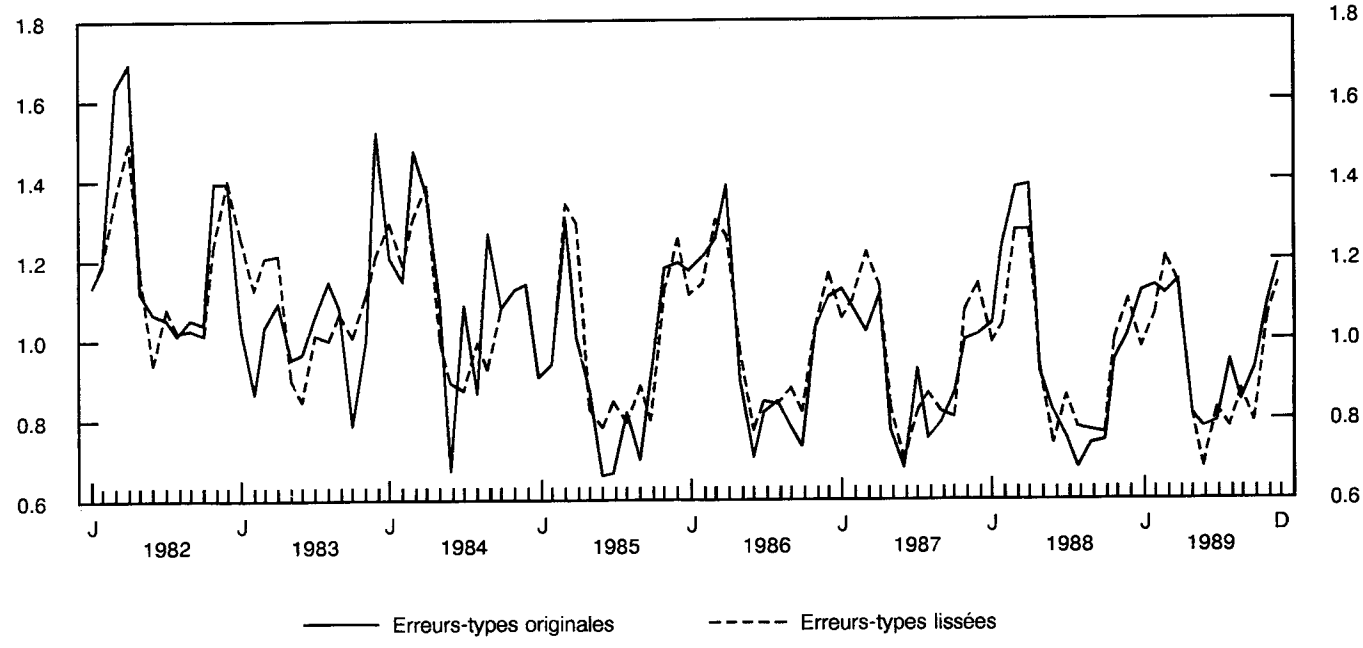

Figure 2. Erreurs-types originales et lissées des estimateurs de l'enquête (× 100) – Île-du-Prince-Édouard

Comme il a été mentionné à la section 2.3, plutôt que d'utiliser les estimations originales des erreurs-types dues au plan dans les modèles ajustés aux erreurs propres aux panels, nous utilisons des valeurs lissées, ce qui réduit les effets des erreurs d'échantillonnage sur les estimateurs de l'enquête. La figure 2 présente le graphique des deux ensembles d'estimateurs pour l'Île-du-Prince-Édouard (Î.-P.-É.), qui est la plus petite province de la région de l'Atlantique, et donc celle où les tailles des échantillons sont les plus faibles. Comme on peut le voir, l'effet du lissage est d'éliminer les estimations brutes extrêmes, mais, par ailleurs, les valeurs lissées suivent en gros le même profil que les estimations brutes. Les graphiques des autres provinces présentent des profils semblables, mais les écarts entre les estimations brutes et les estimations lissées sont moins prononcés en raison du fait que les tailles des échantillons sont plus élevées dans ces provinces.

Nous concluons cette section en définissant les modèles postulés pour les effets saisonniers dans les quatre provinces. Notre modèle initial supposait des variances fixes pour les termes d'erreur $\eta_{St} = \sum_{j=0}^{11} S_{t+j}$, $t = 1, 2, \ldots$ (voir l'équation 2.1). Les erreurs prédites $\hat{\eta}_{St} = \sum_{j=0}^{11} \hat{S}_{t+j}$ fournies par ce modèle diminuaient en valeur absolue en fonction du temps dans trois des quatre provinces, et augmentaient dans la dernière. Notons qu'en vertu du modèle défini par (2.1), avec des variances constantes des termes d'erreur des états, le filtre de Kalman converge vers un état stable dans lequel les matrices V-C des estimateurs des vecteurs d'états, et donc les $\hat{\eta}_{St}$, sont constants. Donc, nous avons modifié le modèle initial de telle façon que $\mathrm{VAR}(\eta_{St}) = \sigma_S^2 \times g(t)$ où, pour la Nouvelle-Écosse, Terre-Neuve et l'Î.-P.-É., $g(t) = t^{(-3/2)}$, tandis que pour le Nouveau-Brunswick, $g(t) = t^{1/2}$.

3.2 Résultats

3.2.1 Biais liés aux groupes de renouvellement

Le tableau 2 montre les biais liés aux groupes de renouvellement (BGR) et leurs erreurs-types estimées (E.-T.) pour les quatre provinces, en vertu du modèle complet défini par (2.3), (2.5), (2.6) et (2.10).

