

REVUE DE STATISTIQUE APPLIQUÉE

J. FINE

S. FOTSO

Contribution à l'étude du modèle âge-période-cohorte

Revue de statistique appliquée, tome 37, n° 3 (1989), p. 39-56

http://www.numdam.org/item?id=RSA_1989__37_3_39_0

© Société française de statistique, 1989, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

CONTRIBUTION A L'ÉTUDE DU MODÈLE ÂGE-PÉRIODE-COHORTE

J. FINE ET S. FOTSO

Laboratoire de Statistique et Probabilités, UA CNRS 745

Université Paul Sabatier

118, route de Narbonne, 31062 Toulouse Cedex, France

RÉSUMÉ

Nous étudions l'évolution dans le temps des taux par âge d'un phénomène démographique à l'aide du modèle âge-période-cohorte. Ce modèle peut être considéré comme une analyse de variance à trois facteurs sans interaction; cependant, même après avoir supposé nulle la somme des effets de chaque facteur, la matrice d'incidence du modèle est singulière (de plein rang moins un) ce qui pose un problème d'identification. L'estimation des effets se fait généralement par les moindres carrés ordinaires. Nous proposons de les estimer dans le cadre d'un modèle de quasi-vraisemblance et apportons des éléments de réponse au problème d'identification.

Mots-clés : Modèle Age-Période-Cohorte, Quasi-vraisemblance, Ridge-régression, Contrainte d'identification.

SUMMARY

We use the so-called age-period-cohort model to study the rates of a demographic phenomenon from data presented in the form of a two-way table where rows correspond to age groups and columns to periods of time. This model may be considered as a three-way anova model without interaction; but, even under the hypothesis of a null sum of the effects for each factor, the incidence matrix of the model is non regular, full rank minus one, and this raises an identification problem. The effects are generally estimated by the ordinary least squares criterium. We propose here an estimation procedure based on a quasi-vraisemblance model and we give some answers to the identification problem.

1. Introduction

Pour analyser l'évolution dans le temps des taux par âge d'un phénomène démographique (taux d'activité par exemple) le modèle Age-Période-Cohorte (A.P.C.) est fréquemment utilisé. Il s'agit de décomposer le taux en un effet général, un effet dû à l'âge, un effet dû à la période et un effet dû à la cohorte (ou génération). L'introduction d'un effet de cohorte a été rendu nécessaire par le paradoxe maintenant bien connu des démographes : pour chaque période les taux peuvent décroître avec l'âge alors que pour chaque cohorte les taux sont croissants avec l'âge (cf. exemple 3.4.). Le modèle A.P.C. peut être considéré comme une analyse de variance à deux facteurs, âge et période, avec une interaction assez particulière supposée constante pour une même cohorte, ou encore comme une analyse de variance à trois facteurs sans interaction. Cependant, les

trois facteurs sont liés. Aussi, même après avoir supposé que la somme des effets de chaque facteur est nulle, la matrice d'incidence des différents niveaux des facteurs est encore singulière, (de plein rang moins un), ce qui pose un problème d'identification.

L'estimation des effets est généralement obtenue par les moindres carrés ordinaires avec une contrainte linéaire "arbitraire" sur les effets pour lever le problème d'identification. Les résultats peuvent dépendre très fortement de la contrainte choisie. Dans le paragraphe 2 nous présentons le modèle A.P.C. et les propriétés des estimateurs des moindres carrés ordinaires à partir de l'article de synthèse sur le sujet de Kupper et al. (1983). Dans le cas où on ne connaît pas de contrainte sur les paramètres nous proposons deux critères pour faire un choix parmi l'ensemble des estimateurs. Le premier est d'être de variance minimale dans la classe des estimateurs de biais relatif minimum, le deuxième est d'être d'erreur quadratique moyenne relative minimale.

Nous proposons (§3) d'améliorer l'estimation des effets en utilisant le critère des moindres carrés généralisés. En effet, les taux peuvent être considérés comme des variables aléatoires dont les espérances vérifient le modèle A.P.C. et dont on peut expliciter les variances. Cependant les variances dépendent des moyennes et dépendent donc des paramètres à estimer ce qui conduit à utiliser le modèle de quasi-vraisemblance proposé par Mc Cullagh en 1983. Une utilisation directe de ce modèle est impossible à cause de la non-diagonalité de la matrice de variance-covariance des erreurs et de la non-régularité de la matrice d'incidence du modèle. Aussi, une solution adaptée à notre problème est proposée.

L'approche envisagée n'élimine pas le problème d'identification, cependant, elle apporte des arguments pour le choix d'une contrainte "naturelle" dans les modèles A.P.C. ce qui constitue un autre aspect de notre contribution à l'étude de ces modèles (§4). Les arguments sont en fait valables pour tous les modèles linéaires où existe un problème d'identification, les modèles A.P.C. ne servant que de support concret à la discussion.

2. Modèle A.P.C. et estimateur des moindres carrés orthogonaux

2.1. Modèle APC

On considère, pour une population, les taux, par âge et sur plusieurs périodes, d'individus soumis à un phénomène démographique (par exemple taux de femmes actives (cf. ex §3.4.)). On dispose d'une matrice de dimension (a,p) : $t = (t_i^j)$, $i=1,\dots,a$, $j=1,\dots,p$, où t_i^j est le taux à l'âge i et la période j . On suppose que les différences entre deux âges consécutifs et les différences entre deux périodes consécutives sont une même constante (en général égale à 1) de sorte que l'on trouve sur des diagonales parallèles les taux concernant les cohortes; on a ainsi $a-1+p$ cohortes et les individus qui ont l'âge i à la période j font partie de la cohorte $a-i+j$.

$$t = \begin{matrix} & & \text{périodes} \\ & & 1 \quad j \quad p \\ \text{âges } i & \left(\begin{matrix} \ddots & & \ddots \\ & t_i^j & \\ \ddots & & \ddots \end{matrix} \right) & a-1+p \\ a & & 1 \quad a-i+j & \text{cohortes} \end{matrix}$$

Dans l'exemple proposé, une cohorte correspond à une génération d'individus, l'âge est l'âge réel des individus. Plus généralement, une cohorte est l'ensemble des individus qui débutent la même année le phénomène étudié, l'âge est alors la durée depuis cette année de référence. Par exemple, si le phénomène étudié est l'insertion professionnelle des diplômés d'une école alors une cohorte est l'ensemble des étudiants ayant eu leur diplôme la même année, l'âge est la durée d'attente du premier emploi.

On considère tout d'abord le modèle suivant :

$$t_i^j = \delta + \alpha_i + \beta_j + \gamma_i^j + \varepsilon_i^j$$

où δ est l'effet général, α_i l'effet de l'âge i , β_j l'effet de la période j , γ_i^j l'interaction âge x période (i,j) et ε_i^j une erreur aléatoire d'espérance nulle, les effets et l'interaction étant supposés non aléatoires.

