# Comparaison d'estimateurs alternatifs dans l'enquête emploi en continu

Dominique Place

INSEE
Unité méthodes statistiques

JMS 2009

# Plan

# Principe

Dans une enquête répétée avec réinterrogation, les estimations entre les différentes périodes sont corrélées.

La prise en compte des périodes passées peut permettre d’améliorer la précision de l’estimation de la période courante ainsi que celles des évolutions et des moyennes annuelles.

# Les trois types d'estimateurs composites

- les estimateurs linéaires optimaux à fenêtre fixe (Jessen, 1942, Bell, 1998) :
  une combinaison linéaire optimale d'estimateurs élémentaires définis sur un nombre fixe de périodes, utilisés dans l'enquête emploi australienne
- les estimateurs à composante récursive (Gurney et Daly, 1965) :
  les estimateurs K et AK, utilisés dans la *Current Population Survey* (USA)
- les estimateurs par calage modifié (Singh et Merkouris, 1995, Fuller, 1999) :
  aux variables de calage habituelles, sont ajoutées des variables relatives à la période précédente, avec imputation pour les entrants, utilisés dans l'EPA au Canada

Comparaison de ces trois types d'estimateurs effectuée par Ph. Bell sur l'enquête emploi australienne (2001)

# L'estimateur linéaire optimal à fenêtre fixe

- Estimateur élémentaire pour le trimestre $t$ et le rang d'interrogation $Ri$ : $\hat{y}_t^{Ri}$
  Estimateur transversal naturel (calé) : $\hat{Y}_t = \sum_{i=1}^{6} \hat{y}_t^{Ri}/6$
- Les estimateurs linéaires considérés sont :
  $$\hat{y}_t^f = \sum_{u=0}^{f-1} \sum_{i=1}^{6} a_i^u \hat{y}_{t-u}^{Ri} = A' \hat{Y}_{tf}$$
  où $f$ est la fenêtre fixée et $A = (a_1^0, \ldots, a_6^0, \ldots, a_1^f, \ldots, a_6^f)'$
- on impose des contraintes sur les coefficients pour avoir un biais nul : $\sum_{i=1}^{6} a_i^0 = 1$, $\sum_{i=1}^{6} a_i^u = 0$ pour $u \geq 1$
  exprimées matriciellement par $C'A = c$.
- Hypothèse liée aux contraintes : absence de biais de rotation.

# L'estimateur linéaire optimal à fenêtre fixe

## Résultat

*Soit $V$ la matrice de variance-covariance du vecteur $\hat{Y}_{tf}$ des estimateurs élémentaires, alors la minimisation de $Var(A'\hat{Y}_{tf})$ sous les contraintes précédentes conduit à : $A = V^{-1}C(C'V^{-1}C)^{-1}c$*

En pratique, les coefficients ne sont pas réactualisés à chaque période et ils sont fixés pour minimiser une situation moyenne :

- soit en utilisant une matrice de variance-covariance moyenne
- soit en prenant la moyenne de coefficients optimaux sur plusieurs trimestres passés, ce qui semble préférable.

# Prise en compte du biais de rotation

On suppose : $E\left(\hat{y}_t^{Ri}\right) = (1 + b_i)E(\hat{Y}_t) = (1 + b_i)Y_t$ avec $\sum_{i=1}^{6} b_i = 0$

Les $b_i$ sont par exemple les coefficients de Bailar.
Les contraintes deviennent :

$$\sum_{i=1}^{6}(1 + b_i)a_i^0 = 1, \sum_{i=1}^{6}(1 + b_i)a_i^u = 0 \text{ pour } u \geq 1$$

ce qui revient à modifier la matrice $C$.

# L'estimateur AK

- Ils dépendent de deux paramètres $A$ et $K$ :

$$\hat{Y}_t^{AK} = (1-K)\hat{Y}_t + K(\hat{Y}_{t-1}^{AK} + \hat{\Delta}_t) + A\beta_t$$

où : $$\hat{\Delta}_t = \frac{1}{5}\left(\sum_{i=2}^{6} \hat{y}_t^{Ri} - \sum_{i=1}^{5} \hat{y}_{t-1}^{Ri}\right)$$

$$\beta_t = \left(\hat{y}_t^{R1} - \frac{1}{5}\sum_{i=2}^{6} \hat{y}_t^{Ri}\right)$$

- Pour prendre en compte les biais de rotation, il faut modifier $\hat{\Delta}_t$ et $\beta_t$ en divisant chaque estimateur élémentaire par $1 + b_i$.
- A et K sont choisis pour minimiser la variance asymptotique.

