

# Comparaison d'autres estimateurs pour l'Enquête sur la population active

Philip Bell<sup>1</sup>

## Résumé

L'auteur examine un choix d'estimateurs applicables à une enquête auprès des ménages périodique comportant un chevauchement contrôlé entre les enquêtes successives. L'auteur décrit la façon d'améliorer le MELSB (meilleur estimateur linéaire sans biais) fondé sur une fenêtre fixe de moments en appliquant la technique de la régression généralisée. Cet estimateur amélioré est comparé à l'estimateur *AK* de Gurney et Daly (1965) et à l'estimateur de régression modifiée de Singh, Kennedy, Wu et Brisebois (1997), à l'aide de données de l'enquête sur la population active de l'Australie.

Mots clés : Estimateur composite; meilleur estimateur linéaire sans biais; régression modifiée; enquêtes répétées.

## 1. Introduction

L'auteur examine un choix d'estimateurs applicables à une enquête-ménage périodique comportant un chevauchement contrôlé entre les enquêtes successives. Le thème commun des estimateurs est le recours à des données de moments antérieurs afin d'améliorer les estimations courantes, en bénéficiant de corrélations dans l'échantillon chevauchant. L'auteur considère tous les estimateurs de ce genre comme des estimateurs composites.

Les estimateurs sont évalués en fonction de l'enquête sur la population active (EPA) de l'Australie. Dans l'EPA, on contrôle le chevauchement en divisant l'échantillon du premier degré (régions géographiques) en huit « groupes de renouvellement » parmi lesquels on sélectionne des logements. Chaque mois, les mêmes logements sont tirés de sept des groupes de renouvellement, de nouveaux logements étant sélectionnés à même le groupe restant. L'échantillon est constitué de personnes âgées de 15 ans ou plus demeurant dans les logements sélectionnés.

Ce plan d'échantillonnage entraîne un chevauchement élevé de l'échantillon entre deux mois consécutifs au sein des sept « groupes de renouvellement appariés ». Le fait d'utiliser uniquement des données de ces groupes de renouvellement plutôt que de l'échantillon tout entier permet de diminuer l'erreur d'échantillonnage pour une estimation du mouvement d'un mois à l'autre. À l'aide de techniques d'estimation composite, on peut exploiter cette situation de façon à préparer des estimations comportant une erreur d'échantillonnage plus faible.

À la section 2, l'auteur présente l'EPA de l'Australie et son estimateur courant de « régression généralisée ». L'auteur aborde également la question du biais d'accoutumance (appelé biais de groupe de renouvellement par Bailar 1975).

À la section 3, l'auteur présente l'estimateur « composite *AK* » proposé par Gurney et Daly (1965). Cette méthode est

utilisée dans la Current Population Survey des États-Unis depuis de nombreuses années. Un prolongement appelé « pondération composite *AK* » est utilisé depuis quelques années; il a été proposé par Fuller (1990) et étudié par Lent, Miller et Cantwell (1994, 1996).

À la section 4, l'auteur présente la méthode « de régression modifiée » de l'estimation composite (Singh et Merkouris 1995; Singh 1996). Dans le présent exposé, l'auteur examine l'estimateur MR2 de Singh et coll. (1997), qui permet la plus forte réduction de l'erreur d'échantillonnage. L'auteur présente également une variante de cette méthode suggérée par Fuller (1999) en vue de l'Enquête sur la population active du Canada.

À la section 5, l'auteur présente un « meilleur estimateur linéaire sans biais » (MELSB) fondé sur des données tirées d'une « fenêtre » comportant un nombre fixe de mois successifs. Cet estimateur a d'abord été présenté par Jessen (1942) dans le cas de 2 occasions. Un MELSB fondé sur toutes les occasions d'une longue série semble peu pratique, bien que Yansaneh et Fuller (1998) en aient élaboré une approximation récursive. L'auteur améliore le MELSB à fenêtre fixe décrit par Bell (1998) à l'aide de la technique de régression généralisée.

À la section 6, on trouve les résultats de l'application des différentes méthodes à l'estimation des personnes ayant un emploi et des personnes sans emploi dans l'EPA. Les erreurs-types sont estimées pour des indicateurs à plus long terme comme la tendance et le mouvement de la tendance, de même que pour des estimations du niveau mensuel et de son mouvement. L'auteur examine les biais éventuels, de même que les indications d'un changement de profil saisonnier.

L'auteur termine l'exposé en comparant les avantages et les inconvénients des différents types d'estimateurs en fonction de l'EPA. L'estimateur MELSB amélioré est jugé efficace et, lorsqu'il est appliqué à l'EPA, il n'est pas exposé à un biais appréciable.

1. Philip Bell, Bureau of Statistics de l'Australie. Courriel : philip.bell@abs.gov.au.

## 2. Estimateurs courants pour l'Enquête sur la population active

### 2.1 Aperçu de l'EPA

L'EPA comporte un plan d'échantillonnage à plusieurs degrés, le premier étant un échantillon de petites régions géographiques appelées CD (« Census collector's Districts »). On tire un nouvel échantillon de CD tous les cinq ans, et les CD sont classés dans huit « groupes de renouvellement ». Les logements tirés d'un CD restent dans l'échantillon durant huit enquêtes, puis ils sont remplacés par d'autres logements tirés du même CD. On appelle renouvellement ce remplacement des logements, tous les logements d'un groupe de renouvellement étant remplacés en même temps. Des intervieweurs cherchent à recueillir des données pour toutes les personnes admissibles dans les logements sélectionnés.

Dans l'EPA, la situation d'activité de la personne (ayant un emploi, sans emploi, ne relevant pas de la population active) revêt un intérêt particulier. Le nombre de personnes se rapportant à chaque situation d'activité, pour diverses catégories de personnes, est un élément clé qu'il s'agit d'estimer dans l'enquête. Les nombreux utilisateurs des données d'enquête accordent une importance encore plus grande aux estimations du mouvement des chiffres d'un moment au suivant. On peut affirmer que les indications à plus long terme de la direction de la série sont encore plus importantes (par exemple le mouvement de la tendance X11 ou d'une tendance semblable plus lisse (Bell 1999)).

Le plan d'échantillonnage garantit que la probabilité inconditionnelle de sélection  $\pi_{t,i}$  est connue pour chaque personne échantillonnée  $i$  à un moment  $t$ . Cela permet d'obtenir un estimateur simple pour un total de population dû à Horvitz et Thompson (1952). Si  $Y_t$  est l'élément de population qu'il s'agit d'estimer à un moment  $t$ , et si  $y_{it}$  est le même élément signalé par la  $i^{\text{ième}}$  unité à un moment  $t$ , l'estimateur de Horvitz-Thompson est

$$\hat{y}_t^H = \sum_i w_{it}^\pi y_{it} \quad (1)$$

pour  $w_{it}^\pi = \pi_{it}^{-1}$ , connus comme les poids de sélection.

### 2.2 L'estimateur de régression généralisée (RG)

La régression généralisée est une méthode de rajustement ou de « calage » d'une série de poids unitaires dont l'addition donne une série d'attributs de la population appelés repères. Pour un choix approprié de repères, les poids résultants donnent une meilleure estimation en tenant compte de renseignements externes.

Dans l'EPA, nous commençons par un calage des poids de Horvitz-Thompson de façon que leur addition donne des repères démographiques indiquant le nombre de personnes dans la population pour 560 strates après sélection (14 régions géographiques classées selon le sexe et 20 groupes d'âge). Les poids d'une strate après sélection particulière

sont répartis proportionnellement de façon que leur addition donne le repère de strate. Cet estimateur par quotient stratifié après sélection est un cas particulier de l'estimateur de régression généralisée ou estimateur RG.

