



Etude de la précision des estimateurs de l'ESPA continue suisse selon le choix du schéma de rotation des échantillons

Jean-Pierre Renfer

Service des méthodes statistiques
Office fédéral de la statistique

Journées de Méthodologie Statistique
Paris, 23-25 mars 2009





Plan de la présentation

1. Objectifs de cette étude
2. Critères de précision EUROSTAT
3. Estimateurs considérés
4. Estimation de la variance de ces estimateurs
5. Comparaison des schémas de rotation envisagés
6. Conclusions et variante privilégiée



1. Objectifs de cette étude

- ▶ Evaluer l'impact de différents schémas de rotation sur la précision des estimateurs;
- ▶ Obtenir un ordre de grandeur de la taille d'échantillon pour satisfaire les critères de précision EUROSTAT;
- ▶ Choisir un schéma de rotation pour l'ESPA suisse révisée.



2. Critères de précision EUROSTAT

a) Règlement (CE) N° 577/98 du 9 mars 1998, article 3 pt 1 :

Pour un groupe de chômeurs représentant 5% de la population d'âge actif, l'erreur-type relative pour l'estimation des moyennes annuelles (ou des estimations de printemps, dans le cas d'une enquête annuelle effectuée au printemps) ne doit pas excéder 8% au niveau de la NUTS II de la sous-population en cause ($8\% \cdot 5\% = 0.4\%$ en absolu).

Les régions de moins de 300'000 habitants ne sont pas soumises à cette condition.



b) Règlement (CE) N° 577/98 du 9 mars 1998, article 3, pt 2 :

Dans le cas d'une enquête continue, pour des sous-populations représentant 5% de la population d'âge actif, l'erreur-type relative pour l'estimation des variations entre deux trimestres successifs, au niveau national, ne doit pas excéder 2% de la sous-population en cause. Pour les Etats membres dont la population est comprise entre un million et vingt millions d'habitants, la condition précédente est allégée de telle sorte que l'erreur-type relative pour l'estimation des variations trimestrielles ne doit pas excéder 3% de la sous-population en cause ($3\% \cdot 5\% = 0.15\%$ absolu).



$$\begin{aligned} \text{Var}(p_{t_2} - p_{t_1}) &= \text{Var}(p_{t_1}) + \text{Var}(p_{t_2}) - 2\text{Cov}(p_{t_1}, p_{t_2}) \\ &\approx \text{Var}(p_{t_1}) + \text{Var}(p_{t_2}) - 2\rho q \frac{[\text{Var}(p_{t_1}) + \text{Var}(p_{t_2})]}{2} \\ &= [\text{Var}(p_{t_1}) + \text{Var}(p_{t_2})](1 - \rho q) \\ &\approx \frac{2p(1-p)}{n}(1 - \rho q) \end{aligned}$$

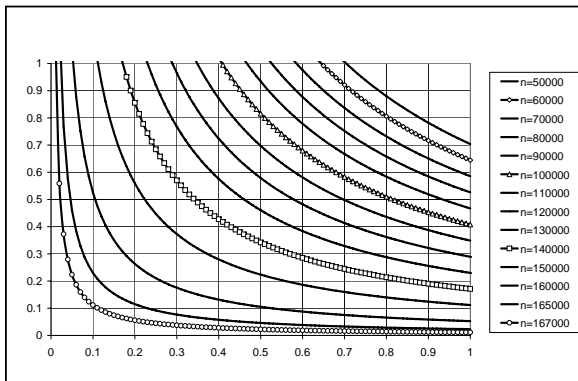
où p_{t_i} est la proportion d'intérêt au temps i , q le recouvrement et ρ le coefficient de corrélation.

La taille minimale d'échantillon par trimestre (**avec $q=0$**) est :

$$n = \frac{2 \cdot 0.05 \cdot 0.95}{(0.05 \cdot 0.03)^2} \simeq 42222 \text{ trimestriellement au niveau national}$$



Courbes de niveau de la fonction $f(\rho, q) = 168888(1 - \rho q)$.