# Estimation des variances et covariances

- Utilisation de formules analytiques pour un plan à deux phases avec linéarisation
- Difficulté majeure : dans un secteur, une seule aire est enquêtée à chaque trimestre et les remplacements d'aires se font à l'intérieur des secteurs
  $\longrightarrow$ pour les covariances intertemporelles, prise en compte des remplacements d'aires par imputation
- Autres difficultés : changements dans la composition des ménages, "panel" de logements et non d'individus, influence de la non-réponse

# Estimation des variances et covariances

- Utilisation de formules analytiques pour un plan à deux phases avec linéarisation
- Difficulté majeure : dans un secteur, une seule aire est enquêtée à chaque trimestre et les remplacements d'aires se font à l'intérieur des secteurs
  $\longrightarrow$ pour les covariances intertemporelles, prise en compte des remplacements d'aires par imputation
- Autres difficultés : changements dans la composition des ménages, “panel” de logements et non d'individus, influence de la non-réponse

$\Longrightarrow$ de nombreuses sources de biais difficiles à contrôler pour les covariances

# Variance d'un estimateur élémentaire

$$\widehat{Var}\left(\hat{y}_t^{Ri}\right) = \widehat{Var}(\hat{Y}_t) + 5nS_t^2$$

avec :

$n$ nombre total d'aires enquêtées sur un trimestre

$$S_t^2 = \frac{1}{n-1} \sum_{\text{aires } k} (\check{y}_{kt} - \bar{\check{y}}_t)^2$$

$\check{y}_{kt}$ somme pondérée des valeurs individuelles sur l'aire $k$,

$\check{y}_{tk} = \sum_{i \in k} y_{it} / \pi_k$

# Covariances intertemporelles

- sans remplacement d'aires :

$$\widehat{Cov}\left(\hat{y}_t^{Ri}, \hat{y}_{t'}^{Rj}\right) = \widehat{Cov}_{tt'}^{cyl} + \left(6 - \frac{1}{o_{tt'}}\right) nS_{tt'}$$

- avec remplacement d'aires :

$$\widehat{Cov}\left(\hat{y}_t^{Ri}, \hat{y}_{t'}^{Rj}\right) = \widehat{Cov}_{tt'}^{cyl} + \left(6 - \frac{1}{o_{tt'}}\right) nS_{tt'} - 6(1 - o_{tt'})\widehat{Cov}_{\text{intra}}$$

avec : $\widehat{Cov}_{tt'}^{cyl}$ estimation de la covariance entre $t$ et $t'$ sur l'échantillon cylindré d'aires ;

$$S_{tt'} = \frac{1}{n-1} \sum_{\substack{\text{aires } k \\ \text{non remplacees}}} (\check{y}_{kt} - \bar{\check{y}}_t)(\check{y}_{kt'} - \bar{\check{y}}_{t'}) ;$$

$o_{tt'}$ taux de recouvrement des échantillons entre $t$ et $t'$ ;

$\widehat{Cov}_{\text{intra}}$ covariance intra-secteurs entre $t$ et $t'$

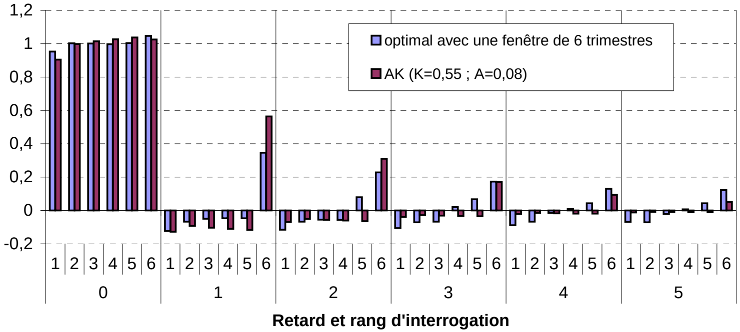
# Corrélations moyennes

| Retard | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Nombre de chômeurs | | | | | |
| Sans remplacement d'aires | 0,57 | 0,49 | 0,45 | 0,43 | 0,36 |
| Avec remplacement d'aires | 0,08 | 0,08 | 0,07 | 0,09 | 0,07 |
| Nombre d'actifs occupés | | | | | |
| Sans remplacement d'aires | 0,76 | 0,71 | 0,68 | 0,66 | 0,61 |
| Avec remplacement d'aires | 0,08 | 0,08 | 0,08 | 0,10 | 0,07 |