Tableau 2

Biais liés aux groupes de renouvellement et erreurs-types pour les quatre provinces ($\times$ 100)

Panels	Nouvelle-Écosse		Nouveau-Brunswick		Terre-Neuve		Î.-P.-É.	
	BGR	E.-T.	BGR	E.-T.	BGR	E.-T.	BGR	E.-T.
1	−0.20	0.10	−0.02	0.11	−0.47	0.13	0.32	0.17
2	0.18	0.09	0.40	0.10	0.42	0.12	0.18	0.15
3	0.32	0.08	0.24	0.09	0.47	0.12	0.31	0.15
4	0.06	0.07	0.01	0.09	0.18	0.12	0.03	0.15
5	−0.03	0.08	−0.15	0.10	−0.10	0.13	−0.25	0.16
6	−0.34	0.08	−0.50	0.11	−0.50	0.14	−0.60	0.16

Les biais liés aux groupes de renouvellement affichent un profil relativement uniforme entre les provinces. Ainsi, les biais relatifs au 3e et au 6e panel sont tous hautement significatifs selon la statistique t classique, affichant un signe positif pour le 3e panel et un signe négatif pour le 6e panel. Les biais relatifs au 4e et au 5e panel ont eux aussi le même signe dans toutes les provinces, et sont tous non significatifs.

Pour le 2e panel, tous les biais sont positifs, mais celui de l'Î.-P.-É. est non significatif (l'Î.-P.É. est la province ayant la taille d'échantillon la plus faible). C'est également pour l'Î.-P.-É. que le signe du biais du 1er panel est différent de celui des autres provinces.

Comme il a été indiqué à la section 2.3, il y a plus d'une raison possible pour expliquer l'existence de biais liés aux groupes de renouvellement, mais les résultats présentés dans le tableau incitent fortement à penser que quelle que soit la raison, les biais observés pour certains des panels sont réels, et non pas attribuables uniquement aux erreurs d'échantillonnage. Un inconvénient de la présente analyse, toutefois, est que les biais liés aux groupes de renouvellement sont supposés fixes dans le temps. Un modèle plus souple est proposé à la section 4.

3.2.2 Qualité de l'ajustement

A. TEST DE LA NORMALITÉ

Définissons par $I_{ta}^{(j)} = (y_{ta}^{(j)} - y_{ta|(t-1)}^{(j)})$ l'erreur de prédiction qu'on obtient en prédisant l'estimateur du j-ième panel un mois à l'avance, et posons $I'_{ta} = (I_{ta}^{(1)}, \ldots, I_{ta}^{(6)})$. L'utilisation de l'estimation du maximum de vraisemblance dans cette étude suppose que les vecteurs I_{ta} sont des écarts aléatoires normaux (voir la section 2.4). Pour tester cette hypothèse, nous avons calculé la distribution empirique des erreurs de prédiction standardisées $\{(SI)_{ta}^{(j)} = [I_{ta}^{(j)}/\widehat{SD}(I_{ta}^{(j)})], t = (k+1), \ldots, N\}$, et nous l'avons comparée à la distribution normale réduite au moyen de la statistique de test de Kolmogorov-Smirnov. La statistique de test a été calculée pour chacun des six panels et chacune des quatre provinces, et a donné des valeurs P supérieures à 0.15 dans 21 cas sur 24. (Les tests ont été faits au moyen de la procédure PROC UNIVARIATE du logiciel SAS. Dans cette procédure, si la taille de l'échantillon est plus grande que cinquante, comme dans notre cas, les données sont testées à l'égard d'une distribution normale avec moyenne et variance égales à la moyenne et à la variance de l'échantillon.) L'application de la même procédure de test aux erreurs de prédiction standardisées $\{(SI)_{ta} = [I_{ta}/\widehat{SD}(I_{ta})], t = (k+1), \ldots, N\}$ où $I_{ta} = [\sum_{j=1}^{6} I_{ta}^{(j)}/6]$, donne des valeurs P supérieures à 0.15 dans chacune des quatre provinces.

Les estimateurs des écarts-types des erreurs de prédiction ayant servi aux tests sont ceux produits par le filtre de Kalman, sans prise en considération de la composante de variance résultant de l'estimation des paramètres (voir la section 2.5). Cette dernière composante est négligeable même pour l'Î.-P.-É., qui compte les échantillons ayant la taille la plus faible parmi les quatre provinces. Nous reviendrons sur cette observation à la section 3.4.

B. ERREURS DE PRÉDICTION AFFECTANT DIFFÉRENTS PRÉDICTEURS

Le tableau 3 présente des statistiques sommaires comparant le profil des erreurs de prédiction entre les quatre provinces, selon les résultats fournis par trois différents ensembles d'estimateurs des vecteurs d'états: 1) Les estimateurs obtenus en vertu des modèles séparés (MS) définis par (2.2), (2.5) et (2.6); 2) les estimateurs obtenus en vertu du modèle conjoint (MC) défini par (2.2), (2.5), (2.6) et (2.10); 3) les estimateurs obtenus avec les contraintes imposées au modèle conjoint (ROB) pour les fins de la robustesse (2.11). Nous définissons ci-dessous les statistiques sommaires en utilisant encore une fois la notation $I_{ta}^{(j)} = (y_{ta}^{(j)} - \hat{y}_{ta|(t-1)}^{(j)})$ pour l'erreur de prédiction obtenue lorsqu'on prédit l'estimateur du j-ième panel un mois à l'avance.