Dans la mesure où l'on a une seule observation par couple (i,j) il n'est pas possible de "distinguer" l'interaction de l'erreur. De plus, il semble naturel d'interpréter l'interaction âge x période γ_i^j comme un effet de cohorte γ_k avec $k=a-i+j$; aussi on considère le modèle suivant, appelé modèle A.P.C. :

$$t_i^j = \delta + \alpha_i + \beta_j + \gamma_k + \varepsilon_i^j \quad \text{avec} \quad k = a - i + j.$$

On peut considérer que le modèle est, soit un modèle à deux facteurs, avec une interaction assez particulière car supposée constante dans une même cohorte, soit un modèle à trois facteurs sans interaction. Il est à noter que l'on peut dans ce modèle séparer les effets de l'erreur.

Soit T le vecteur colonne ($ap,1$) dont les éléments sont les éléments de t rangés lignes par lignes, c'est-à-dire : $T' = (t_1^1 \dots t_1^p \ t_2^1 \dots t_2^p \dots t_a^1 \dots t_a^p)$, $\tilde{\rho}$ le vecteur colonne ($2(a+p), 1$) défini par $\tilde{\rho}' = (\delta \ \alpha_1 \dots \alpha_a \ \beta_1 \dots \beta_p \ \gamma_1 \dots \gamma_{a+p-1})$ et \tilde{X} la matrice de dimension ($ap, 2(a+p)$) dont les colonnes sont $\mathbf{1}, \tilde{A}_1, \dots, \tilde{A}_a, \tilde{B}_1, \dots, \tilde{B}_p, \tilde{C}_1, \dots, \tilde{C}_{a+p-1}$ où $\mathbf{1}$ est un vecteur composé uniquement de 1, \tilde{A}_i (resp. \tilde{B}_j , resp. \tilde{C}_k) est l'indicatrice du niveau i du facteur âge (resp. du niveau j du facteur période, resp. du niveau k du facteur cohorte).

Le modèle A.P.C. s'écrit alors matriciellement sous la forme :

$$E(T) = \tilde{X} \tilde{\rho}$$

Cependant on a $\sum_{i=1}^a \tilde{A}_i = \sum_{j=1}^p \tilde{B}_j = \sum_{k=1}^{a+p-1} \tilde{C}_k = \mathbf{1}$ aussi on pose

les contraintes habituelles $\sum_{i=1}^a \alpha_i = \sum_{j=1}^p \beta_j = \sum_{k=1}^{a+p-1} \gamma_k = 0$, les effets α_a, β_p et

γ_{a+p-1} seront déduits des autres; ce qui est équivalent à écrire le modèle sous la forme suivante :

$$E(T) = X\rho$$

où ρ est le vecteur colonne $(2(a+p)-3, 1)$ défini par

$$\rho' = (\delta \alpha_1 \dots \alpha_{a-1} \beta_1 \dots \beta_{p-1} \gamma_1 \dots \gamma_{a+p-2})$$

et X est la matrice de dimension $(ap, 2(a+p)-3)$ dont les colonnes sont $\mathbf{1}, A_1, \dots, A_{a-1}, B_1, \dots, B_{p-1}, C_1, \dots, C_{a+p-2}$ avec $A_i = \tilde{A}_i - \tilde{A}_a, B_j = \tilde{B}_j - \tilde{B}_p, C_k = \tilde{C}_k - \tilde{C}_{a+p-1}$.

Cependant, le modèle n'est pas identifié car la matrice X est singulière. En effet elle est de plein rang moins un, c'est-à-dire qu'il existe exactement une relation linéaire, à une constante multiplicative près, entre les colonnes de X dont la forme est donnée en 1980 par Kupper et Janis (cf. Kupper et al, 1983) :

$$\sum_{i=1}^{a-1} \left(i - \frac{a+1}{2}\right) A_i - \sum_{j=1}^{p-1} \left(j - \frac{p+1}{2}\right) B_j + \sum_{k=1}^{a+p-2} \left(k - \frac{a+p}{2}\right) C_k = 0$$

c'est-à-dire $Xv = 0$ avec

$$v' = \left(0, 1 - \frac{a+1}{2}, \dots, a-1 - \frac{a+1}{2}, -1 + \frac{p+1}{2}, \dots, -p+1 + \frac{p+1}{2}, 1 - \frac{a+p}{2}, \dots, a+p-2 - \frac{a+p}{2}\right)$$

Les modèles $E(T)=X\rho$ et $E(T)=X(\rho+\lambda v), \lambda \in \mathbb{R}$, sont donc indiscernables. Il est pourtant d'usage de supposer dans les analyses A.P.C. qu'il existe un unique vecteur ρ mesurant les effets, l'objectif des analyses étant de l'estimer le mieux possible. Nous discutons cette hypothèse dans le paragraphe 4.

2.2. Estimateur des moindres carrés ordinaires

L'estimateur de ρ par les moindres carrés ordinaires conduit à la résolution du système d'équations normales :

$$X'X\hat{\rho} = X'T$$

Pour trouver un estimateur parmi les solutions de ce système on impose à ces solutions une contrainte d'identification, en général une contrainte linéaire, i.e. : $c'\hat{\rho} = 0$ où c est de dimension $(2(a+p)-3, 1)$. La contrainte $c'\rho = 0$ est une contrainte d'identification si pour tout θ de $\text{Im}(X)$ il existe un unique ρ tel que l'on ait : $c'\rho = 0$ et $\theta = X\rho$.

Une condition nécessaire et suffisante pour que la contrainte : $c'\rho=0$ soit une contrainte d'identification est : $c'v \neq 0$.

A la contrainte d'identification : $c'\hat{\rho} = 0$ imposée aux solutions du système correspond la solution :

$$\hat{\rho}_c = (X'X + cc')^{-1}X'T$$

Puisque $Xv = 0$ on a $(X'X + cc')v = cc'v$ et $(X'X + cc')^{-1}c = \frac{1}{c'v}v$

d'où $E(\hat{\rho}_c) = \rho - \frac{c'\rho}{c'v}v$;

on en déduit que $\hat{\rho}_c$ est un estimateur sans biais de ρ si et seulement si le vecteur ρ des effets vérifie : $c'\rho = 0$ (Kupper et al., 1983).

Dans le cas où $c = v$ on a $\hat{\rho}_v = (X'X + vv')^{-1}X'T$ quantité égale à $(X'X)^+ X'T$ où $(X'X)^+ (\neq (X'X + vv')^{-1})$ est l'inverse généralisé de Moore-Penrose de $X'X$. On a la relation : $\hat{\rho}_c = \hat{\rho}_v + \lambda v$ avec $\lambda = v'\hat{\rho}_c/v'v$.

2.3. Le problème d'identification

L'estimation des effets A.P.C. dépend donc fortement du choix de c . Cependant le problème d'identification concerne le modèle et ne peut être résolu à partir des données. En particulier lorsqu'on teste la nullité des effets d'un facteur on teste en réalité le fait que les effets sont dans une relation affine donnée par les composantes de v correspondant à ce facteur.