Soit  $x_{it}$  un vecteur de ligne de variables auxiliaires pour l'unité  $i$  à un moment  $t$ , et  $\hat{x}_t = \sum_i b_{it} x_{it}$  des estimations du vecteur de ligne correspondant de valeurs repères  $X_t$ , d'après certains poids initiaux  $b_{it}$ . L'estimateur RG fondé sur ces poids initiaux est alors donné par

$$\hat{y}_t^G = \hat{y}_t + (X_t - \hat{x}_t) \hat{\beta} \quad (2)$$

$$\text{pour } \hat{\beta} = \left( \sum_i b_{it} x'_{it} x_{it} \right)^{-1} \sum_i b_{it} x'_{it} y_{it}. \quad (3)$$

Donc,  $\hat{y}_t^G = \sum_i w_{it}^G y_{it}$  pour

$$w_{it}^G = b_{it} \left( 1 + (X_t - \hat{x}_t) \left( \sum_i b_{it} x'_{it} x_{it} \right)^{-1} x'_{it} \right). \quad (4)$$

Dans une estimation par quotient stratifiée après sélection, les vecteurs de ligne  $x_{it}$  contiennent des zéros sauf dans la colonne qui correspond à la strate après sélection de l'unité, et  $b_{it}$  sont les poids de sélection  $w_{it}^\pi$ . Dans ce cas, les paramètres de régression sont simplement les moyennes de strates après sélection, estimées à l'aide des poids de sélection.

### 2.3 Estimations par groupe de renouvellement

Chaque groupe de renouvellement est constitué d'un échantillon représentatif de logements, et peut donc fournir une estimation distincte. On numérote les groupes de renouvellement à tel moment selon le nombre de fois que les logements du groupe de renouvellement ont été échantillonnés. On écrit  $R(t, i) = r$  si l'unité  $i$  se trouve dans le groupe de renouvellement échantillonné pour la  $r^{\text{ième}}$  fois à un moment  $t$ . L'estimation de Horvitz-Thompson de  $Y_t$  d'après le groupe de renouvellement  $r$  est

$$\hat{y}_t^{Hr} = \sum_{i:R(t,i)=r} \delta w_{it}^\pi y_{it}. \quad (5)$$

On peut utiliser la régression généralisée pour améliorer ces estimateurs en calant les poids de façon que leur addition donne une série de repères. Malheureusement, la taille plus faible de l'échantillon dans un groupe de renouvellements unique risque d'exiger l'utilisation d'un plus petit nombre de repères que pour le cas global. Dans la situation de l'EPA, l'auteur du présent exposé a appliqué une même étape de régression généralisée à l'ensemble de l'échantillon de façon que, pour tout l'échantillon, l'addition des poids donne les repères pour les 540 strates après sélection courantes, tandis que, dans chaque groupe de renouvellement, l'addition des poids donne un huitième des repères pour 71 strates après sélection groupées. Les poids résultants, lorsqu'ils sont appliqués à un groupe de renouvellements donné  $r$  et multipliés par huit, donnent les estimations de groupe de renouvellement  $\hat{y}_t^{Rr}$ .

## 2.4 Biais d'accoutumance

Idéalement, les estimations de groupe de renouvellement devraient comporter la même espérance  $Y_t$ , mais dans la pratique elles comportent des espérances légèrement différentes et donc un biais différent. La différence est due en partie à des pratiques de collecte; par exemple, les logements échantillonnés pour la première fois sont interviewés dans le cadre d'une visite sur place, tandis que d'autres groupes de renouvellement sont surtout interviewés par téléphone. Il n'est pas clair quel groupe de renouvellement est influencé le moins par ce genre de biais « d'accoutumance ». L'estimation globale comporte un biais d'accoutumance qui est un mélange des biais de chaque groupe de renouvellement. Nous avons recours à de bonnes techniques d'enquête pour réduire ce biais au minimum. À noter que tous les estimateurs composites recevront un apport différent des groupes de renouvellement, et comporteront donc différents biais d'accoutumance.

## 3. Estimation composite $AK$

### 3.1 Estimateur composite $AK$

L'estimateur composite  $AK$  (Gurney et Daly 1965) est conçu de façon à insister davantage sur le mouvement des groupes de renouvellement appariés (c'est-à-dire les groupes de renouvellement dans lesquels les mêmes logements ont été sélectionnés pour le mois courant et le mois précédent). L'estimateur compte trois composantes. La première est une moyenne des estimations de groupe de renouvellement pour les données du mois courant (moment  $t$ ). La deuxième est l'estimateur composite  $AK$  du mois précédent, plus une estimation de mouvement fondée uniquement sur les groupes de renouvellement appariés. La troisième composante est la différence entre les estimations du groupe de renouvellement non apparié et des groupes appariés. La part de chaque composante qu'il convient d'utiliser est donnée par deux paramètres  $A$  et  $K$ , comme suit :

$$\begin{aligned} \hat{y}_t^{AK} = & (1-K) \frac{1}{8} \sum_{r=1}^8 \hat{y}_t^{Rr} \\ & + K \left( \hat{y}_{t-1}^{AK} + \frac{1}{7} \sum_{r=2}^8 \hat{y}_t^{Rr} - \frac{1}{7} \sum_{r=1}^7 \hat{y}_{t-1}^{Rr} \right) \\ & + A \left( \hat{y}_t^{R1} - \frac{1}{7} \sum_{r=2}^8 \hat{y}_t^{Rr} \right). \end{aligned} \quad (6)$$

### 3.2 Choix des valeurs paramétriques

Le paramètre clé est  $K$ , qui indique la part de l'estimation nouvelle qui se fonde sur le mouvement des groupes de renouvellement appariés. La valeur optimale de  $A$  et de  $K$  à utiliser dépend de la variable qu'il s'agit d'estimer. Des valeurs  $K$  plus élevées se prêtent davantage à l'emploi qu'au

chômage, puisque l'emploi comporte une plus forte corrélation d'un mois à l'autre.

L'addition des estimations composites  $AK$  des personnes ayant un emploi, sans emploi et « ne relevant pas de la population active » ne donne pas exactement la population totale à moins que l'on n'utilise les mêmes paramètres pour toutes les estimations. On choisit donc un compromis pour  $A$  et  $K$ . Les résultats présentés ici se fondent sur  $A = 0,06$  et  $K = 0,7$ . On a trouvé ces valeurs en faisant l'essai d'une gamme de valeurs de  $A$  et de  $K$  et en choisissant les valeurs qui donnent des estimations optimales des personnes ayant un emploi. Dans la présente étude, aucune valeur de  $A$  et de  $K$  n'a donné des estimations des personnes sans emploi appréciablement meilleures que ces valeurs.

Notre étude empirique n'a pas manifesté un niveau particulièrement bon d'erreurs d'échantillonnage pour l'estimateur  $AK$ . Cela est peut-être attribuable au calage fin qui a servi à obtenir les estimations de groupe de renouvellement; il est possible que le recours à des catégories plus vastes améliore le bilan des erreurs d'échantillonnage.

### 3.3 Propriétés de l'estimateur $AK$

L'estimateur  $AK$  insiste davantage sur le mouvement au sein des groupes de renouvellement appariés. Ainsi, le groupe de renouvellement qui contient des logements admissibles pour la première fois contribue moins que dans l'estimateur RG. L'estimateur  $AK$  comporte donc un biais d'accoutumance différent de celui de l'estimateur RG.

L'estimateur  $AK$  est récursif : il faut l'estimateur du mois précédent pour préparer celui du mois courant. Cela est peu pratique lorsqu'il s'agit de préparer des estimations pour un nouvel élément ou une nouvelle catégorie. De plus, le besoin d'utiliser les mêmes valeurs de  $A$  et de  $K$  pour tous les éléments peut entraîner des estimations sous-optimales pour un élément donné.

À cause de ces préoccupations, la Current Population Survey des États-Unis a été changée en fonction d'une variante appelée « pondération composite  $AK$  » (Lent, Miller et Cantwell 1994). Dans la pondération composite  $AK$ , on prépare des estimations distinctes de l'emploi et du chômage pour un certain nombre de catégories publiées importantes, à l'aide de la stratégie composite  $AK$  avec des paramètres optimaux pour l'estimation en question. Les données courantes sont alors calées de façon que l'addition des poids unitaires donne ces estimations  $AK$  de même que des repères démographiques. Toutes les estimations sont alors préparées à partir du fichier de données courant à l'aide de ces nouveaux « poids composites  $AK$  ».