3. Estimateurs considérés

| Totaux | Variations |
|--|---|
| 1. \hat{Y}_m | 4. ${}_1\hat{\Delta}_m = \hat{Y}_m - \hat{Y}_{m-1}$ |
| 2. $\hat{Y}_t = \frac{4}{13}(\hat{Y}_m + \hat{Y}_{m+1}) + \frac{5}{13}(\hat{Y}_{m+2})$ | 5. ${}_{12}\hat{\Delta}_m = \hat{Y}_m - \hat{Y}_{m-12}$ |
| 3. $\hat{Y}_a = \frac{\hat{Y}_m + \hat{Y}_{m+1} + \dots + \hat{Y}_{m+11}}{12}$ | 6. ${}_1\hat{\Delta}_t = \hat{Y}_t - \hat{Y}_{t-1}$ |
| | 7. ${}_4\hat{\Delta}_t = \hat{Y}_t - \hat{Y}_{t-4}$ |
| | 8. ${}_1\hat{\Delta}_a = \hat{Y}_a - \hat{Y}_{a-1}$ |



4. Estimation de la variance de ces estimateurs

Pour les totaux :

$$1. \widehat{V}(\widehat{Y}_m) = \widehat{V}(\widehat{Y}_m);$$

$$2. \widehat{V}(\widehat{Y}_t) \approx \frac{1}{3} \widehat{V}(\widehat{Y}_m);$$

$$3. \widehat{V}(\widehat{Y}_a) = \widehat{V} \left(\frac{\widehat{Y}_m + \dots + \widehat{Y}_{m+11}}{12} \right) = \frac{1}{144} \sum_{i=1}^{12} \sum_{j=1}^{12} \widehat{C}(\widehat{Y}_i, \widehat{Y}_j)$$



| | termes 0 | 1 t | 2 t | 3 t |
|---|-----------|-----------|-----------|----------|
| $\widehat{C}(\widehat{Y}_1, \widehat{Y}_2) + \widehat{C}(\widehat{Y}_1, \widehat{Y}_3) + \dots + \widehat{C}(\widehat{Y}_1, \widehat{Y}_{12})$ | 2 | 3 | 3 | 3 |
| $+ \widehat{C}(\widehat{Y}_2, \widehat{Y}_3) + \widehat{C}(\widehat{Y}_2, \widehat{Y}_4) + \dots + \widehat{C}(\widehat{Y}_2, \widehat{Y}_{12})$ | 1 | 3 | 3 | 3 |
| $+ \widehat{C}(\widehat{Y}_3, \widehat{Y}_4) + \widehat{C}(\widehat{Y}_3, \widehat{Y}_5) + \dots + \widehat{C}(\widehat{Y}_3, \widehat{Y}_{12})$ | 0 | 3 | 3 | 3 |
| $+ \widehat{C}(\widehat{Y}_4, \widehat{Y}_5) + \widehat{C}(\widehat{Y}_4, \widehat{Y}_6) + \dots + \widehat{C}(\widehat{Y}_4, \widehat{Y}_{12})$ | 2 | 3 | 3 | 0 |
| $+ \widehat{C}(\widehat{Y}_5, \widehat{Y}_6) + \widehat{C}(\widehat{Y}_5, \widehat{Y}_7) + \dots + \widehat{C}(\widehat{Y}_5, \widehat{Y}_{12})$ | 1 | 3 | 3 | 0 |
| $+ \widehat{C}(\widehat{Y}_6, \widehat{Y}_7) + \widehat{C}(\widehat{Y}_6, \widehat{Y}_8) + \dots + \widehat{C}(\widehat{Y}_6, \widehat{Y}_{12})$ | 0 | 3 | 3 | 0 |
| $+ \widehat{C}(\widehat{Y}_7, \widehat{Y}_8) + \widehat{C}(\widehat{Y}_7, \widehat{Y}_9) + \dots + \widehat{C}(\widehat{Y}_7, \widehat{Y}_{12})$ | 2 | 3 | 0 | 0 |
| $+ \widehat{C}(\widehat{Y}_8, \widehat{Y}_9) + \widehat{C}(\widehat{Y}_8, \widehat{Y}_{10}) + \dots + \widehat{C}(\widehat{Y}_8, \widehat{Y}_{12})$ | 1 | 3 | 0 | 0 |
| $+ \widehat{C}(\widehat{Y}_9, \widehat{Y}_{10}) + \widehat{C}(\widehat{Y}_9, \widehat{Y}_{11}) + \widehat{C}(\widehat{Y}_9, \widehat{Y}_{12})$ | 0 | 3 | 0 | 0 |
| $+ \widehat{C}(\widehat{Y}_{10}, \widehat{Y}_{11}) + \widehat{C}(\widehat{Y}_{10}, \widehat{Y}_{12})$ | 2 | 0 | 0 | 0 |
| $+ \widehat{C}(\widehat{Y}_{11}, \widehat{Y}_{12})$ | 1 | 0 | 0 | 0 |
| | | | | |
| Total | 12 | 27 | 18 | 9 |

$$\widehat{V}(\widehat{Y}_a) \approx \frac{1}{12} \widehat{V}(\widehat{Y}_m) + \frac{1}{72} \{9(q\widehat{\rho})_{m,m+3} + 6(q\widehat{\rho})_{m,m+6} + 3(q\widehat{\rho})_{m,m+9}\} \widehat{V}(\widehat{Y}_m)$$