$MB_a = \sum_{t=k+1}^{N} (\sum_{j=1}^{6} I_{ta}^{(j)}/6)/(N-k)$ – biais moyen de la prédiction de l'estimateur moyen de l'enquête $\bar{y}_{ta} = \sum_{j=1}^{6} y_{ta}^{(j)}/6$.

$MAB_a = \sum_{j=1}^{6} | \sum_{t=k+1}^{N} I_{ta}^{(j)}/(N-k) | /6$ – biais absolu moyen de la prédiction des estimateurs propres aux panels.

$SQRE_a = \{\sum_{t=k+1}^{N} [1/6 \sum_{j=1}^{6} I_{ta}^{(j)}/\bar{y}_{ta})]^2/(N-k)\}^{1/2}$ – racine carrée de l'erreur de prédiction relative carrée moyenne quand on prédit l'estimateur moyen de l'enquête.

Les statistiques sommaires ci-dessus sont présentées séparément pour la période d'observation de l'échantillon allant de juillet 1983 à décembre 1988, et pour la période postérieure allant de janvier 1989 à décembre 1989. Dans ce dernier cas, les données ont été ajoutées un mois à la fois, c'est-à-dire que pour prédire l'estimateur de l'enquête de février 1989, par exemple, nous avons utilisé les données observées jusqu'à janvier 1989, et ainsi de suite.

Les principales conclusions qui découlent du tableau 3 sont les suivantes:

1) Les résultats obtenus dans le cas des trois ensembles de prédicteurs sont en général très semblables, ce qui indique que pour les données analysées, l'utilisation du modèle conjoint ne produit qu'une légère amélioration par rapport à l'utilisation des modèles séparés, et qu'il n'y a pas de variations brusques du niveau des séries au cours des années examinées.

2) Les erreurs de prédiction des estimateurs de l'enquête sont faibles tant à l'intérieur qu'à l'extérieur de la période d'observation de l'échantillon, ce qui laisse supposer un bon ajustement du modèle. Notons que sauf pour l'Î.-P.-É., les erreurs de prédiction relatives mesurées par la statistique $SQRE_a$ sont toutes inférieures à 7%.

3) Les biais des erreurs de prédiction dans la période postérieure à l'observation de l'échantillon sont plus élevés que dans la période d'observation de l'échantillon, les écarts étant particulièrement prononcés au Nouveau-Brunswick et à l'Î.-P.-É. Ce résultat à lui seul pourrait laisser croire à une certaine défaillance du modèle au cours de l'année 1989. Toutefois, l'examen des erreurs de prédiction mensuelles des panels dans les quatre provinces pour cette année-là (non présentées dans le tableau) indique que même si les erreurs sont en général positives, les biais relativement élevés résultent principalement d'une ou deux erreurs extrêmes, qui ont un effet important sur les statistiques sommaires moyennes, puisqu'il n'y a que 12 mois de données. Il est à signaler, en outre, que les taux de chômage estimés dans les quatre provinces au cours de l'année 1989 se situent entre 0.11 et 0.18, de telle sorte qu'un biais de prédiction de .005, ou même de .009 comme dans le cas de l'Î.-P.-É., n'est pas élevé. De toute évidence, le modèle peut être modifié pour tenir compte de ces biais, si ceux-ci persistent avec l'ajout de nouvelles données. Par ailleurs, rappelons que l'analyse qui précède ne concerne que le biais des erreurs de prédiction, puisque le biais des estimateurs fondés sur le modèle des valeurs de la population correspondantes est régi par les contraintes imposées pour assurer la robustesse (2.11).

Tableau 3

Erreurs de prédiction pour les quatre provinces – statistiques sommaires (× 100)

	Nouvelle-Écosse			Nouveau-Brunswick			Terre-Neuve			Île-du-Prince-Édouard		
	MS	MC	ROB	MS	MC	ROB	MS	MC	ROB	MS	MC	ROB
						7.83 – 12.88						
MB	–.11	–.07	–.06	–.12	–.09	–.06	–.25	–.18	–.08	.06	.14	.15
MAB	.12	.11	.10	.14	.12	.11	.29	.24	.20	.20	.23	.23
SQRE	5.76	5.62	5.70	5.48	5.47	5.47	7.03	6.91	6.96	9.34	9.13	9.17
						1.89 – 12.89						
MB	.14	.11	.04	.47	.47	.46	.36	.33	.17	.84	.85	.86
MAB	.32	.32	.30	.51	.51	.50	.39	.37	.29	.84	.85	.86
SQRE	6.39	6.27	6.82	6.25	6.25	6.32	5.92	5.90	5.61	9.45	9.26	9.30

Compte tenu des résultats très semblables obtenus pour les trois ensembles de prédicteurs examinés, et afin de faire ressortir la performance des contraintes relatives à la robustesse, nous avons délibérément appliqué une correction à la baisse de 33% aux taux de chômage de la période allant de mars 1985 à mars 1987, une correction à la baisse de 25% des taux de la période d'avril 1987 à novembre 1988 et une correction à la hausse de 33% des taux de la période de décembre 1988 à décembre 1989. Ces opérations ont pour effet d'introduire des mouvements soudains des données aux mois $t = 39$, $t = 64$ et $t = 84$. La figure 3 présente les erreurs de prédiction un mois d'avance globales (APE), $I_t^a = \sum_{a=1}^{4} w_{ta} [\sum_{j=1}^{6} (y_{ta}^{(j)} - \hat{y}_{ta|(t-1)}^{(j)})/6]$, selon le modèle conjoint avec et sans les contraintes relatives à la robustesse, ainsi que selon les modèles séparés.