Malgré la difficulté d'utilisation des modèles A.P.C., due au problème d'identification, ces modèles ont un pouvoir explicatif plus grand que les modèles A.P., A.C. et P.C. pris séparément et ils apportent un élément de réponse au paradoxe signalé dans l'introduction et qui a motivé leur utilisation. En 1981, Mason et Smith obtiennent, en utilisant une contrainte justifiée à partir d'une connaissance a priori du problème, des résultats permettant de bien comprendre le mécanisme du déclin de la mortalité par tuberculose pulmonaire au Massachussets et aux Etats-Unis. Cependant, pour les Etats-Unis en particulier, certaines contraintes que l'on pense a priori bien justifiées ne donnent pas les résultats escomptés. En fait se posent ici deux problèmes : celui de la limitation des types de contraintes (on choisit généralement la nullité d'un effet ou l'égalité entre deux effets d'un même facteur) et celui du choix d'une contrainte parmi plusieurs contraintes; dans ce dernier cas quelle est la crédibilité des résultats obtenus avec une contrainte particulière? Rodgers (1982) montre que le fait d'imposer plusieurs contraintes simultanément ne résoud pas le problème de l'estimation des effets A.P.C.. Heckman et Robb (1985) d'une part, Jöreskog et Sörbom (1985) d'autre part proposent une approche par des variables latentes dans laquelle les facteurs Age, Période et Cohorte sont remplacés par des variables inobservables linéairement indépendantes. Cependant ce point de vue exige un gros potentiel d'informations sur le phénomène.

De nombreux auteurs, étant donné les succès obtenus par des analyses A.P.C., proposent de ne pas s'arrêter au problème d'identification, de poursuivre l'étude de tels modèles et de leurs généralisations. En particulier, t_i^j peut être remplacé dans le modèle par une fonction de t_i^j ; par exemple $f(t_i^j) = \ln(t_i^j)$, $f(t_i^j) = \ln(1 + t_i^j)$ ou $f(t_i^j) = \ln(t_i^j / (1 - t_i^j))$. Fienberg et Mason (1985) proposent des modèles polynomiaux pour obtenir des interactions d'ordre plus élevés. Nous renvoyons à l'article de Kupper et al. (1983) pour une plus ample bibliographie. Avant de proposer un modèle de quasi-vraisemblance pour les analyses A.P.C. (§3), nous proposons, dans le cas où on ne connaît pas de contrainte sur les paramètres, deux critères pour le choix d'un estimateur

dans l'ensemble des solutions possibles. Nous reviendrons dans le paragraphe 4 sur le problème d'identification.

2.4. Cas où on ne connaît pas de contrainte sur ρ

Soit $\mathcal{M}_{\mathbb{R}}(m,n)$ l'espace des matrices de dimension (m,n) à coefficients réels et pour tout N de $\mathcal{M}_{\mathbb{R}}(m,n)$ la norme définie par $\|N\| = (\text{tr}(NN'))^{1/2}$. On rappelle les propriétés caractéristiques suivantes des inverses généralisés de Moore-Penrose (cf. par exemple : Fourgeaud et Lenclud, 1978) :

soit $A \in \mathcal{M}_{\mathbb{R}}(m,n)$

$$A^+ = \arg \inf \{ \|G\| / G \in \mathcal{M}_{\mathbb{R}}(n,m) \text{ et } \|AG - I_m\| \text{ minimum} \}$$

$$A^+ = \arg \inf \{ \|G\| / G \in \mathcal{M}_{\mathbb{R}}(n,m) \text{ et } \|GA - I_n\| \text{ minimum} \}$$

Soit $E(T) = X\rho$ avec X de plein rang moins 1

On suppose que $\text{Var}(T) = \sigma^2 I$

a) Estimateur de variance minimale dans la classe des estimateurs de biais relatif minimal

Soit $B=NT$ un estimateur de ρ ; on a :

$$\|E(B) - \rho\| = \|(NX - I)\rho\| \leq \|NX - I\| \|\rho\|, \text{ d'où la définition :}$$

parmi les estimateurs de ρ de la forme $B=NT$, $A (=MT)$ est de biais relatif minimal si $\|MX - I\| = \min_N \|NX - I\|$; A est de variance minimale si $\text{tr}(\text{var } A) \leq \text{tr}(\text{var } (B))$.

On a $E(\hat{\rho}_v) = E((X'X)^+X'T) = X^+X\rho$ car $(X'X)^+X' = X^+$, inverse généralisée de Moore Penrose de X , et $\text{var}(\hat{\rho}_v) = \rho^2(X'X)^+$. D'après la 2^e propriété caractéristique de X^+ rappelée ci-dessus on a, d'une part, $\|X^+X - I\| = \min_N \|NX - I\|$, donc $\hat{\rho}_v$ est de biais relatif minimal, d'autre part, parmi les matrices N telles que $\|NX - I\| = \|X^+X - I\|$, $\|X^+\| \leq \|N\|$ donc

$$\text{tr}(\text{var}(\hat{\rho}_v)) = \sigma^2 \|X^+\|^2 \leq \sigma^2 \|N\|^2 = \text{tr}(\text{var}(NT)).$$

L'estimateur $\hat{\rho}_v$ est donc de variance minimale dans la classe des estimateurs de biais relatif minimal. C'est l'estimateur que l'on obtient si on a la contrainte sur ρ : $v'\rho = 0$. On montre de plus que, parmi les estimateurs obtenus par une contrainte d'identification de la forme : $c'\rho = 0$, c'est-à-dire parmi les estimateurs de la forme NT avec $N = (X'X + cc')^{-1}X'$, c'est le seul, à une constante multiplicative près, qui appartienne à la classe des estimateurs de biais relatif minimum.

On a en effet : $E(B) - \rho = (NX - I)\rho = -\frac{c'\rho}{c'v}v$ et $\|NX - I\| = \|\frac{vc'}{c'v}\|$
 $= \frac{1}{|\cos(v,c)|}$ donc $\|NX - I\|$ est minimum si et seulement si il existe $\lambda \in \mathbb{R}^*$ tel que $c = \lambda v$.

b) Estimateur d'erreur quadratique moyenne relative minimale

Soit $B=NT$ un estimateur de ρ , on a

$$\begin{aligned} E(\|B - \rho\|^2) &= \|E(B) - \rho\|^2 + \text{tr}(\text{var}(B)) = \|(NX - I)\rho\|^2 + \sigma^2 \text{tr}(NN') \\ &\leq \|NX - I\|^2 \|\rho\|^2 + \sigma^2 \text{tr}(NN') \end{aligned}$$

On cherche à définir un critère d'erreur quadratique moyenne minimale de façon analogue à ce qui a été fait précédemment.

On pose $f(N, \rho) = \|NX - I\|^2 \|\rho\|^2 + \sigma^2 \text{tr}(NN')$; a priori on peut choisir le critère $\min_N f(N, \rho)$ c'est-à-dire on cherche M tel que $f(M, \rho) \leq f(N, \rho)$ pour tout ρ .