Le fait de pouvoir préparer toutes les estimations sous forme de somme pondérée des données d'un même mois est un grand avantage de la stratégie de pondération composite  $AK$ . Un autre avantage est que les plus importantes estimations sont des estimations composites  $AK$  avec un choix de  $AK$  presque optimal. Un inconvénient est que seules les plus importantes estimations sont de véritables estimations composites. Toutes les autres estimations (y

compris les estimations de personnes ne relevant pas de la population active) ne sont typiquement pas très améliorées relativement aux estimations RG standard (Lent, Miller et Cantwell 1996).

#### 4. Estimation de régression modifiée

##### 4.1 Aperçu de la régression modifiée

La méthode de régression modifiée est une autre façon de fournir des estimations composites que l'on peut obtenir comme agrégats pondérés du fichier de données d'enquête courant. La méthode s'applique à un ensemble d'éléments clés déterminé à l'avance, pour lequel elle réalise un nombre particulièrement faible d'erreurs d'échantillonnage.

La technique de régression modifiée fait appel à une régression généralisée pour le fichier de données du mois courant après avoir annexé de nouvelles variables auxiliaires  $z_{it}$  à chaque unité  $i$  à un moment  $t$ . Ici  $z_{it}$  est un vecteur de ligne comportant un élément pour chacun des éléments clés. Pour ceux-ci, on a des « pseudo-repères »  $Z_t$  fondés sur les estimations du mois précédent pour les éléments clés. L'estimateur de régression modifiée est alors donné par une étape de régression généralisée qui applique les repères démographiques aussi bien que les pseudo-repères.

$$\hat{y}_t^M = \hat{y}_t^H + ((X_t, Z_t) - (\hat{x}_t^H, \hat{z}_t^H)) \beta_t^M \quad (7)$$

pour

$$\beta_t^M = \left( \sum_i w_{it}^\pi (x_{it}, z_{it})' (x_{it}, z_{it}) \right)^{-1} \sum_i w_{it}^\pi (x_{it}, z_{it})' y_{it} \quad (8)$$

donc,  $\hat{y}_t^M = \sum_i w_{it}^M y_{it}$  pour

$$w_{it}^M = w_{it}^\pi \left\{ \frac{1 + ((X_t, Z_t) - (\hat{x}_t^H, \hat{z}_t^H))}{\left( \sum_i w_{it}^\pi (x_{it}, z_{it})' (x_{it}, z_{it}) \right)^{-1} (x_{it}, z_{it})'} \right\} \quad (9)$$

La clé de la méthode est la définition des variables auxiliaires. Soit  $D$ , l'ensemble des unités des groupes de renouvellement appariés (ceux comportant des logements sélectionnés aux deux moments) à un moment  $t$ . Soit  $y_{it}^*$ , le vecteur d'éléments clés pour l'unité  $i$  à un moment  $t$  et  $Y_t^*$ , les totaux de population correspondants. Pour  $i \in D$ , notons  $y_{t-1,i}^*$ , la valeur du mois précédent pour le vecteur d'éléments clés ou, si aucune valeur n'a été déclarée, imputons  $y_{t-1,i}^* - j$ 'ai utilisé  $y_{t-1,i}^* = y_{t-1,i}^*$  comme l'a suggéré Singh (1996).

J'examine des estimations de régression modifiée pour  $z_{it}$  du type ci-dessous, pour  $a \in [0, 1]$ :

$$z_{it} = (1-a) \frac{8}{7} y_{t-1,i}^* + a \begin{pmatrix} y_{it}^* \\ -\frac{8}{7} (y_{it}^* - y_{t-1,i}^*) \end{pmatrix} \text{ pour } i \in D$$

$$= a y_{it}^* \text{ pour } i \notin D. \quad (10)$$

Compte tenu de cette définition, nous avons

$$\hat{z}_t^H = (1-a) \hat{y}_{t-1}^{+HD} + a (\hat{y}_t^{*H} - (\hat{y}_t^{*HD} - \hat{y}_{t-1}^{+HD})), \quad (11)$$

où  $\hat{y}_{t-1}^{+HD} = 8/7 \sum_{i \in D} w_{t,i}^* y_{t-1,i}^\pi$  et  $\hat{y}_t^{*HD} = 8/7 \sum_{i \in D} w_{t,i}^\pi y_{t,i}^*$  sont des estimations de  $Y_{t-1}^*$  et de  $Y_t^*$  respectivement en fonction d'unités en  $D$  seulement et des poids de sélection du mois courant. Pour  $a=0$ ,  $\hat{z}_t^H$  est simplement l'estimation  $\hat{y}_{t-1}^{+HD}$ . Pour  $a=1$ ,  $\hat{z}_t^H$  est l'estimation de Horvitz-Thompson du mois courant moins une estimation de mouvement fondée sur les groupes de renouvellement appariés  $\hat{y}_t^{*HD} - \hat{y}_{t-1}^{+HD}$ . Les valeurs  $a=0$  et  $a=1$  donnent les méthodes MR1 et MR2 respectivement de Singh et coll. (1997). L'utilisation d'une valeur  $a$  intermédiaire a été suggérée par Fuller (1999).

Un pseudo-repère  $Z_t$  approprié serait une estimation de  $Y_{t-1}^*$  rajustée de façon à correspondre aux poids du mois courant. Suivant Singh et coll. (1997), j'ai utilisé une étape de régression généralisée afin de rajuster l'estimateur de régression modifiée du mois précédent pour que l'addition donne les repères du mois courant :

$$Z_t = \hat{y}_{t-1}^{*M} + (X_t - \hat{x}_{t-1}^M) \beta_t^{\text{adj}} \quad (12)$$

$$\text{pour } \beta_t^{\text{adj}} = \left( \sum_i w_{t-1,i}^M x_{t-1,i}' x_{t-1,i} \right)^{-1} \sum_i w_{t-1,i}^M x_{t-1,i}' y_{t-1,i}^* \quad (13)$$

À noter que  $Z_t \approx \hat{y}_{t-1}^{*M}$  puisque  $\hat{x}_{t-1}^M = X_{t-1} \approx X_t$ . Cela termine la définition des estimateurs de régression modifiée.

##### 4.2 Propriétés des estimateurs de régression modifiée

Le mouvement  $\hat{y}_t^{*HD} - \hat{y}_{t-1}^{+HD}$  en (11) se fonde en réalité uniquement sur l'échantillon apparié (c'est-à-dire les unités qui déclarent aux deux moments), puisque les autres unités des groupes de renouvellement appariés  $D$  contribuent zéro au mouvement (pour l'imputation utilisée ici). Il peut en résulter des estimateurs de régression modifiée ayant une erreur d'échantillonnage moins élevée qu'un estimateur  $AK$ , car ce « mouvement de l'échantillon apparié » n'est pas influencé par les unités qui ne sont pas présentes pour les deux mois.

Malheureusement, cela entraîne également une possibilité de biais lorsque des personnes qui ne sont pas représentées dans l'échantillon apparié manifestent un comportement différent de celui des personnes de

l'échantillon apparié. Ce peut être effectivement le cas : l'échantillon apparié exclut les personnes qui ont changé de logement entre les deux mois, et il est possible que le changement de logement soit lié à un changement d'emploi. Ce « biais de l'échantillon apparié » vient s'ajouter à tout biais d'accoutumance.