Il ressort clairement de la figure 3 que lorsque les contraintes sont imposées, les APE pour les périodes qui suivent les trois mois affichant des mouvements soudains sont moins élevées que les APE obtenues sans les contraintes. Par exemple, en mars 1985 ($t = 39$), les APE sont très élevées en valeur absolue avec et sans les contraintes, ce qui est évident étant donné que les prédicteurs ne sont fondés que sur les données allant jusqu'à février 1985. Les APE correspondant aux prédicteurs robustes reviennent toutefois à leur niveau normal beaucoup plus vite que les APE relatives aux prédicteurs non robustes. Une situation semblable peut être observée dans les deux autres périodes. Un autre résultat notable du graphique est que dans les périodes qui suivent les mois affichant des mouvements soudains, le modèle conjoint a une meilleure performance que les modèles séparés, même sans l'imposition des contraintes visant la robustesse. Par conséquent, en regroupant l'information des différentes provinces, le modèle conjoint s'adapte plus rapidement au nouveau niveau de la série. Pour d'autres illustrations de la performance des contraintes relatives à la robustesse, voir Pfeffermann et Burck (1990).

C. COMPARAISONS AVEC LES ESTIMATEURS PRODUITS PAR LA PROCÉDURE X-11

À titre d'évaluation finale du bien-fondé du modèle, nous comparons les estimations des effets saisonniers et des niveaux de la tendance produits par le modèle aux estimations produites par la procédure X-11 (Dagum 1980). Cette dernière est reconnue pour sa moins grande dépendance à l'égard des hypothèses particulières du modèle. Elle est la procédure couramment utilisée partout dans le monde pour la désaisonnalisation. La figure 4 présente les effets saisonniers moyens pour les quatre provinces obtenus avec la procédure X-11 et en vertu du modèle. La figure 5 montre les estimations correspondantes des niveaux de la tendance. Les moyennes sont calculées à l'aide des poids (w_{ta}) employés dans les analyses précédentes. Les estimations fondées sur le modèle indiquées dans les deux figures sont les estimations lissées qui, comme celles de la procédure X-11, utilisent toutes les données de la période d'observation de l'échantillon.

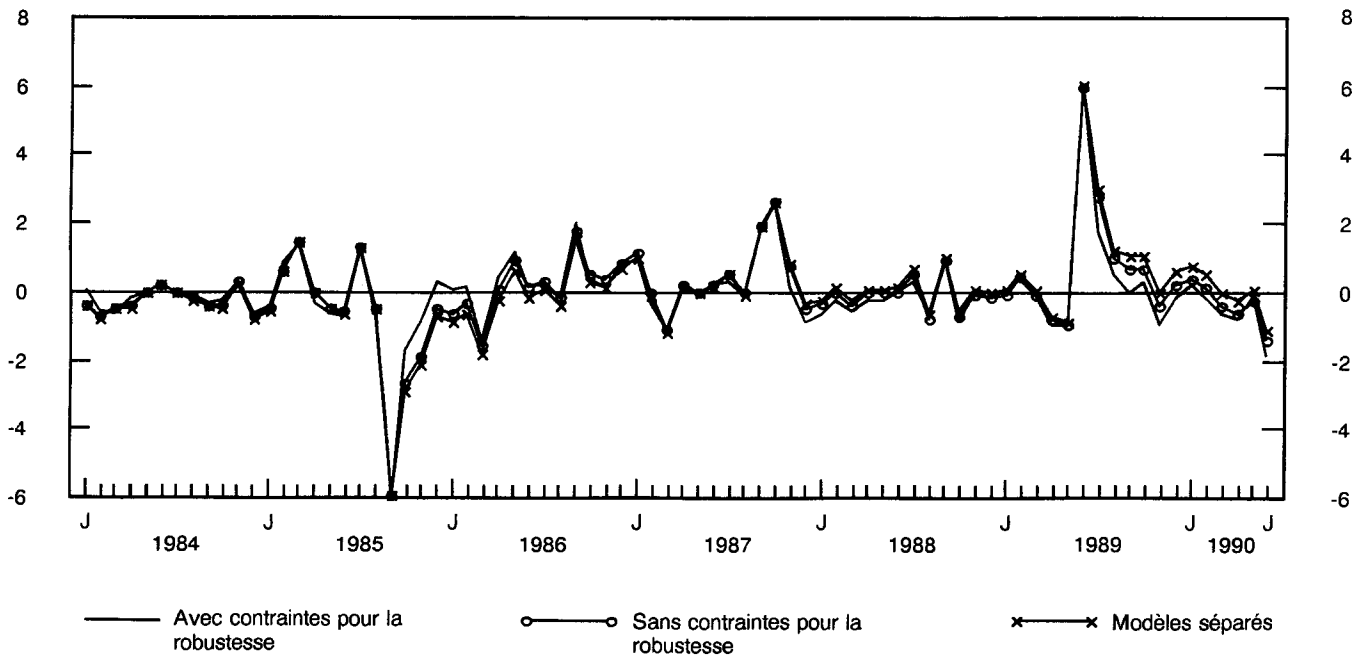

Figure 3. Erreurs de prédiction un mois d'avance globales pour les trois ensembles de prédicteurs (× 100) – Données contaminées