Le problème ainsi posé n'a pas de solution; on a en effet le théorème suivant : (Fourgeaud et Lenclud, 1978) :

Soit α et β des réels strictement positifs de somme égale à 1, l'unique solution du problème : $\min_N \{\alpha \|NX - I\|^2 + \beta \text{tr}(NN')\}$ est :

$$M = (X'X + h I_\rho)^{-1} X' \text{ avec } h = \beta/\alpha$$

Il est clair que si $\alpha + \beta \neq 1$, on est conduit à la même solution. Ce résultat appliqué à notre problème donne pour estimateur de ρ : $M_\rho T$ avec $M_\rho = (X'X + \sigma^2 / \|\rho\|^2 I)^{-1} X'T$, estimateur qui dépend de ρ et ne répond pas au problème posé.

Le critère proposé ci-dessus n'a pas de solution; on propose alors la définition suivante :

parmi les estimateurs de ρ de la forme $B=NT$, $A(=MT)$ est d'erreur quadratique moyenne relative minimale si on a :

$$M = \arg \min_N f(N, \rho)$$

Le minimum est alors atteint pour M_{ρ_v} ; on note $\hat{\rho}_{EQM}(= M_{\rho_v} T)$ l'estimateur correspondant de ρ :

$$\hat{\rho}_{EQM} = (X'X + \sigma^2 / \|\rho_v\|^2 I)^{-1} X'T.$$

L'estimateur $\hat{\rho}_{EQM}$ est un estimateur "ridge" (Hoerl et Kennard, 1970); on montre (cf. Fourgeaud et Lenclud) que l'inverse généralisé de Moore Penrose de X vérifie la propriété suivante :

$$X^+ = \lim_{\epsilon \rightarrow 0_+} (X'X + \epsilon I)^{-1} X'$$

aussi lorsque σ^2 est petit l'estimateur $\hat{\rho}_{EQM}(= (X'X + \sigma^2 / \|\rho_v\|^2 I)^{-1} X'T)$ et l'estimateur $\hat{\rho}_v(= (X'X)^+ X'T = X^+ T)$ sont proches.

Si l'on suppose que $\text{Var}(T) = \sigma^2 V$ où V est une matrice régulière, soit P tel que $PP' = V$, $U = P^{-1}T$, $Y = P^{-1}X$ alors $E(U) = Y\rho$ avec Y de plein rang moins 1 et $\text{Var}(U) = \sigma^2 I$, on est ramené au modèle précédent. Les estimateurs de variance minimale dans la classe des estimateurs de biais relatif minimal et d'erreur quadratique moyenne relative minimale sont respectivement :

$$\hat{\rho}_v = (X'V^{-1}X + vv')^{-1}X'V^{-1}T \text{ et } \hat{\rho}_{EQM} = (X'V^{-1}X + \sigma^2 / \|\rho_v\|^2 I)^{-1}X'V^{-1}T$$

3. Modèle de quasi-vraisemblance pour le modèle A.P.C.

3.1. Modèle

Soit Ω la population considérée et A^j l'ensemble des individus de Ω vérifiant le phénomène démographique à la période j , $j=1, \dots, p$. Les cohortes, notées Ω_k , $k=1, \dots, a+p-1$, forment une partition de Ω . Pour simplifier l'exposé on suppose que les cohortes sont de même taille et que l'échantillon sur lequel sont calculés les taux est composé de $a+p-1$ échantillons indépendants de taille N . Alors $N t_i^j$ suit une loi binomiale de paramètres N et m_i^j où $m_i^j = P(A^j / \Omega_{a-i+j})$. On a donc $E(t_i^j) = m_i^j$ et $\text{Var}(t_i^j) = \frac{1}{N} m_i^j (1 - m_i^j)$. On peut montrer de plus que t_i^j et $t_{i'}^{j'}$ sont indépendantes si $i - j \neq i' - j'$ c'est-à-dire si les taux concernent des cohortes différentes; sinon

$$\text{cov}(t_i^j, t_{i'}^{j'}) = \frac{1}{N} (P(A^j \cap A^{j'} / \Omega_{a-i+j}) - m_i^j m_{i'}^{j'}).$$

Nous renvoyons à Fotso (1988) pour une explicitation du modèle probabiliste et de l'échantillonnage utilisés.

Tous les éléments de la matrice de covariance de T ne peuvent pas être estimés directement par les t_i^j . C'est le cas de $P(A^j \cap A^{j'} / \Omega_{a-i+j})$; c'est la probabilité qu'un individu de la cohorte $a-i+j$ vérifie le phénomène aux périodes j et j' . Cette probabilité peut être correctement estimée dans certains cas (par exemple si les données résultent d'enquêtes sur panel ou d'enquêtes rétrospectives). Si l'on ne dispose que des taux t_i^j nous proposons d'écrire cette probabilité en fonction de t_i^j et $t_{i'}^{j'}$ (avec $i - j = i' - j'$) en utilisant un moyen terme entre le maximum et le minimum d'entrées ou de sorties (par rapport au phénomène) entre les deux périodes (cf. Fotso 1988). Nous admettons donc dans la suite que tous les éléments de la matrice de covariance de T (supposée définie positive ce qui n'est pas très restrictif) sont des fonctions connues des m_i^j .

Le modèle considéré est alors le suivant, noté (M) :

$$\begin{aligned} E(T) &= \mu = X\rho \\ V(T) &= V_\mu / N \end{aligned}$$

où X est une matrice de plein rang moins un dont on sait expliciter un vecteur v engendrant le noyau de X et V_μ est une matrice dont les éléments sont des fonctions connues de μ .

En fait, on connaît la loi des éléments de T , cependant, la matrice V_μ n'étant pas diagonale, il est difficile d'écrire la fonction de vraisemblance; aussi, nous utilisons la notion de quasi-vraisemblance qui nécessite seulement la connaissance des moments d'ordre un et deux des variables et des relations fonctionnelles qui existent entre eux (cf. Mc Cullagh, 1983).

3.2. Modèle de quasi-vraisemblance

La notion de quasi-vraisemblance a été définie par Wedderburn en 1974 dans le cas unidimensionnel. Im en 1982 l'étend au cas multidimensionnel et étudie quelques propriétés et les relations avec la vraisemblance. Indépendamment, Mc Cullagh en 1983 propose la notion de quasi-vraisemblance dans le cas multidimensionnel pour l'étude de modèles plus généraux que les modèles linéaires. Nous présentons ici la notion de quasi-vraisemblance sous la forme qui nous est utile. On considère le modèle suivant :

$E(T) = \mu = X\rho$ avec $T(n,1)$ aléatoire et $X(n,p)$ de rang $p (\leq n)$.

$V(T) = \sigma^2 V_\mu$ avec V_μ régulière; les éléments de V_μ sont des fonctions connues de μ ; σ^2 réel positif inconnu.

Pour alléger l'écriture V_μ est notée V dans la suite de ce paragraphe. Nous ne nous intéressons pas à l'estimation de σ puisque dans le modèle (M) σ^2 est connu ($= 1/N$).