Une autre difficulté est liée à l'estimateur MR2 (donc,  $a=1$ ). Si la  $k^{\text{ième}}$  variable clé  $y_{i,k}^*$  comporte une forte corrélation d'un mois à l'autre, elle comportera également une forte corrélation avec la  $k^{\text{ième}}$  nouvelle variable auxiliaire  $z_{i,k}$ . Pour une telle variable, l'élément de  $\beta_i^M$  qui correspond à  $z_{i,k}$  aura une certaine valeur  $y_t$  proche de un. Si l'on utilise (7), (11) et  $Z_t \approx \hat{y}_{t-1}^{*M}$ , l'estimateur MR2 adopte la forme

$$\hat{y}_{t,k}^{*M} \approx (1 - y_t) \hat{y}_{t,k}^{*H} + y_t (\hat{y}_{t-1,k}^{*M} + (\hat{y}_{t,k}^{*HD} - \hat{y}_{t-1,k}^{*HD})) + \text{autres termes.} \quad (14)$$

Dans ce cas, il est possible que le mouvement de l'échantillon apparié à un moment donné influence fortement les estimations lors de nombreux moments subséquents. De plus, un biais faible dans le mouvement aura tendance à s'accumuler avec le temps. Ce danger a été reconnu par Fuller (1999) et qualifié de « problème de dérive ». C'est ce qui l'a poussé à suggérer la forme d'estimateur donnée ici, avec une valeur de  $a$  inférieure à 1.

En résumé, la régression modifiée offre des avantages semblables à ceux de la stratégie de pondération composite  $AK$ , mais avec une erreur d'échantillonnage possiblement inférieure. La méthode n'est pas difficile à appliquer, et elle permet d'éviter la nécessité d'un calage distinct des groupes de renouvellement en fonction des pères.

## 5. Meilleure estimation linéaire sans biais (MELSB)

### 5.1 MELSB pour une fenêtre fixe

On obtient l'estimateur MELSB pour une fenêtre fixe (noté  $\hat{y}_t^B$ ) en choisissant une combinaison linéaire « optimale » des estimations de groupe de renouvellement  $\hat{y}_t^{Rr}$  (définie en 2.3) à partir d'une fenêtre de  $l+1$  mois, comme suit :

$$\hat{y}_t^B = \sum_{s=t-l}^t \sum_{r=1}^8 a_{sr} \hat{y}_s^{Rr} \quad (15)$$

où les paramètres  $a_{sr}$  sont choisis de façon à réduire ( $\hat{y}_t^B$ ) au minimum moyennant les contraintes  $\sum_{r=1}^8 a_{sr} = 1$  pour  $s=t$  et  $\sum_{r=1}^8 a_{sr} = 0$  pour  $s=t-l, \dots, t-1$ . Ces contraintes garantissent que  $\hat{y}_t^B$  sera sans biais pour  $Y_t$ , à la condition que les estimations de groupe de renouvellement soient sans biais, c'est-à-dire que  $E(\hat{y}_s^{Gr}) = Y_s$  pour  $s = t-l, \dots, t$ .

La réduction au minimum suppose que les variances et les covariances des estimations de groupe de renouvellement soient connues. Dans la pratique, celles-ci sont estimées en fonction de données chronologiques. Le problème se laisse alors écrire sous une forme matricielle : nous cherchons à choisir le vecteur de colonne  $a$  (avec les éléments  $a_{sr}$  pour  $s=t-l, \dots, t$  et  $r=1, \dots, 8$ ) de façon à réduire une forme quadratique  $a'Va$  à au minimum sous réserve des contraintes  $C'a=c$ . Le résultat standard pertinent (Rao 1973, page 65) est que le minimum survient pour  $a = V^{-1}Cq$  où  $q$  est une solution de  $(C'V^{-1}C)q = c$ . Dans la présente étude, la matrice  $V$  a été remplacée par une matrice de corrélation, suivant l'hypothèse selon laquelle toutes les estimations de groupe de renouvellement dans la fenêtre comportaient la même variance.

### 5.2 Structure de corrélation des estimations de groupe de renouvellement

Puisque différents modes de corrélation donnent diverses estimations MELSB, le choix d'un profil de corrélation comporte des aspects semblables à ceux du choix des paramètres  $A$  et  $K$  pour la stratégie composite  $AK$ . Il est souhaitable d'utiliser la même combinaison linéaire pour toutes les estimations afin de garantir l'additivité des estimations.

J'ai adopté un modèle à quatre paramètres pour le mode de corrélation :

$$\begin{aligned} \text{corr}(\hat{y}_t^{Gr}, \hat{y}_s^{Gr'}) &= \rho_{|t-s|}^W \quad \text{pour } r-r' = t-s \\ &= \rho_{|t-s|}^B \quad \text{pour } r-r' = t-s+8m \\ &\quad \text{pour le nombre entier } m \neq 0 \\ &= 0 \quad \text{autrement.} \end{aligned} \quad (16)$$

Ainsi, la corrélation entre les estimations pour un décalage  $k$  du même groupe de renouvellement est  $\rho_k^W$  si le groupe de renouvellement comporte les mêmes logements aux deux moments, et  $\rho_k^B$  autrement. Il n'y a pas de corrélation entre les estimations de différents groupes de renouvellement. Un modèle à quatre paramètres est utilisé :

$$\rho_k^W = (1 - r_U^2) (\theta_P^k r_P^2 + \theta_B^k (1 - r_P^2)) \quad (17)$$

et

$$\rho_k^B = (1 - r_U^2) \theta_B^k (1 - r_P^2). \quad (18)$$

Bell et Carolan (1998) ont discuté ce modèle. Les valeurs paramétriques utilisées dans cet exposé étaient  $\theta_P = 0,87697$ ,  $\theta_B = 0,94$ ,  $r_U = 0,3101$  et  $r_P = 0,90456$ . Ces valeurs sont le résultat d'un ajustement du modèle en fonction d'autocorrélations estimatives pour les estimations de groupe de renouvellement de la proportion d'employés.

Il importe de noter que les estimations de type MELSB sont sans biais peu importe l'exactitude du modèle de corrélation supposé. Le modèle utilisé ici cherche à être optimal pour des estimations des personnes ayant un

emploi, mais fonctionne bien aussi pour des personnes sans emploi. L'utilisation d'autres valeurs pour les paramètres du modèle n'a donné aucune amélioration appréciable de l'erreur-type pour les personnes sans emploi.

### 5.3 Estimations MELSB améliorées

Une difficulté liée aux estimations MELSB ci-dessus est qu'il faut des estimations RG au niveau du groupe de renouvellement. La taille plus petite de l'échantillon au niveau du groupe de renouvellement risque de limiter les repères utilisables, comme on l'a vu pour l'AK. Pour le MELSB, toutefois, une autre stratégie est possible.

On définit l'estimateur B1 en établissant un estimateur MELSB fondé sur les estimateurs de Horvitz-Thompson au niveau du groupe de renouvellement, puis en appliquant la technique de régression généralisée afin d'améliorer cet estimateur. Il s'agit de procéder comme suit. On définit  $y_{ii}^{\#} = a_{iR(t,i)} y_{ii}$  et  $x_{ii}^{\#} = a_{iR(t,i)} x_{ii}$ , où  $a_{iR(t,i)}$  est le multiplicateur MELSB applicable au groupe de renouvellement dans lequel se trouve l'unité  $i$  à un moment  $t$ . On peut alors écrire l'estimateur MELSB fondé sur les estimateurs de Horvitz-Thompson sous la forme

$$\hat{y}_t^{\text{BH}} = \sum_{s=t-l}^t \sum_i w_{si}^{\pi} y_{si}^{\#} \quad (19)$$

En calant en fonction des repères, on obtient l'estimateur MELSB amélioré B1 :

$$\hat{y}_t^{\text{B1}} = \hat{y}_t^{\text{BH}} + (X_t - \hat{x}_t^{\text{BH}}) \hat{\beta} \quad (20)$$

$$\text{pour } \hat{\beta} = \left( \sum_{s=t-l}^t \sum_i w_{si}^{\pi} x_{si}^{\#} x_{si}^{\#} \right)^{-1} \sum_{s=t-l}^t \sum_i w_{si}^{\pi} x_{si}^{\#} y_{si}^{\#} \quad (21)$$

$$\text{donc, } \hat{y}_t^{\text{B1}} = \sum_{s=t-l}^t \sum_i w_{si}^{\text{B1}} y_{si} \quad (22)$$

pour

$$w_{si}^{\text{B1}} = \left\{ 1 + (X_t - \hat{x}_t^{\text{BH}}) \left[ \sum_{u=s-l}^s \sum_j w_{uj}^{\pi} a_{uR(uj)} x'_{uj} x'_{uj} \right]^{-1} a_{sR(s,i)} x'_{si} \right\}^{-2} \quad (23)$$

### Propriétés des estimateurs MELSB et B1

Les estimations de type MELSB et B1 sont des sommes de données unitaires pondérées pour une fenêtre de mois. Chaque estimation n'exige que des données de cette fenêtre et se laisse préparer indépendamment des estimations pour les mois précédents, de sorte que la méthode n'est pas récursive.