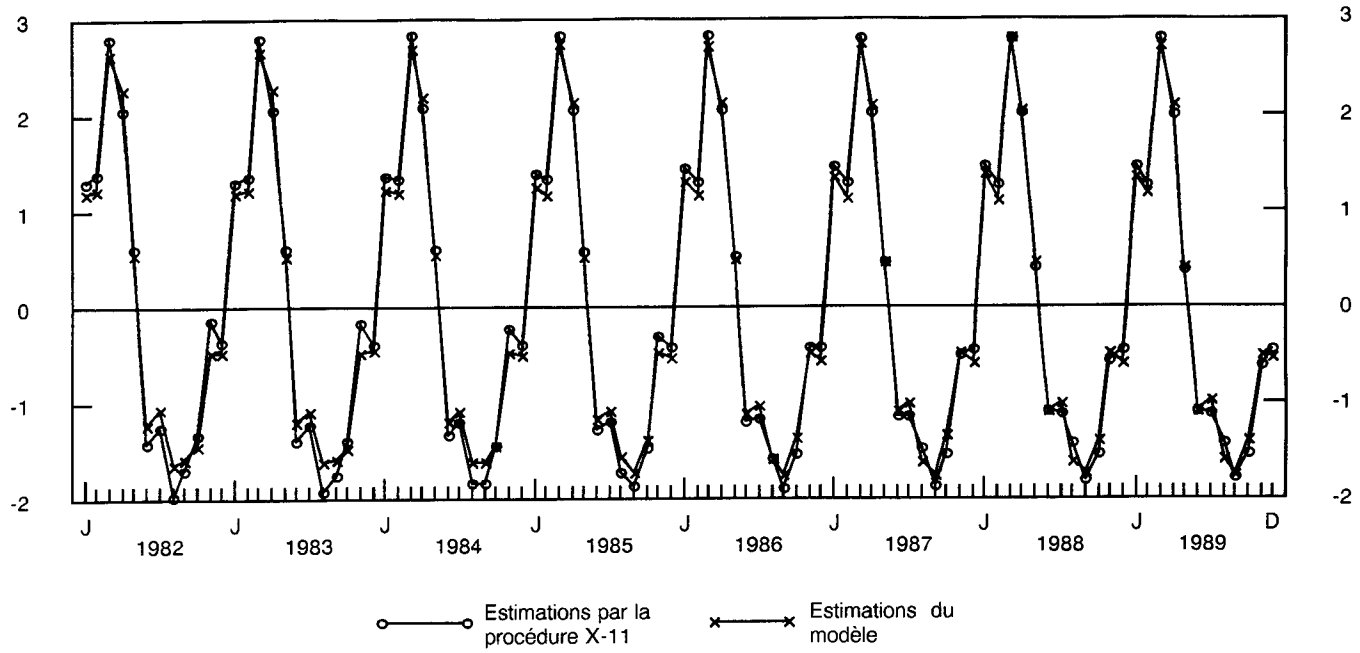

Figure 4. Moyennes pondérées des effets saisonniers par la procédure X-11 et selon le modèle (× 100)

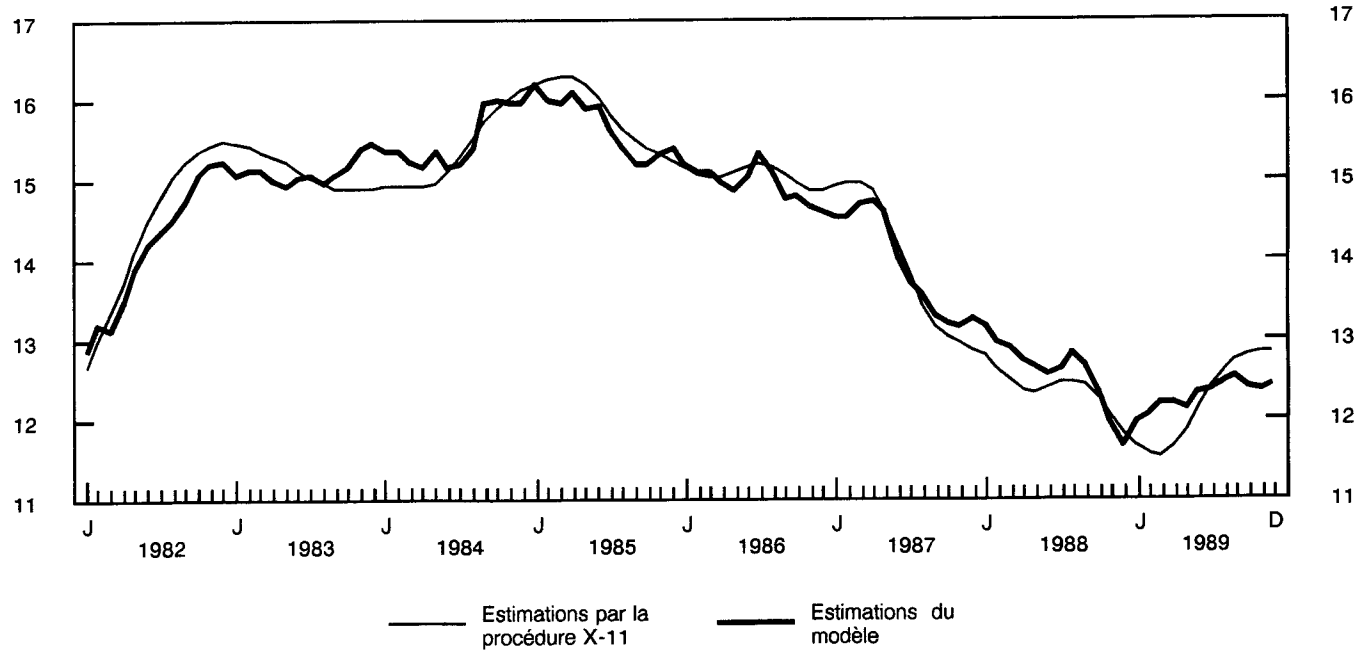

Figure 5. Moyennes pondérées des niveaux de la tendance obtenues par la prodédure X-11 et selon le modèle