3.2.1. Fonction de quasi-vraisemblance

Le logarithme de la quasi-vraisemblance du modèle est défini, de manière implicite, par le système d'équations différentielles :

$$\frac{\partial \ell(T; \mu)}{\partial \mu} = V^{-1} (T - \mu)$$

Comme on a : $\mu = X\rho$, on a aussi :

$$U_\rho = \frac{\partial \ell}{\partial \rho} = \frac{\partial \ell}{\partial \mu} \cdot \frac{\partial \mu}{\partial \rho} = X'V^{-1} (T - \mu).$$

La fonction U ainsi définie est appelée la fonction score.

Par analogie à la notion liée à la vraisemblance, une solution $\hat{\rho}$ de l'équation : $U_\rho = 0$, est appelée estimateur de maximum de quasi-vraisemblance (M.Q.V.). Une explicitation de la quasi-vraisemblance ℓ est inutile pour trouver un tel estimateur.

3.2.2. Propriétés asymptotiques

On note $J_\rho = \frac{\partial^2 \ell}{\partial \rho^2}$ et $i_\rho = E(J_\rho) = X'V^{-1}X$

Si T admet des moments d'ordre 3 et si la suite de terme général $n^{-1}i\rho$ a une limite D_ρ définie positive lorsque n tend vers l'infini alors on a les résultats suivants (cf. Mc Cullagh 1983) :

– la loi de $n^{-1/2} U_\rho$ est de la forme $N_\rho (0, n^{-1} \sigma^2 i_\rho) + 0 (n^{-1/2})$ où $0 (n^{-1/2})$ désigne l'ordre de grandeur de la différence entre les fonctions de répartition de $n^{-1/2} U_\rho$ et de la loi normale $N_\rho (0, n^{-1} \sigma^2 I_\rho)$.

– $n^{-1/2} U_\rho$ converge en loi lorsque n tend vers l'infini vers la loi $N_\rho (0, \sigma^2 D_\rho)$

– $n^{-1} \mathcal{J}_\rho$ converge en probabilité lorsque n tend vers l'infini vers D_ρ

– il existe une suite $(\hat{\rho}_n)_n$ d'estimateurs M.Q.V. telle que :

(i) $E(\hat{\rho}_n) = \rho + 0 (n^{-1})$

(ii) $\hat{\rho}_n$ converge en probabilité lorsque n tend vers l'infini vers ρ .

(iii) la loi de $\sqrt{n}(\hat{\rho}_n - \rho)$ est $N_p (0, \sigma^2 n i_\rho^{-1}) + 0 (n^{-1/2})$

(iv) $\sqrt{n}(\hat{\rho}_n - \rho)$ converge en loi lorsque n tend vers l'infini vers la loi $N_\rho (0, \sigma^2 D_\rho^{-1})$.

3.2.3. Détermination de l'estimateur M.Q.V.

Pour résoudre l'équation : $U_\rho = 0$, on utilise la méthode de Newton-Raphson modifiée dans laquelle la matrice jacobienne de $U (=U'_\rho)$ est remplacée par son espérance mathématique $-X'V^{-1}X$. Ceci conduit à l'algorithme suivant : partant de ρ_0 suffisamment proche de la solution $\hat{\rho}$ de $U_\rho = 0$ et supposant les i premiers itérés calculés, le $(i+1)$ -ème, ρ_{i+1} , est obtenu comme solution du système : $X'V_{\rho_i}^{-1} X\rho_{i+1} = X'V_{\rho_i}^{-1} T$. Comme X est de plein rang on a :

$\rho_{i+1} = (X'V_{\rho_i}^{-1}X)^{-1}X' V_{\rho_i}^{-1} T$ et la suite de terme général ρ_i converge vers $\hat{\rho}$. On pourra choisir pour ρ_0 l'estimateur des moindres carrés ordinaires.

3.2.4. Cas où X est de rang $p-1$

On suppose maintenant que X est de rang $(p-1)$ et que pour tout n le noyau de X est engendré par un même vecteur v

a) Cas d'une contrainte connue sur ρ

Si on a : $c'\rho = 0$ alors que le modèle est équivalent à :

$$\begin{aligned} E(T_1) &= \mu_1 = X_1\rho \\ \text{Var}(T_1) &= \sigma^2 W_{\mu_1} \end{aligned}$$

où $T_1 = \begin{pmatrix} T \\ 0 \end{pmatrix}$, $X_1 = \begin{pmatrix} X \\ c' \end{pmatrix}$ et $W_{\mu_1} = \begin{pmatrix} V_\mu & 0 \\ 0 & 0 \end{pmatrix}$ sont respectivement de dimension $(n+1, 1)$, $(n+1, p)$ et $(n+1, n+1)$ et X_1 de rang p.

La matrice W_{μ_1} n'est pas inversible; cependant on peut définir la notion de quasi-vraisemblance avec un inverse généralisé. Si on pose $W_{\mu_1}^- = \begin{pmatrix} V_\mu^{-1} & 0 \\ 0 & 0 \end{pmatrix}$ alors

on définit le logarithme de la quasi-vraisemblance du modèle par :

$$\frac{\partial \ell}{\partial \mu_1} = W_{\mu_1}^{-1} (T_1 - \mu_1) = \begin{pmatrix} V_{\mu}^{-1} & (T - \mu) \\ 0 & \end{pmatrix};$$

la fonction de score est alors : $U_{\rho} = X_1' W_{\mu_1}^{-1} (T_1 - \mu_1) = X' V_{\mu}^{-1} (T - \mu)$; l'estimateur M.Q.V. de ρ vérifie : $U(\hat{\rho}) = 0$ et $c' \hat{\rho} = 0$. On utilise les mêmes notations qu'en 3.2.2. On a alors les mêmes propriétés asymptotiques; cependant, i_{ρ} n'est pas inversible et i_{ρ}^{-1} est remplacé par : $(X' V^{-1} X + c c')^{-1} X' V^{-1} X (X' V^{-1} X + c c')^{-1} (= (X' V^{-1} X)^+$ dans le cas où $c=v$).

On détermine alors l'estimateur M.Q.V. de ρ par l'algorithme suivant : partant de ρ_0 proche de $\hat{\rho}$, le (i+1)-ème itéré est : $\rho_{i+1} = (X' V_{\rho_i}^{-1} X + c c')^{-1} X' V_{\rho_i}^{-1} T$ et la suite de terme général ρ_i converge vers $\hat{\rho}$.

b) Pas de contrainte connue sur ρ

Dans ce cas, compte tenu de l'étude faite en 2.4. a) nous proposons l'estimateur correspondant à la contrainte $v' \rho = 0$. On a alors les propriétés asymptotiques décrites précédemment en remplaçant ρ par ρ_v . Cependant, on peut réécrire les propriétés en fonction de ρ en utilisant la relation : $\rho_v = \rho - \frac{v' \rho}{v' v} v$.