Le même mois de données contribuera différents poids à l'estimation pour différents moments. Une unité contribuera un poids appréciable à son estimation du mois courant, et un poids proche de zéro, souvent négatif, aux estimations d'autres mois. Le travail de préparation d'un tableau est le

même que pour la RG multipliée par la taille de la fenêtre. Il existe également une possibilité d'estimations négatives pour de petites cellules ne comportant aucune unité courante.

À noter que, dans l'estimateur B1, on ne force pas l'addition des poids appliqués aux mois autres que le mois courant de donner une valeur de zéro. En vertu des hypothèses du modèle, l'estimation  $\hat{y}_t^{\text{B1}}$  demeure inconditionnellement sans biais, puisque  $\hat{y}_t^{\text{BH}}$  et  $\hat{x}_t^{\text{BH}}$  sont sans biais pour  $Y_t$  et  $X_t$  respectivement. Dans la pratique, le mois courant contribue 99,5 % environ du poids total. J'estime que le biais qui en résulte est faible et sans danger (entraînant un léger lissage des estimations avec le temps).

Pour toute estimation dans laquelle il existe une corrélation appréciable des données d'un mois à l'autre, les estimations MELSB et B1 devraient comporter une erreur d'échantillonnage moindre que l'estimation RG. C'est là un avantage théorique relativement à une méthode servant à améliorer un ensemble d'estimations déterminé à l'avance (comme la régression modifiée ou l'AK avec pondération composite). Dans la pratique, cet avantage n'est pas nécessairement très important, car pour l'EPA on s'intéresse surtout à un petit nombre d'estimations bien définies.

L'utilisateur doit également déterminer la période de temps ou la « fenêtre » des estimations. Le fait d'utiliser trop de moments suppose beaucoup de temps d'ordinateur, tandis qu'un nombre trop faible réduit les avantages. La fenêtre de sept mois utilisée ici a été suffisante pour offrir presque tous les avantages; des fenêtres plus petites comportent des erreurs-types appréciablement plus élevées.

## 6. Comparaison des méthodes

### 6.1 Méthode de comparaison

Les estimations pour la période allant de juillet 1993 à janvier 1999 ont été préparées en fonction de données allant de janvier 1993 à janvier 1999. On a obtenu des estimations classées selon le mois, l'État, le sexe, l'état matrimonial et la situation par rapport au marché du travail. On a également obtenu des estimations pour le mouvement de décalage un, la moyenne trimestrielle et le mouvement de moyennes trimestrielles successives.

L'erreur-type de ces estimations a été calculée à l'aide de la technique « jackknife avec suppression d'un groupe » (Kott 1998). Les unités géographiques qui constituent la première étape du tirage de l'échantillon ont été divisées systématiquement en  $G=30$  groupes, et « des groupes répétés » ont été établis en fonction de l'échantillon complet à l'exclusion des unités de l'un de ces groupes. Chaque estimation étudiée a également été préparée pour chacun des groupes répétés  $G$ . Lorsqu'on écrit  $e$  pour l'estimation et  $e_{(g)}$  pour l'estimation de la répétition  $g$ , l'estimation de type jackknife avec suppression d'un groupe de l'erreur-type est donnée par

$$SE_{(e)} = \sqrt{\frac{G-1}{G} \sum_{g=1}^G (e_{(g)-e})^2}. \quad (24)$$

On a obtenu les estimations et les erreurs-types pour chacun des estimateurs ci-dessous (énumérés en abrégé pour faciliter la suite de l'exposé) :

RG : L'estimation de régression généralisée utilisée couramment dans l'EPA

AK : L'estimation AK avec  $K = 0,7$ ,  $A = 0,06$

ME : L'estimateur MELSB fondé sur une fenêtre de 7 mois

B1 : Le MELSB amélioré fondé sur une fenêtre de 7 mois

MR2 : L'estimateur MR2 (régression modifiée avec  $a = 1$ )

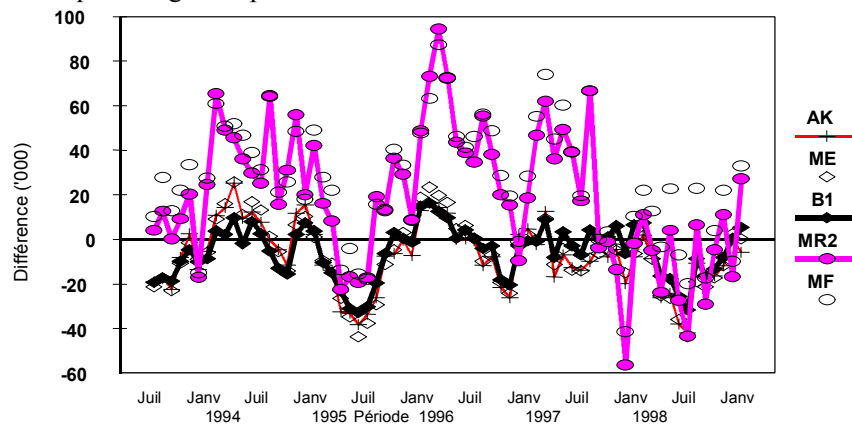
MF : La variante de Fuller de la régression modifiée ( $a = 0,7$ )

Les estimateurs de régression modifiée supposent un choix des variables clés qu'il s'agit d'optimiser. Pour

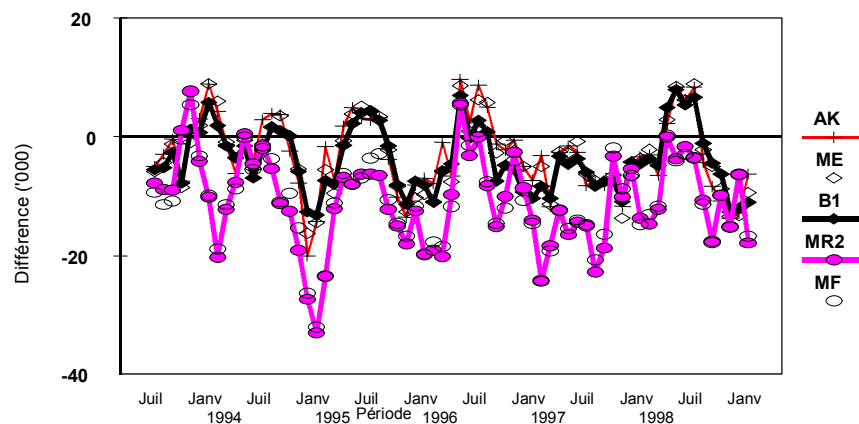
préparer les estimations de régression modifiée du présent rapport, des variables  $z$  ont été préparées pour les estimations de l'emploi et du chômage pour chaque État et chaque sexe. Cela donne au total 32 variables auxiliaires supplémentaires, en plus des 540 repères de strate après sélection habituels utilisés dans la régression généralisée.