Comme on peut le voir, les effets saisonniers produits par les deux méthodes sont très voisins. Les estimations des niveaux de la tendance sont aussi voisines, mais la courbe de la tendance X-11 est plus lisse que la courbe du modèle. Une correspondance étroite semblable entre la procédure X-11 et le modèle est obtenue pour chacune des quatre provinces séparément, y compris, notamment, l'Î.-P.-É., où les tailles des échantillons sont relativement faibles.

3.3 Comparaison entre les estimateurs fondés sur le plan et les estimateurs fondés sur le modèle

Il est mentionné, dans l'introduction, qu'une des principales raisons qui incitent à modéliser les estimateurs bruts de l'enquête est que le modèle produit des estimations des valeurs de la population qui, au moins pour les petites régions, sont plus exactes (lorsque le modèle tient) que celles de l'enquête. Nous avons calculé les deux ensembles d'estimations pour les quatre provinces et constaté que, comme prévu, les estimations produites par les deux méthodes se comportent de façon très semblable, mais que les estimateurs fondés sur le plan sont moins stables, affichant généralement des pointes et des creux plus prononcés. Un important aspect de la comparaison entre les deux ensembles d'estimations, c'est leur aptitude à estimer les variations annuelles des valeurs de la population. Une telle comparaison a l'avantage de ne pas être obscurcie par les effets saisonniers. La figure 6 montre les résultats obtenus pour l'Î.-P.-É. Les estimations fondées sur le modèle sont les valeurs lissées du modèle conjoint qui utilisent toutes les données de tous les mois. Comme on peut le voir, les estimations produites par le modèle sont beaucoup plus stables et ne varient que légèrement d'un mois à l'autre, comparativement aux estimations fondées sur le plan. La figure 7 montre les erreurs-types (E.-T.) des estimateurs des taux de chômage pour l'Î.-P.-É., calculées en vertu du plan (valeurs lissées, voir la figure 2) et en vertu du modèle conjoint. Sont également indiquées l'erreur-type résultant de l'ajustement du modèle séparé défini par (2.2), (2.5) et (2.6) et l'erreur-type correspondante après prise en considération du fait que des estimations des paramètres sont utilisées à la place des valeurs réelles inconnues. Voir la section 2.5 pour plus de détails. (Nous avons calculé cette dernière erreur-type uniquement pour le modèle séparé afin d'économiser du temps d'ordinateur.)

Il y a trois aspects importants qui ressortent des graphiques:

1) Les erreurs-types des estimateurs fondés sur le modèle, en vertu du modèle conjoint, ne sont que légèrement inférieures aux erreurs-types obtenues pour le modèle séparé, mais beaucoup plus faibles que celles des estimateurs de l'enquête.

2) Les erreurs-types des estimateurs fondés sur le modèle suivent le même profil que les erreurs-types des estimateurs de l'enquête, ce qui est une conséquence directe de la prise en considération, dans la définition du modèle, des changements des variances des erreurs de l'enquête au fil du temps. Voir la section 2.3 pour plus de détails.

3) La prise en considération du fait qu'on utilise des valeurs estimées des paramètres dans le calcul de l'erreur-type des estimateurs fondés sur le modèle n'a qu'un effet minime sur l'erreur-type calculée. Rappelons que l'Î.-P.-É. est la province ayant les plus faibles tailles d'échantillon. Pour les autres provinces, l'effet de la prise en considération de l'utilisation d'estimations des paramètres est encore plus faible.

4. RÉSUMÉ

Le présent article montre que des données recueillies en vertu d'un plan de sondage complexe, comprenant plusieurs degrés de sélection et des groupes de renouvellement, peuvent être modélisées avec succès par un modèle relativement simple. Le modèle comprend deux parties: le modèle du recensement, qui représente les valeurs de la population, et le modèle des erreurs de l'enquête, qui décrit les relations qui existent dans les séries chronologiques des erreurs de l'enquête. L'utilisation du modèle donne des estimateurs plus exacts des valeurs de la population et de leurs composantes comme la tendance et la saisonnalité, et elle permet d'estimer l'erreur-type de ces estimateurs d'une façon relativement simple. On peut modifier les équations du modèle de manière à assurer la robustesse des estimateurs fondés sur le modèle et ainsi se protéger contre des défaillances possibles du modèle.

Le modèle utilisé dans cet article peut être prolongé dans diverses directions. Le premier effort devrait consister à appliquer simultanément le modèle à un plus grand nombre de provinces ou à d'autres petites régions, pour s'assurer que les estimateurs globaux de l'échantillon $\sum_{a=1}^{A} w_{ta}\,\bar{y}_{ta}$ sont suffisamment proches des valeurs correspondantes de la population. Voir l'analyse présentée à la section 2.7. L'incorporation au modèle d'un mécanisme de détection des valeurs aberrantes, afin de mieux mesurer la performance et l'adéquation du modèle, est une autre addition valable.