Un estimateur "ridge" calculé à l'aide d'une procédure itérative, noté $\hat{\rho}_{EQM}$, peut permettre d'améliorer l'estimateur précédent, noté $\hat{\rho}_v$, dans le sens d'une réduction de l'erreur quadratique moyenne; cet estimateur est alors la limite de la suite :

$$\rho_0 = \hat{\rho}_v; \rho_{i+1} = (X' V_{\rho_i}^{-1} X + (1/N) \| \rho_i \|^2 I)^{-1} X' V_{\rho_i}^{-1} T$$

3.3. Application au modèle A.P.C.

On considère le modèle (M) défini en 3.1.

a) Cas d'une contrainte connue sur ρ

Si on a : $c' \rho = 0$ on propose alors l'estimateur $\hat{\rho}_c$ limite de la suite :

$$\rho_0 = (X' X + c c')^{-1} X' T; \rho_{i+1} = (X' V_{\rho_i}^{-1} X + c c')^{-1} X' V_{\rho_i}^{-1} T.$$

On a alors :

$$E(\hat{\rho}_c) = \rho + 0 (1/ap);$$

$$\text{Var}(\hat{\rho}_c) = N^{-1} (X' V_{\hat{\rho}_c}^{-1} X + c c')^{-1} X' V_{\hat{\rho}_c}^{-1} X (X' V_{\hat{\rho}_c}^{-1} X + c c')^{-1} + 0 (1/ap)$$

b) Pas de contrainte connue sur ρ

On propose alors les deux estimateurs suivants :

$\hat{\rho}_v$, c'est-à-dire le cas a) avec $c=v$; on a alors : $E(\hat{\rho}_v) = \rho + 0 (1/ap)$ et $\text{Var}(\hat{\rho}_v) = N^{-1} (X' V_{\hat{\rho}_v}^{-1} X)^+ + 0 (1/ap)$

$\hat{\rho}_{EQM}$, c'est-à-dire la limite de la suite :

$$\rho_0 = \hat{\rho}_v; \rho_{i+1} = (X'V_{\rho_i}^{-1}X + (1/N) \|\rho_i\|^2 I)^{-1} X'V_{\rho_i}^{-1}T$$

L'amélioration de $\hat{\rho}_v$ par $\hat{\rho}_{EQM}$ n'est sensible que si N n'est pas trop grand.

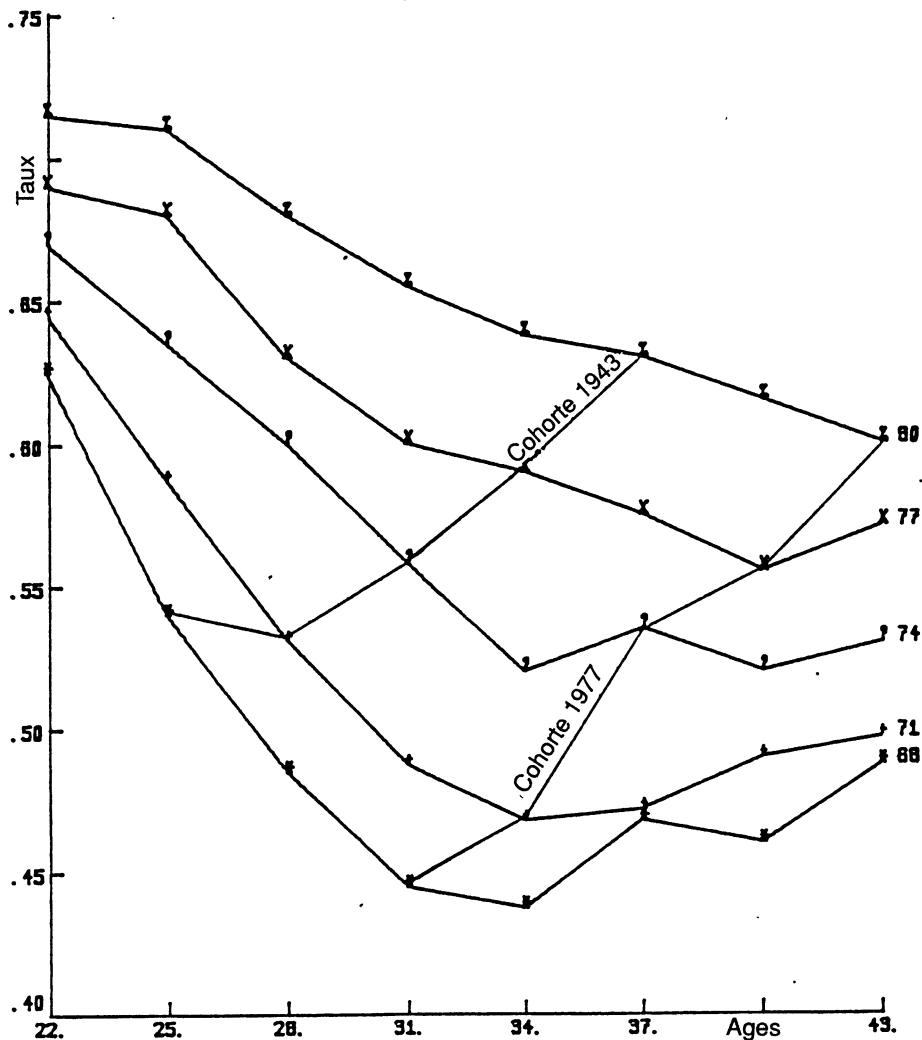
On ne peut cependant pas utiliser directement les résultats asymptotiques du §3.2.4. car la matrice X de dimension (n,p) dans le §3.2.4 est ici de dimension $(p, 2(a+p)-3)$ aussi, lorsque n tend vers l'infini, p tend aussi vers l'infini. On aborde ici un autre problème : celui de l'effet de la discrétisation suivant les âges et les périodes.

3.4. Exemple

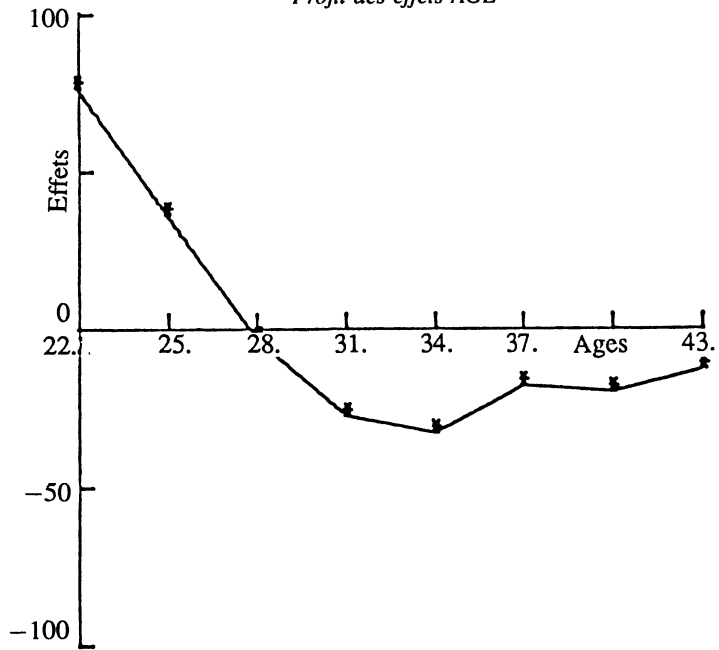
L'exemple suivant concerne les taux par âge d'activité des femmes de 1968 à 1980 en France, exemple extrait de l'article de Zighera (1982).