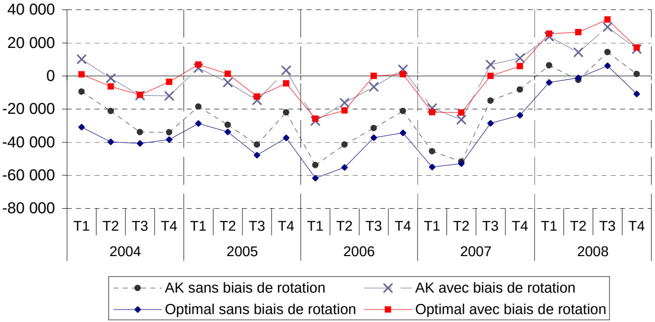
A cause du calage, les corrélations obtenues sont plus faibles que les corrélations individuelles.

# Multiplicateurs des poids pour le nombre de chômeurs

Une pondération plus importante des sixièmes rangs d'interrogation des trimestres passés



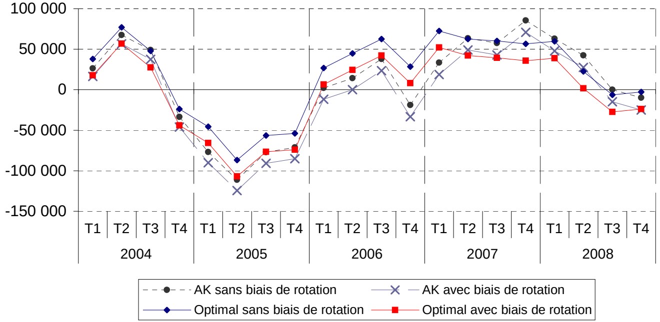
# Écarts sur le nombre de chômeurs




La fenêtre est de 6 trimestres pour les estimateurs linéaires optimaux.

Choix de K et A : K=0,6 ; A= 0,08 sans prise en compte du biais de rotation
K=0,55 ; A= 0,08 avec prise en compte du biais de rotation

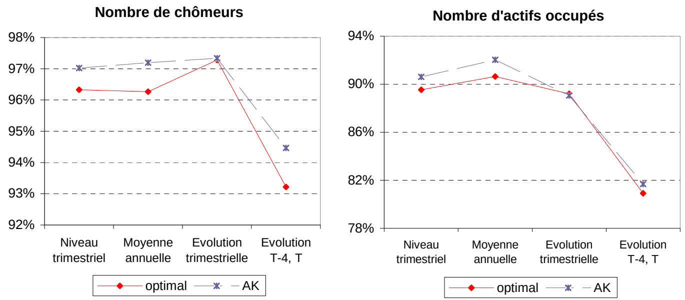
# Écarts sur le nombre d'actifs occupés




Choix de K et A : K=0,75 ; A= 0,08

Les profils restent semblables mais les écarts sont les plus réduits avec l'estimateur linéaire optimal tenant compte du biais de rotation.

# Écarts types (% des écarts types actuels)




Rapports moyens calculés sur la période 2005T2 - 2008T4

# Conclusion

- Résultats du même ordre de grandeur que ceux obtenus pour l'enquête emploi australienne (mensuelle) :
  gains en précision limités sur le nombre de chômeurs, mais plus intéressants sur le nombre d'actifs employés.

# Conclusion

- Résultats du même ordre de grandeur que ceux obtenus pour l'enquête emploi australienne (mensuelle) :
  gains en précision limités sur le nombre de chômeurs, mais plus intéressants sur le nombre d'actifs employés.
- Les biais de rotation ne jouent que sur la hauteur des courbes et non sur leur profil.

# Conclusion

- Résultats du même ordre de grandeur que ceux obtenus pour l'enquête emploi australienne (mensuelle) :
  gains en précision limités sur le nombre de chômeurs, mais plus intéressants sur le nombre d'actifs employés.
- Les biais de rotation ne jouent que sur la hauteur des courbes et non sur leur profil.
- En pratique, on peut calculer des estimations composites sur des grandeurs cibles (grandeurs par catégorie, pour les grandes régions...) et les incorporer aux poids trimestriels diffusés en reprenant le calage avec ces grandeurs.