Deux autres extensions pourraient consister à abandonner l'hypothèse d'une variance constante pour le terme d'erreur ϵ_t dans le modèle du recensement et à laisser varier dans le temps les biais liés aux groupes de renouvellement . La première extension découle de l'observation faite à la section 3.1 selon laquelle les variances des erreurs de l'enquête sont l'objet d'effets saisonniers, affichant un profil saisonnier semblable à celui des estimations brutes. L'ajustement des équations (2.4) dans les quatre provinces révèle également l'existence d'une légère tendance dans les variances qui, encore une fois, est semblable à celle des estimations brutes de l'enquête. Ainsi, les variances des erreurs de l'enquête semblent dépendre de la grandeur des estimateurs de l'enquête, ce qui laisse croire que les variances $\sigma_t^2 = V(\epsilon_t)$ changent avec le niveau des valeurs de la population. Comme première approximation, on pourrait supposer que σ_t^2 est proportionnel à la variance correspondante de l'erreur de l'enquête.

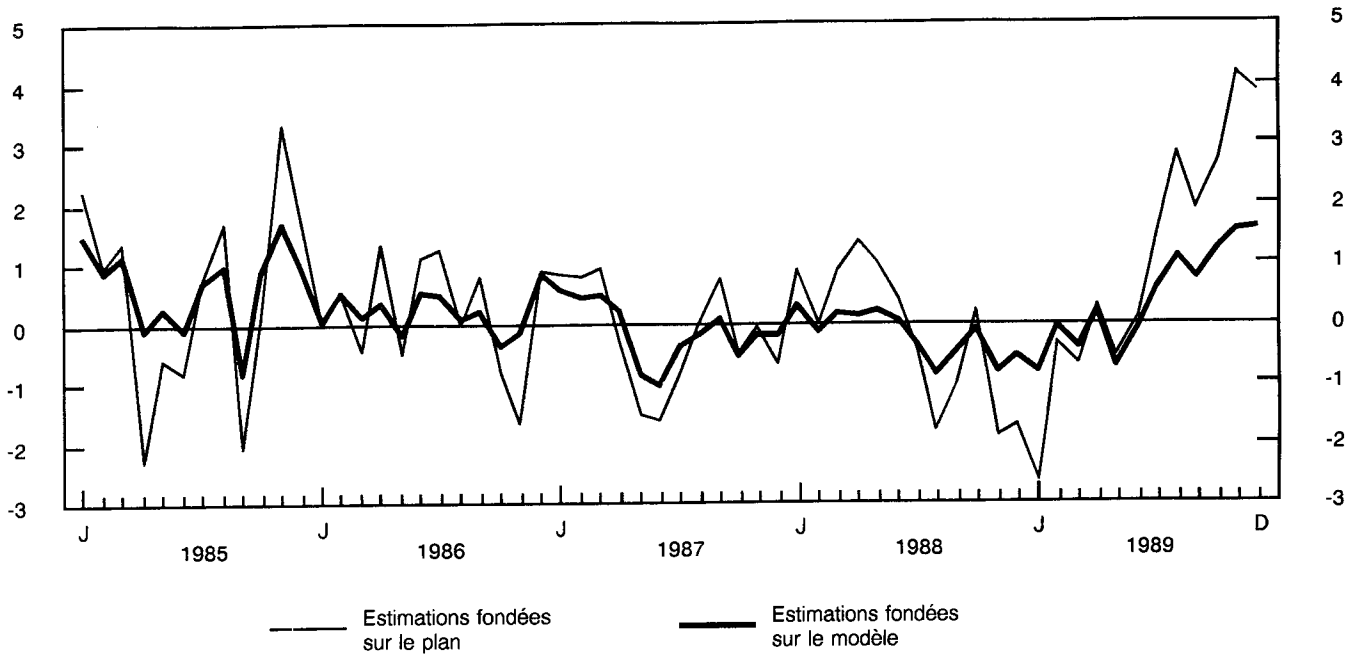

Figure 6. Variations annuelles des estimations fondées sur le plan et des estimations fondées sur le modèle des taux de chômage de l'Î.-P.-É. (× 100)