## 6.2 Différences par rapport à l'estimation RG

On peut utiliser l'estimateur RG courant comme base de comparaison pour les autres estimateurs. Au lieu de présenter des diagrammes d'estimations de niveau, je montre les différences des autres estimations par rapport aux estimations RG courantes. Les diagrammes 1 et 2 indiquent ces différences pour les estimations des personnes ayant un emploi et des personnes sans emploi respectivement. Afin de situer l'ampleur de ces différences, notons que les erreurs-types publiées pour l'estimation courante étaient de 25 200 pour les personnes ayant un emploi et de 7 900 pour les personnes sans emploi en janvier 1999 (semblables pour les autres mois).



**Diagramme 1.** Différence des autres estimations par rapport à l'estimation RG, personnes ayant un emploi (milliers), juillet 1993 à janvier 1999.



**Diagramme 2.** Différence des autres estimations par rapport à l'estimation RG, personnes sans emploi (milliers), juillet 1993 à janvier 1999.

Les estimations AK, ME et B1 sont assez semblables, puisque, pour l'ensemble des trois méthodes, l'apport d'une unité dépend de son groupe de renouvellement. Dans les deux diagrammes, les estimateurs AK, ME et B1 semblent donner, en moyenne, des valeurs plus faibles que les estimations RG. Cela indique un changement du biais d'accoutumance, résultat du fait que l'on accorde un poids moindre au groupe de renouvellement échantillonné pour la première fois. Les estimations varient vers le haut et vers le bas relativement à leur différence moyenne pour de courtes périodes.

Les estimations MR2 et MF tendent à différer des autres estimations, puisqu'elles soulignent l'apport des unités de l'échantillon apparié. Pour les personnes ayant un emploi, les estimateurs MR2 et MF sont appréciablement plus grands en moyenne que les estimations RG, jusqu'à septembre 1997. Il y a alors une baisse des différences correspondant à l'intégration d'un nouvel échantillon à partir de septembre 1997. Pour des raisons qui ne sont pas claires, au cours de cette période l'échantillon apparié manifeste un comportement différent de celui de l'échantillon global. Cela influe sur la différence entre ces séries de régression modifiée et la série RG. Ce qui risque d'être

inquiétant est le fait que le changement de niveau influe sur le niveau de la série MR2 au cours d'une période subséquente considérable, manifestation possible du « problème de dérive ».

Pour les personnes sans emploi, les estimations MR2 et MF tendent à être inférieures aux estimations RG. Il n'y a aucune indication d'un « problème de dérive » pour les personnes sans emploi, ce qui n'a rien de surprenant compte tenu des corrélations moindres mises en jeu.

### 6.3 Différences moyennes selon le mois civil

Afin de quantifier le changement probable de biais suivant le passage à un nouvel estimateur, on a calculé la différence moyenne au cours de la période de chaque estimation relativement à l'estimation RG. Il est possible que la différence soit saisonnière, et l'on a donc obtenu des moyennes distinctes pour chaque mois de l'année civile, de même que pour l'ensemble. On trouvera dans le diagramme 3 les différences moyennes au cours de la période allant de juillet 1993 à janvier 1999 pour les personnes ayant un emploi.

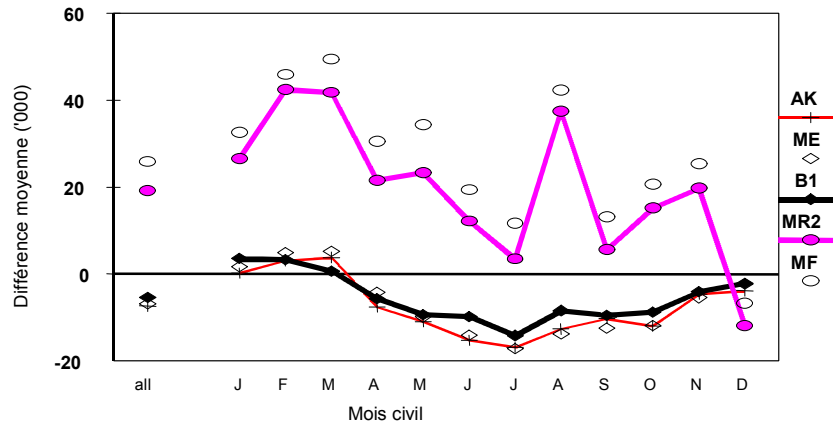


Diagramme 3. Différence moyenne par rapport à l'estimation RG, globalement et selon le mois civil, personnes ayant un emploi (milliers).

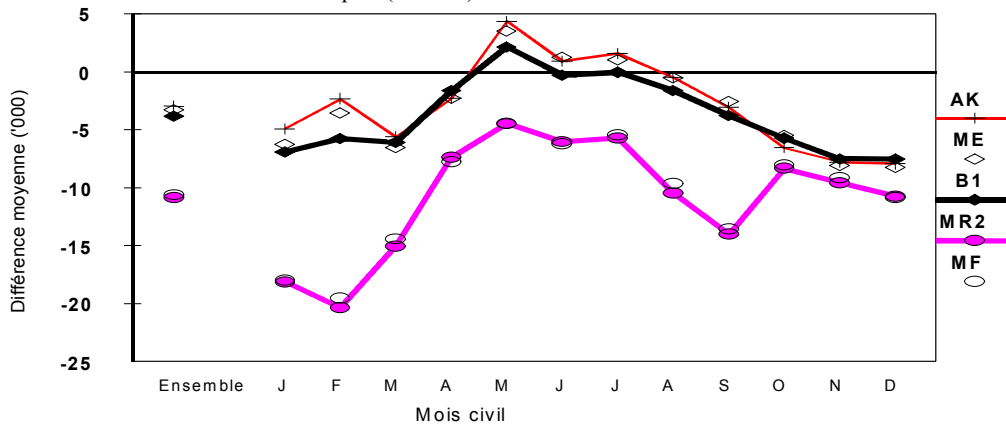


Diagramme 4. Différence moyenne par rapport à l'estimation RG, globalement et selon le mois civil, personnes sans emploi (milliers).



Le diagramme indique que les estimations de personnes ayant un emploi auraient été supérieures en moyenne si l'on avait utilisé l'estimateur MR2 ou MF. Cette différence vers le haut pour les estimateurs de régression modifiée est peut-être en réalité un aspect de la période particulière, puisque la différence semble avoir disparu depuis septembre 1997.

L'autre aspect des estimations MR2 et MF est le fait que la différence pour les personnes ayant un emploi est très saisonnière. Par exemple, le mouvement de décembre à janvier des estimations MR2 est supérieur de 40 000 environ au mouvement des estimations RG. Cela semble indiquer que l'échantillon apparié tend à manquer des personnes qui avaient un emploi en décembre, mais non pas en janvier. Le même caractère saisonnier se manifeste lorsqu'on observe directement des estimations de l'échantillon apparié. Le mouvement des groupes de renouvellement appariés ne manifeste pas cet important biais saisonnier.

Pour les estimations AK, ME et B1, les différences manifestent un certain caractère saisonnier, mais elles sont beaucoup plus faibles.

Le diagramme 4 indique les différences moyennes des diverses estimations par rapport à l'estimation RG pour les personnes sans emploi au cours de la même période. Ici, il semble y avoir une différence négative pour tous les estimateurs, mais elle est moins marquée pour les estimations AK, ME et B1 que pour les estimations MR2 et MF. Le changement de caractère saisonnier pour un passage des estimateurs RG à MR2 et MF est encore une fois plus marqué que pour le passage aux autres estimateurs.

#### 6.4 Erreurs-types

Les erreurs-types (ET) des estimations dans l'ensemble, selon l'état matrimonial et selon le sexe, sont présentées dans les diagrammes qui suivent. On les obtient comme pourcentage des estimations de l'ET pour la même estimation à l'aide de la méthode RG (c'est-à-dire les ET de l'EPA courante), et on établit alors la moyenne de ces pourcentages pour la période en fonction de laquelle ils ont

été préparés (juin 1993 à janvier 1999 pour les estimations de niveau). Les diagrammes 5, 6, 7 et 8 indiquent les ET tant pour les personnes ayant un emploi que pour les personnes sans emploi en fonction du niveau, du mouvement, de la moyenne trimestrielle et du mouvement de la moyenne trimestrielle respectivement.