	1968	1971	1974	1977	1980
22	0.62	0.65	0.67	0.69	0.71
25	0.54	0.59	0.64	0.68	0.71
28	0.48	0.53	0.60	0.63	0.68
31	0.44	0.49	0.56	0.60	0.65
34	0.44	0.47	0.52	0.59	0.64
37	0.47	0.47	0.53	0.57	0.63
40	0.46	0.49	0.52	0.55	0.61
43	0.49	0.50	0.53	0.57	0.60

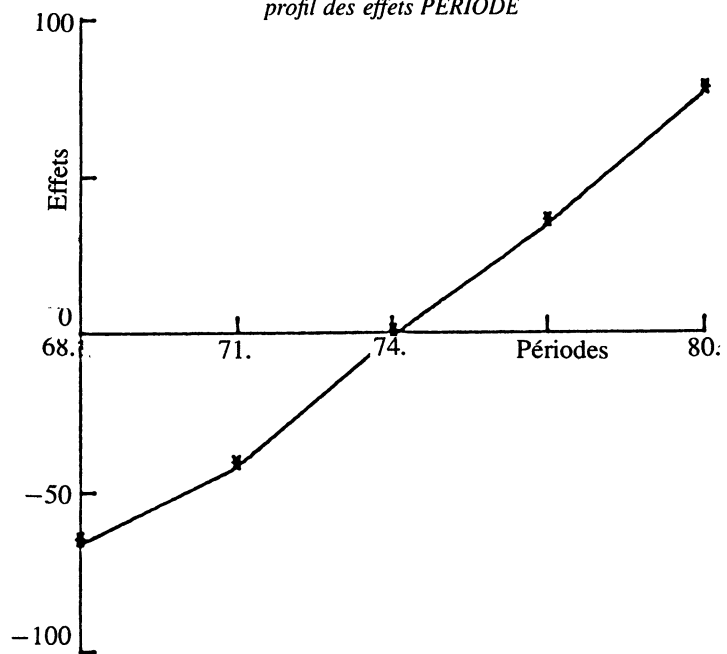
Graphique 1
 Profil des périodes



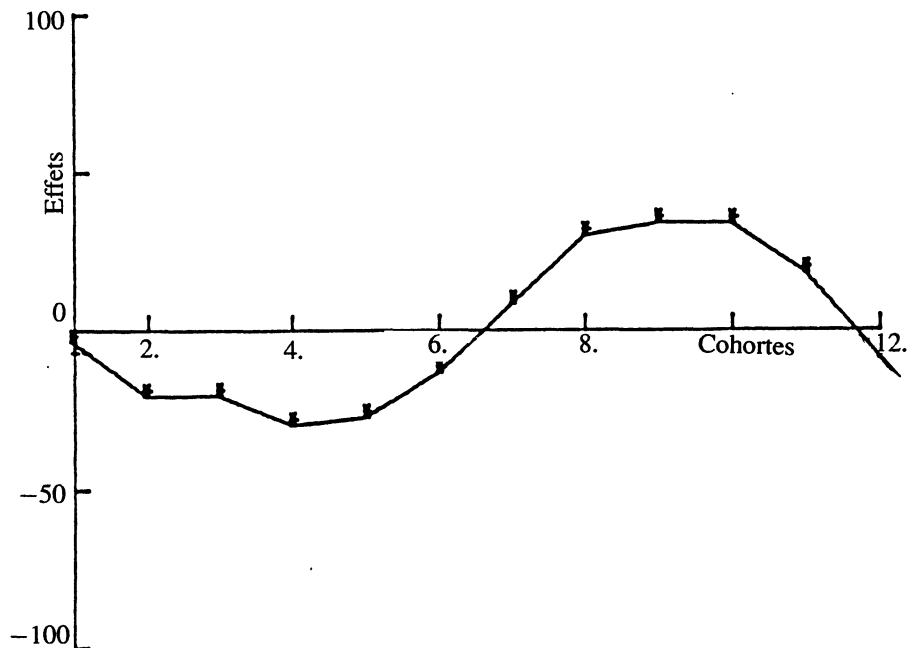
Graphique 2
Profil des effets AGE



Graphique 3
profil des effets PERIODE



Graphique 4
Profil des effets COHORTE



On remarque, graphique 1, que pour chaque période les taux diminuent avec l'âge. Une mauvaise interprétation du graphique conduit à penser que les femmes ont un taux d'activité qui diminue avec l'âge. Or, pour chaque cohorte, le taux d'activité des femmes augmente. Sur le graphique seules les cohortes 1943 et 1937 ont été représentées. Ce paradoxe est maintenant bien connu des démographes et a rendu nécessaire l'introduction des effets de cohortes dans les modèles analysant de tels tableaux.

L'analyse A.P.C. de ces données conduit aux résultats suivants représentés par les graphiques 2, 3 et 4. On utilise l'estimateur $\hat{\rho}_v$ présenté au §. 3.3.b.

Effet général et effets des âges, périodes et cohortes ($\times 10^3$)

Effet général :

570

Effet des âges :

âges	22	25	28	31	34	37	40	43
effets	77	37	-3	-27	-33	-18	-20	-13

Effet des périodes :

périodes	1968	1971	1974	1977	1980
effets	-66	-43	-1	34	76

Effet des cohortes :

cohortes	1925	1928	1931	1934	1937	1940
effets	-4	-21	-21	-30	-27	-13
cohortes	1943	1946	1949	1952	1955	1958
effets	8	30	34	34	18	-8

Les effets des âges sont décroissants entre 22 et 31 ans puis relativement stables entre 31 et 43 ans; les effets des périodes croissent linéairement; quant aux effets de cohortes ils sont les plus faibles pour les cohortes 1928-1931-1934-1937, les plus élevés pour les cohortes 1946-1949-1952, ces effets diminuent pour les cohortes les plus jeunes 1955-1958.

L'amélioration apportée par l'estimateur M.Q.V. sur l'estimateur des moindres carrés ordinaires n'est pas sensible; en effet la matrice V dont les éléments diagonaux sont de la forme $p_i(1 - p_i)$ est presque proportionnelle à l'identité.

Il est possible de proposer d'autres lois pour les éléments de T (dans le cas où le phénomène est une maladie, des lois poissonniennes peuvent être envisagées) ou de travailler sur des fonctions de T .

4. A propos du problème d'identification

Nous reprenons le modèle A.P.C. tel qu'il est présenté au paragraphe 2.1. :

$E(T) = X\rho$ avec X de dimension $(ap, 2(a+p)-3)$ et de rang $2(a+p)-4$; il existe v , que l'on sait expliciter (cf. §2.1.), tel que $Xv=0$.