Pour les variations :

$$4. \hat{V}({}_1\hat{\Delta}_m) = \hat{V}(\hat{Y}_m - \hat{Y}_{m-1}) \stackrel{!}{=} \hat{V}(\hat{Y}_m) + \hat{V}(\hat{Y}_{m-1}) \approx 2\hat{V}(\hat{Y}_m);$$

$$5. \hat{V}({}_{12}\hat{\Delta}_m) = \hat{V}(\hat{Y}_m - \hat{Y}_{m-12}) \approx 2(1 - (q\hat{\rho})_{m,m-12})\hat{V}(\hat{Y}_m);$$

$$6. \hat{V}({}_1\hat{\Delta}_t) = \hat{V}(\hat{Y}_t - \hat{Y}_{t-1}) \approx \frac{2}{3}(1 - (q\hat{\rho})_{m,m-3})\hat{V}(\hat{Y}_m);$$

$$7. \hat{V}({}_4\hat{\Delta}_t) = \hat{V}(\hat{Y}_t - \hat{Y}_{t-4}) \approx \frac{2}{3}(1 - (q\hat{\rho})_{m,m-12})\hat{V}(\hat{Y}_m);$$

$$8. \hat{V}({}_1\hat{\Delta}_a) = \hat{V}\left(\left[\frac{\hat{Y}_m + \dots + \hat{Y}_{m+11}}{12}\right] - \left[\frac{\hat{Y}_{m-1} + \dots + \hat{Y}_{m-12}}{12}\right]\right)$$



5. Comparaison des schémas de rotation

Variante 3-3-3-3 :

nombre de vagues : 5

q à 3 mois : 80%

q à 6 mois : 60%

q à 9 mois : 40%

q à 12 mois : 20%

| T ₁ | T ₂ | T ₃ | T ₄ | T ₅ | T ₆ | ... |
|----------------|----------------|----------------|----------------|----------------|----------------|-----|
| | | | | | | |
| X | | | | | | |
| X | X | | | | | |
| X | X | X | | | | |
| X | X | X | X | | | |
| X | X | X | X | X | | |
| | X | X | X | X | X | |

Variante 3-9-3 :

nombre de vagues : 4

q à 3 mois : 50%

q à 6 mois : 0%

q à 9 mois : 25%

q à 12 mois : 50%

q à 15 mois : 25%

| T ₁ | T ₂ | T ₃ | T ₄ | T ₅ | T ₆ | ... |
|----------------|----------------|----------------|----------------|----------------|----------------|-----|
| | | | | | | |
| X | | | | | | |
| X | X | | | | | |
| | X | X | | | | |
| | | X | X | | | |
| X | | | X | X | | |
| X | X | | | X | X | |



Variante 3-9 :

nombre de vagues : 3

q à 3 mois : 33%

q à 6 mois : 0%

q à 9 mois : 33%

q à 12 mois : 33%

| T ₁ | T ₂ | T ₃ | T ₄ | T ₅ | T ₆ | ... |
|----------------|----------------|----------------|----------------|----------------|----------------|-----|
| | | | | | | |
| X | | | | | | |
| | X | | | | | |
| | | X | | | | |
| X | | | X | | | |
| X | X | | | X | | |
| | X | X | | | X | |

Variante 12-12-12-12 :

nombre de vagues : 2

q à 3 mois : 0%

q à 6 mois : 0%

q à 9 mois : 0%

q à 12 mois : 50%

| T ₁ | T ₂ | T ₃ | T ₄ | T ₅ | T ₆ | ... |
|----------------|----------------|----------------|----------------|----------------|----------------|-----|
| | | | | | | |
| X | | | | | | |
| | X | | | | | |
| | | X | | | | |
| | | | X | | | |
| X | | | | X | | |
| | X | | | | X | |