Pour toutes ces estimations, l'estimateur B1 de type MELSB comporte une erreur d'échantillonnage plus faible que les estimateurs AK ou ME. Puisque l'estimation B1 semble manifester un biais et un caractère saisonnier de biais comparables, les estimateurs AK et ME utilisés dans la présente étude ne semblent pas pouvoir concurrencer l'estimateur B1.

Par contre, les estimateurs de régression modifiée MR2 et MF manifestent des erreurs d'échantillonnage nettement inférieures à celles de l'estimateur B1 pour les personnes ayant un emploi, pour les estimations dans l'ensemble et pour les estimations selon le sexe. Ce sont là des estimations clés de la régression modifiée – d'autres estimations clés comme les estimations selon l'État ont aussi montré de meilleures erreurs-types. Les estimations selon l'état matrimonial ne sont pas des estimations clés, et elles comportent des erreurs-types plus élevées pour les estimations MR2 et MF que pour l'estimateur B1.

En ce qui concerne les personnes sans emploi, l'amélioration des ET pour un recours aux estimateurs MR2 et MF est moins constante, et elle disparaît complètement pour les estimations de la moyenne trimestrielle. L'estimateur B1 réduit l'erreur-type de façon plus constante, même si les gains sont moins élevés pour les personnes sans emploi que pour les personnes ayant un emploi.

#### 6.5 Série désaisonnalisée et série de tendances

Le Bureau of Statistics de l'Australie fait appel au progiciel X11 (Shiskin, Young et Musgrave 1967) pour préparer des estimations désaisonnalisées visant à éliminer des séries divers effets de l'année civile. Le progiciel permet également de préparer une tendance, qui est un indicateur du comportement sous-jacent de la série.

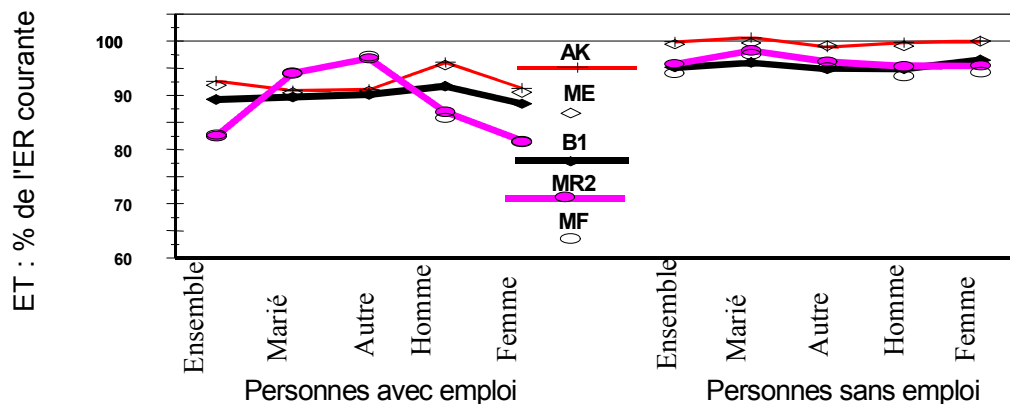


Diagramme 5. Erreur-type du niveau (% de l'ET courante).

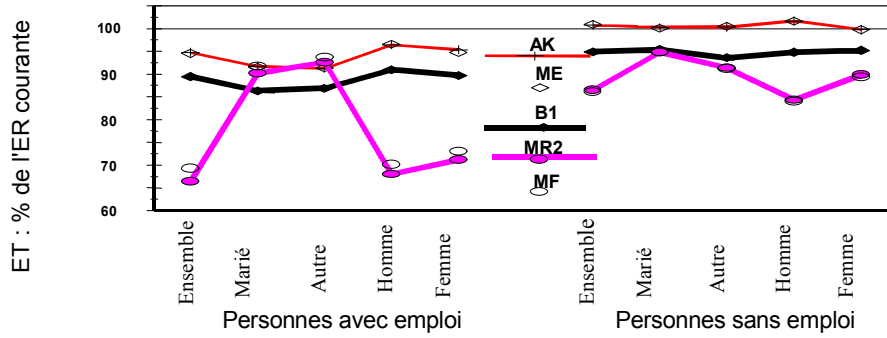


Diagramme 6. Erreur-type du mouvement (% de l'ET courante).

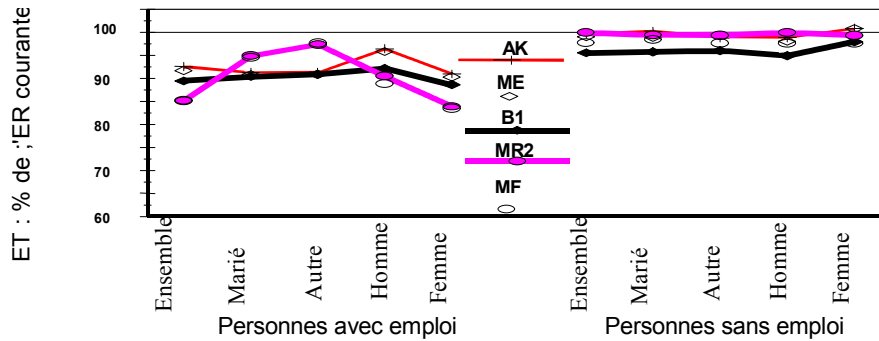


Diagramme 7. Erreur-type de la moyenne trimestrielle (% de l'ET courante).

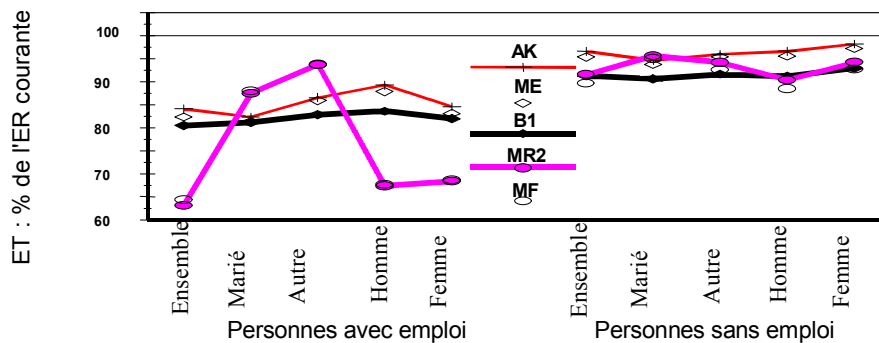


Diagramme 8. Erreur-type du mouvement de la moyenne trimestrielle (% de l'ET courante).

La valeur de la tendance pour tel moment est révisée à mesure que les données pour des moments ultérieurs sont accessibles. J'ai estimé l'erreur-type des estimations de tendance à la fin de la série (tendance de fin) et pour les mêmes moments lorsque douze autres mois de données sont disponibles (tendance à mi-chemin). Les révisions de la tendance (ou du mouvement de tendance) sont définies comme la différence entre la valeur à mi-chemin et la valeur de fin de la tendance (ou du mouvement de tendance). L'ampleur de la révision dépend de la forme de la série réelle et de l'erreur d'échantillonnage des séries estimées. La révision de la tendance quadratique moyenne pour une série d'estimations non biaisées est la somme de deux composantes : la révision de la tendance quadratique

moyenne qui serait survenue même en l'absence d'erreur d'échantillonnage, et la variance de l'estimation de la révision. Ainsi, l'erreur-type de la révision est une mesure de la composante erreur d'échantillonnage de la révision de la tendance quadratique moyenne (voir Bell 1999).

Les chiffres désaisonnalisés se prêtent à une révision semblable. Je présente les erreurs-types pour le niveau et le mouvement des estimations désaisonnalisées à la fin de la série. Les erreurs-types étaient très semblables pour les révisions ultérieures de ces estimations.