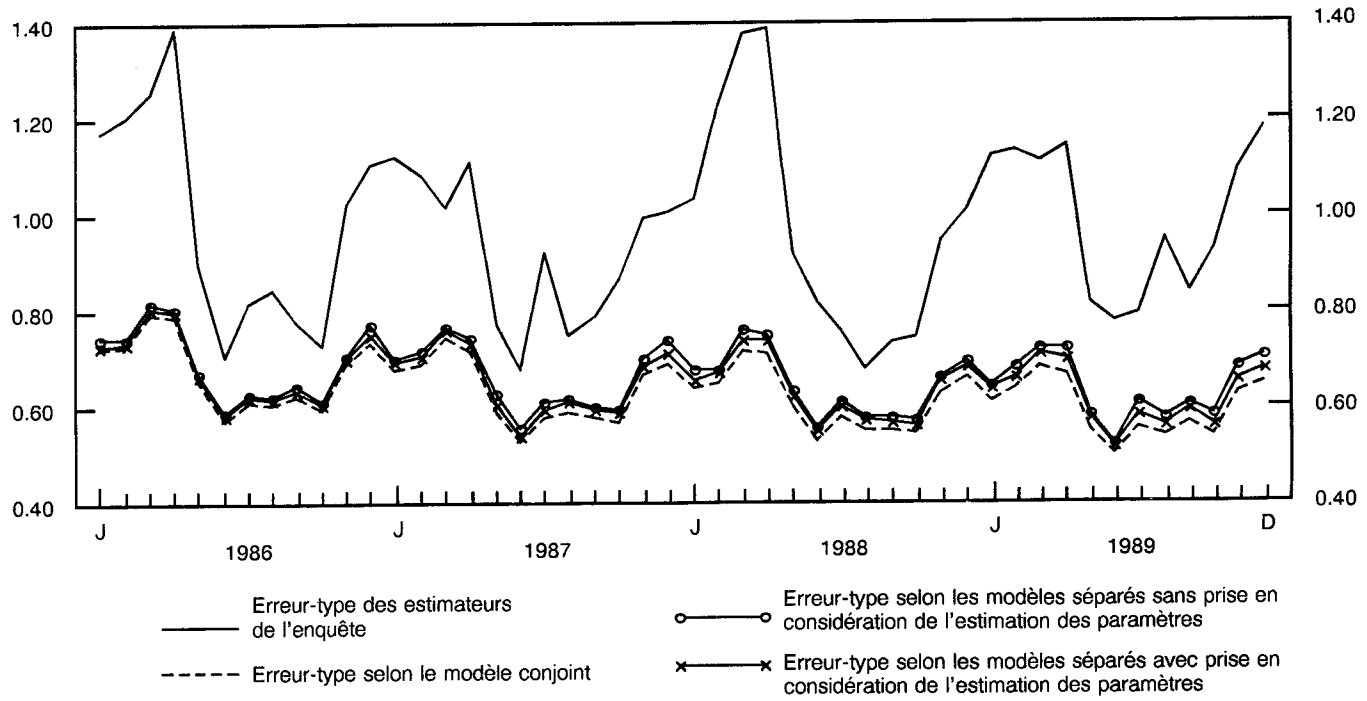

Figure 7. Erreurs-types des estimateurs de l'enquête et des estimateurs fondés sur le modèle avec et sans prise en considération de l'estimation des paramètres (× 100) – Île-du-Prince-Édouard

Le fait de laisser varier dans le temps les biais liés aux groupes de renouvellement est une extension naturelle du modèle, si l'on tient compte du fait que les moyennes des valeurs de la population varient en fonction du temps. Toutefois, la modélisation de l'évolution des biais liés aux groupes pourrait poser des difficultés, en raison d'éventuels problèmes d'identifiabilité touchant les modèles représentant la tendance et les effets saisonniers. Voir l'analyse présentée dans Pfeffermann (1991).

Les deux dernières extensions sont importantes et méritent d'être étudiées, mais notre expérience des données sur l'emploi nous porte à croire qu'elles auront un effet très minime sur les estimateurs du modèle.

REMERCIEMENTS

Les auteurs désirent remercier les arbitres pour leurs commentaires et leurs suggestions utiles. Les travaux relatifs à cette étude ont été réalisés pendant que le premier auteur se trouvait à Statistique Canada dans le cadre de son programme de bourse de recherche.

BIBLIOGRAPHIE

ANSLEY, C.F., et KOHN, R. (1986). Predicted mean square error for state-space models with estimated parameters. *Biometrika*, 73, 467-473.

BATTESE, G.E., HARTER, R.M., et FULLER, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

CHANG, I., TIAO, G.C., et CHEN, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30, 193-204.

DAGUM, E.B. (1980). *La méthode de désaisonnalisation* X-11 *ARMMI*. N° 12-564 au catalogue, Statistique Canada, Ottawa, Ontario, K1A 0T6.

HAMILTON, J.D. (1986). A standard error for the estimated state vector of a state-space model. *Journal of Econometrics*, 387-397.

HARRISON, P.J., et STEVENS, C.F. (1976). Bayesian forecasting (avec discussion). *Journal of the Royal Statistical Society*, Séries B, 38, 205-247.

HARVEY, A.C. (1984). A unified view of statistical forecasting procedures (avec discussion). *Journal of Forecasting*, 3, 245-275.

HARVEY, A.C., et TODD, P.H.J. (1983). Forecasting economic time series with structural and Box-Jenkins models (avec discussion). *Journal of Business and Economic Statistics*, 1, 299-315.

LEE, H. (1990). Estimation des coefficients de corrélation de panel pour l'Enquête sur la population active du Canada. *Techniques d'enquête*, 16, 297-306.

MARAVALL, A. (1985). On structural time series models and the characterization of components. *Journal of Business and Economic Statistics*, 3, 350-355.

MORRIS, N.D., et PFEFFERMANN, D. (1984). A Kalman filter approach to the forecasting of monthly time series affected by moving festivals. *Journal of Time Series*, 5, 255-268.

PFEFFERMANN, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business and Economic Statistics*, 9, 163-175.

PFEFFERMANN, D., et BURCK, L. (1990). Estimation robuste pour petits domaines par la combinaison de données chronologiques et traversales. *Techniques d'enquête*, 16, 229-249.

PFEFFERMANN, D., et BARNARD, C.M. (1991). Some new estimators for small-area means with application to the assessment of farmland values. *Journal of Business and Economic Statistics*, 9, 73-84.

SINGH, M.P., DREW, J.D., GAMBINO, J.G., et MAYDA, F. (1990). Méthodologie de l'enquête sur la population active du Canada. N° 71-526 au catalogue, Statistique Canada, Ottawa, Ontario, K1A 0T6.

TILLER, R.B. (1992). Time series modeling of sample survey data from the U.S. current population survey. *Journal of Official Statistics*, 8, 149-166.