En faisant l'hypothèse que la corrélation successive entre deux trimestres consécutifs est de $\rho = 60\%$ et que cette corrélation diminue de 3% à chaque trimestre, on obtient :

| Estim. | Sans rotation | Avec rotation | | | |
|-------------------------|----------------|---------------------|-------------------|-----------------|-------------------------|
| | $V(\hat{Y}_m)$ | Variante 3-3-3-3 | Variante 3-9-3 | Variante 3-9 | Variante 12-12-12-12 |
| \hat{Y}_m | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| \hat{Y}_t | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \hat{Y}_a | 0.08 | 0.18 | 0.13 | 0.12 | 0.08 |
| ${}_1\hat{\Delta}_m$ | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| ${}_{12}\hat{\Delta}_m$ | 2.00 | 1.79 | 1.49 | 1.66 | 1.49 |
| ${}_1\hat{\Delta}_t$ | 0.67 | 0.35 | 0.47 | 0.53 | 0.67 |
| ${}_4\hat{\Delta}_t$ | 0.67 | 0.60 | 0.50 | 0.55 | 0.50 |
| ${}_1\hat{\Delta}_a$ | 0.17 | 0.27 | 0.17 | 0.17 | 0.12 |



| Variante | Estimateurs | taille d'échantillon annuelle | | |
|-------------|--------------------------|-------------------------------|--------------|--------------|
| | | $n = 80000$ | $n = 100000$ | $n = 120000$ |
| | | in % | in % | in % |
| 3-3-3-3 | \widehat{Y}_a | ± 0.227 | ± 0.203 | ± 0.185 |
| 3-3-3-3 | ${}_1\widehat{\Delta}_t$ | ± 0.314 | ± 0.281 | ± 0.257 |
| | | | | |
| 3-9 | \widehat{Y}_a | ± 0.182 | ± 0.163 | ± 0.148 |
| 3-9 | ${}_1\widehat{\Delta}_t$ | ± 0.390 | ± 0.349 | ± 0.318 |
| | | | | |
| 3-9-3 | \widehat{Y}_a | ± 0.190 | ± 0.170 | ± 0.155 |
| 3-9-3 | ${}_1\widehat{\Delta}_t$ | ± 0.365 | ± 0.326 | ± 0.298 |
| | | | | |
| 12-12-12-12 | \widehat{Y}_a | ± 0.154 | ± 0.138 | ± 0.126 |
| 12-12-12-12 | ${}_1\widehat{\Delta}_t$ | ± 0.436 | ± 0.390 | ± 0.356 |



6. Conclusions et variante privilégiée

- ▶ Le choix du schéma de rotation est crucial;
- ▶ Trouver un compromis entre la variante 3-3-3-3 (${}_1\hat{\Delta}_t$) et la variante actuelle (\hat{Y}_a);
- ▶ Solution raisonnable : variante 3-9-3 ou 3-9;
- ▶ Préférence pour la variante 3-9-3 : $\hat{V}({}_1\hat{\Delta}_t)$ plus petite;

Variante privilégiée, proposée et acceptée par l'OFS : 3-9-3



Bibliographie

Bosredon, J. et Février, P. (2000). *Estimations dans l'enquête emploi en continu*. Actes des journées de méthodologie statistique, 4 et 5 décembre 2000, Méthodologie statistique, n°100, pp. 393-413, Méthodologie statistique, Paris, INSEE.

Caron, N. et Ravalet, P. (2000). *Estimation dans les enquêtes répétées : application à l'enquête emploi en continu*. Actes des journées de méthodologie statistique, Méthodologie statistique, n°100, pp. 327-392, Méthodologie statistique, Paris, INSEE.

Christine, M. (2000). *La construction de l'échantillon de la future enquête emploi en continu à partir du recensement de 1999*. Actes des journées de méthodologie statistique, n°100, pp. 175-229, Méthodologie statistique, Paris INSEE.



Kytir, J. und Stadler, B. (2004). *Die kontinuierliche Arbeitskräfteerhebung im Rahmen des neuen Mikrozensus. Vom "alten" zum "neuen" Mikrozensus.* In : Statistische Nachrichten 6/2004, S. 511-518.

Mirza, H. and Hörngren, J. (2002). The sampling- and the estimation procedure in the Swedish Labour Force Survey. R&D Report 2002:4, Statistics Sweden.

RÈGLEMENT (CE) N° 577/98 DU CONSEIL DU 9 mars 1998 relatif à l'organisation d'une enquête par sondage sur les forces de travail dans la Communauté.

De Vitiis, C., Di Consiglio, L. Falorsi, S. (2005). *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di lavoro.* ISTAT, Servizio Progettazione e Supporto Metodologico.