Les modèles $E(T)=X\rho$ et $E(T)=X(\rho + \lambda v)$, $\lambda \in \mathbb{R}$, sont indiscernables. Le problème d'identification vient du fait que l'on suppose qu'il existe un unique vecteur ρ , que l'on cherche à estimer au mieux, tel que l'on ait : $E(T)=X\rho$. Peut-être faut-il renoncer à cette "unicité". Les effets d'âges, de périodes et de cohortes sont inextricablement liés parce que les variables elles-mêmes sont liées. L'effet α_i d'une variable X_i sur une variable T est interprété comme l'accroissement de T lorsque X_i augmente d'une unité les autres variables explicatives restant constantes. Or, si les variables explicatives sont linéairement dépendantes, il n'est pas possible de faire varier X_i en laissant les autres variables fixes. Cette interprétation de l'effet n'est pas correcte dans le cas où les variables sont liées. Quelle est la définition, l'interprétation, la mesure des effets lorsque les variables sont liées? L'ensemble des solutions du modèle $E(T)=X\rho$ est : $\Theta = \{\rho_v + \lambda v \mid \lambda \in \mathbb{R}\}$ où ρ_v est tel que : $v'\rho_v = 0$. Dans ce qui suit nous tentons de convaincre le lecteur que la mesure la plus "naturelle" des effets d'âges,

de périodes et de cohortes, compte-tenu du fait que les variables sont liées par v , est fournie par ρ_v .

1. Arguments reposant sur le modèle

a) Nous considérons que pour tout élément $\rho (= \rho_v + \lambda v)$ de Θ le terme λv est une part due uniquement à la colinéarité. On a en effet :

$$\lim_{\lambda \rightarrow \infty} \frac{\rho_v + \lambda v}{\|\rho_v + \lambda v\|} = \frac{v}{\|v\|} \text{ et } \lim_{\lambda \rightarrow -\infty} \frac{\rho_v + \lambda v}{\|\rho_v + \lambda v\|} = - \frac{v}{\|v\|}$$

aussi, pour de grandes valeurs de $|\lambda|$, le rapport entre les effets est pratiquement celui existant entre les composantes de v . Comme ce vecteur ne dépend que de la matrice d'incidence X , c'est-à-dire uniquement des dimensions (a,p) du tableau $E(t)$, les éléments eux-mêmes sont négligés au profit d'un effet pervers du modèle. D'autre part, on peut remarquer que, plus on donne de grandes valeurs à $|\lambda|$, plus on donne du poids aux cohortes extrêmes (celles pour lesquelles on a le moins d'information!), le vecteur $\rho_v + \lambda v$ a alors une part importante due à la colinéarité qui est sans information pour le modèle. Les solutions admissibles sont à retenir parmi celles où $|\lambda|$ est petit et le vecteur ρ_v est le seul élément de Θ qui ne contienne aucune part due à la colinéarité.

b) Si l'on considère un tableau dont tous les éléments sont égaux à une même constante les effets A.P.C. devraient être nuls car l'effet moyen général prend en compte cette constante. Dans ce cas, le seul élément de Θ qui annule tous les effets d'âges, de périodes et de cohortes est ρ_v .

2. Arguments reposant sur l'estimation

Bien que le problème d'identification concerne le modèle et non l'estimation on propose ici des arguments en faveur de ρ_v reposant sur les propriétés de ρ_v (qui est un estimateur sans biais de ρ_v).

a) On a vu que, à défaut d'information sur le choix d'un élément ρ de Θ , l'estimateur de ρ de variance minimale dans la classe des estimateurs de biais relatif minimal est $\hat{\rho}_v$.

b) Toujours dans le cas où on n'a pas de contrainte connue sur ρ l'estimateur de régression en composantes principales de ρ est $\hat{\rho}_v$. Cette technique (cf. Hocking, 1976) a le même objectif que la ridge-régression. Dans le cas où la matrice du modèle est presque singulière, l'estimateur des moindres carrés a une forte variance. Si on suppose que certaines des plus petites valeurs propres de $X'X$ sont nulles et que l'on impose à la solution de la régression d'être orthogonale au noyau on obtient un estimateur de ρ biaisé mais de variance plus faible ce qui améliore l'erreur quadratique moyenne. Lorsque X n'est pas de plein rang la méthode se réduit à imposer la contrainte d'orthogonalité au noyau à la solution des équations normales. Dans le modèle A.P.C. l'estimateur en composantes principales de ρ est donc $\hat{\rho}_v$.

Bibliographie

- [1] S.E. FIENBERG et W.M. MASON (1985). – *Spécification and implementation of age, period and cohort models*. Cohort analysis in social research : beyond the identification problem, Springer-Verlag – 45-88.
- [2] S. FOTSO (1988). – *contribution à l'étude du modèle âge-période-cohorte*. (Doctorat de 3^e cycle, Université Paul Sabatier, Toulouse).
- [3] C. FOURGEAUD et B. LENCLUD (1978). – *Econométrie*. Presses Universitaires de France.
- [4] J. HECKMAN et R. ROBB (1985). – *Using longitudinal data to estimate age, period and cohort effects in earnings equations*. Cohort analysis in social research : beyond the identification problem, Springer-Verlag – 137-150.
- [5] R.R. HOCKING (1976) – *The analysis and selection of variables in linear regression*. Biometrics, 32 – 1-49.
- [6] A.E. HOERL et R.W. KENNARD (1970). – *Ridge regression : Biased estimation for non orthogonal problems*. Technometrics, 12 – 55-68.
- [7] S. IM (1982). – *Contribution à l'étude des tables de contingence à paramètres aléatoires*. Utilisation en biométrie. Thèse de 3^e cycle, U.P.S. Toulouse.
- [8] K.G. JÖRESKOG et D. SÖRBOM (1985). – *Simultaneous analysis of longitudinal data from several cohorts*. Cohort analysis in social research : beyond the identification problem, Springer-Verlag – 323-342.
- [9] L.L. KUPPER, J.M. JANIS, I.A. SALAMA, C.N. YOSHIZAWA et B.G. GREENBERG (1983). – *Age, Period, Cohort analysis : an illustration of the problems in assessing interaction in one observation per cell data*. Commun. Statist. Theor. Meth., 12, (23) – 2779-2807.
- [10] P. Mc CULLAGH et J.A. NELDER (1983). – *Generalized linear models*. London – New York, Chapman and Hall.
- [11] W.M. MASON et H.L. SMITH (1985). – *Age, Period, Cohort analysis and the study of deaths from pulmonary tuberculosis*. Cohort analysis in social research : beyond the identification problem, Springer-Verlag – 151-228.
- [12] W.L. RODGERS (1982). – *Estimable functions of age, period and cohort effects*. American Sociological Review, 47 – 774-787.
- [13] R.W.M. WEDDERBURN (1974). – *Quasi-likelihood functions, generalized linear models and the Gauss-Newton method*. Biometrika, 61, 3 – 439-447.
- [14] J.A. ZIGHERA (1982). – *Métiers et générations*. Economie et Statistique, n° 145.