La technique jackknife avec suppression d'un groupe a permis de préparer des estimations de l'erreur-type pour les différentes estimations de tendance et désaisonnalisées. Cette technique suppose la préparation de versions répétées

des estimations. Malheureusement, l'étude a fourni des valeurs répétées pour la série chronologique uniquement pour les moments entre juin 1993 et janvier 1999. Chacune de ces séries chronologiques répétées a été complétée par les neuf années précédentes de données chronologiques afin qu'il y ait suffisamment de données pour l'application du progiciel X11. Puisque les séries de tendances et désaisonnalisées répétées se fondent sur les mêmes valeurs avant juillet 1993, l'estimation de type jackknife de l'ET aura tendance à sous-estimer légèrement l'ET réelle, surtout pour les moments au début de la série. Afin de réduire cet effet au minimum, on établit la moyenne des mesures du changement de l'erreur d'échantillonnage pour les mois à compter de janvier 1995 uniquement (et uniquement jusqu'à janvier 1998, de façon que l'on puisse utiliser les douze mois jusqu'à janvier 1999 pour l'estimation des révisions).

Le tableau 1 présente ces erreurs-types moyennes pour diverses mesures de tendance et désaisonnalisées, relativement à celles que l'on peut obtenir de l'estimateur RG courant, tant pour les personnes ayant un emploi que pour les personnes sans emploi. Dans ce même tableau on trouve les chiffres correspondants pour le niveau, le mouvement, la moyenne trimestrielle et le mouvement de la moyenne trimestrielle, en fonction des diagrammes 5 à 8.

**Tableau 1**  
Erreur-type comme pourcentage de l'erreur-type  
de l'estimateur RG courant

	AK	ME	B1	MR2	MF
Personnes ayant un emploi :					
Niveau	93	92	89	82	83
Mouvement	95	95	89	66	69
Moyenne trimestrielle	93	92	89	85	85
Mouvement de la moyenne trimestrielle	84	82	80	63	64
Désaisonnalisé	94	92	90	87	88
Mouvement de la désaisonnalisation	96	95	91	68	71
Tendance de fin	93	91	89	88	88
Mouvement de la tendance de fin	86	84	82	65	67
Révision de la tendance	88	85	83	66	68
Révision du mouvement de la tendance	89	86	84	67	69
Personnes sans emploi :					
Niveau	100	99	95	96	94
Mouvement	101	101	95	87	86
Moyenne trimestrielle	100	99	95	100	98
Mouvement de la moyenne trimestrielle	97	95	91	92	90
Désaisonnalisé	100	99	95	96	95
Mouvement de la désaisonnalisation	102	102	95	87	86
Tendance de fin	99	98	95	99	97
Mouvement de la tendance de fin	97	95	92	93	91
Révision de la tendance	97	95	91	91	89
Révision du mouvement de la tendance	97	95	92	92	90

J'estime que, pour de nombreuses applications, les indicateurs les plus importants sont ceux qui indiquent la direction sous-jacente de la série à la fin courante, c'est-à-dire le mouvement de la moyenne trimestrielle et le

mouvement de la tendance. Une réduction de l'erreur-type pour ces éléments rend la direction sous-jacente de la série à la fin plus claire, même pour les utilisateurs qui se fient sur une inspection visuelle ou sur un processus de lissage quelconque autre que la tendance X11. Cela offre dès lors de meilleures chances de détecter les points tournants de la série sous-jacente.

Pour ce qui est du mouvement de la tendance, l'estimateur B1 permet une réduction de 18 % de l'erreur-type pour les personnes ayant un emploi et une réduction de 8 % pour les personnes sans emploi. Quant au MR2, ces réductions sont de 35 % et de 7 % respectivement. Les estimateurs composites permettent également de réduire l'apport de l'erreur d'échantillonnage aux révisions de la série des tendances.

## 6.6 Résumé

Le présent exposé décrit une variante de l'estimateur MELSB, l'estimateur B1, qui permet d'appliquer la technique de régression généralisée à une estimation composite fondée sur une fenêtre de sept mois de données. Pour les données de l'Australie, le B1 comporte une erreur d'échantillonnage moins élevée que les estimateurs MELSB ou AK traditionnels pour un choix de mesures comprenant les estimations de tendance et désaisonnalisées. L'auteur a également évalué un estimateur MR2 composite « de régression modifiée » proposé par A.C. Singh et une variante de cet estimateur proposée par W. Fuller. Ces estimateurs ont donné des erreurs d'échantillonnage appréciablement moins élevées que l'estimateur B1 pour un certain nombre de mesures, en particulier celles qui se fondent sur des personnes ayant un emploi.

L'évaluation d'un estimateur composite dépend de nombreux facteurs autres que les erreurs d'échantillonnage. L'estimateur B1 comporte l'inconvénient d'exiger pour les totalisations des agrégations pondérées de sept mois de données, tandis que les estimateurs de régression modifiée fournissent des poids pour un seul mois de données. Par contre, les estimateurs de régression modifiée risquent d'être biaisés si les personnes déclarantes au cours de deux mois successifs (l'échantillon apparié) ne sont pas représentatives des autres personnes (par exemple les personnes qui changent de logement). L'introduction des estimateurs de régression modifiée entraînerait également un changement plus prononcé de l'estimation et du caractère saisonnier que l'introduction de l'estimateur B1.

## Remerciements

Les idées exprimées dans le présent exposé sont celles de l'auteur et ne reflètent pas nécessairement celles du Bureau of Statistics de l'Australie. Les personnes qui citent ou qui utilisent ces idées voudront bien les attribuer clairement à l'auteur.

## Bibliographie

- Bailar, B.A. (1975). The effect of rotation group bias on estimates from panel surveys. *Journal of American Statistical Association*, 70, 23-29.
- Bell, P.A. (1998). Using state space models and composite estimation to measure the effects of telephone interviewing on labour force estimates. Document de travail *Econometrics and Applied Statistics*, numéro de 1351.0, no. 98/2, ABS, Canberra.
- Bell, P.A. (1999). The impact of sample rotation patterns and composite estimation on survey outcomes. Document de travail, *Econometrics and Applied Statistics*, numéro de 1351.0, no. 99/1, ABS, Canberra.
- Bell, P.A., et Carolan, A. (1998). Trend estimation for small areas from a continuing survey with controlled sample overlap. Document de travail, *Econometrics and Applied Statistics*, numéro de 1351.0, no. 98/1, ABS, Canberra.
- Fuller, W.A. (1990). Analyse d'enquêtes à passages répétés. *Techniques d'enquête*, 16, 177-190.
- Fuller, W.A. (1999). Canadian Regression Composite Estimation. Manuscrit non publié.
- Gurney, M., et Daly, J.F. (1965). A multivariate approach to estimation in periodic sample surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 247-257.
- Jessen, R.J. (1942). Statistical investigation of a farm survey for obtaining farm facts. *Iowa Agricultural station research Bulletin*, 304.
- Kott, P.S. (1998). Using the delete-a-group jackknife variance estimator in practice. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 483-486.
- Lent, J., Miller, S. et Cantwell, P. (1994). Composite weights for the Current Population Surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 867-872.
- Lent, J., Miller, S. et Cantwell, P. (1996). Effect of composite weights on some estimates from the Current Population Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 130-139.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. Deuxième édition. New York : John Wiley & Sons, Inc.
- Singh, A.C., et Merkouris, P. (1995). Composite estimation by modified regression for repeated surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 420-425.
- Singh, A.C. (1996). Combining information in survey sampling by modified regression. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 120-129.
- Singh, A.C., Kennedy B., Wu S. et Brisebois F. (1997). Composite estimation for the Canadian Labour Force Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 300-305.
- Shiskin, J., Young, A., et Musgrave, J. (1967). *The X-11 variant of Census Method II Seasonal Adjustment*, Bureau of the Census, U.S. Department of Commerce, Papier technique 15.
- Yansaneh, I.S., et Fuller, W.A. (1998). Méthode optimale d'estimation réursive pour les enquêtes répétitives. *Techniques d'enquête*, 24, 33-42.