Université de Montréal

Modélisation des risques d'accidents avec victimes de nouveaux conducteurs

par

Isabelle Morin

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

juin 2000

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Modélisation des risques d'accidents avec victimes de nouveaux conducteurs

présenté par

Isabelle Morin

a été évalué par un jury composé des personnes suivantes :

Yves Lepage
(président-rapporteur)

Urs Maag
(directeur de recherche)

Robert Cléroux
(membre du jury)

Mémoire accepté le :

31 août 2000

SOMMAIRE

Les études longitudinales jouent un rôle de plus en plus important en recherche, entre autre en sécurité routière. Par exemple, il arrive régulièrement que les accidents des conducteurs soient observés de façon répétée dans le temps. La théorie pour l'analyse de données longitudinales lorsque la réponse est discrète n'est pas encore consolidée. En effet, un certain nombre d'approches ont été introduites et sont encore en discussion dans la littérature. Ce mémoire est donc consacré à l'étude des extensions du modèle linéaire généralisé à l'analyse de données longitudinales de type dichotomique.

Nous proposons d'abord un survol de différentes méthodologies réparties en deux catégories: le modèle marginal et le modèle spécifique aux sujets. Nous nous intéressons ensuite plus particulièrement au modèle logistique-normal (approche spécifique aux sujets) en décrivant la théorie à la base de méthodes permettant de maximiser la vraisemblance. Nous discutons entre autre de l'approche de la quasi-vraisemblance pénalisée (PQL), stratégie permettant d'éviter l'intégration en faisant une approximation des équations score. Puisque cette méthode mène à des résultats biaisés pour des réponses dichotomiques avec peu de répétitions dans le temps, nous décrivons une correction des estimateurs proposée par Lin et Breslow (1996a). Nous étudions finalement l'approximation numérique qui consiste à évaluer l'intégrale par la quadrature gaussienne adaptée (QGA).

Les performances des méthodes PQL, PQL corrigée et QGA sont évaluées à l'aide de simulations. Nous constatons que l'estimateur QGA de la variance

de l'effet aléatoire peut être loin de la vraie valeur pour de petits échantillons lorsque le nombre d'observations dans le temps n'est pas élevé. Nous proposons donc l'utilisation des estimateurs PQL corrigés lorsque le nombre d'individus est petit et l'utilisation des estimateurs QGA dans le cas contraire. Nous terminons ce mémoire en appliquant la méthodologie à des données réelles provenant du domaine de la sécurité routière.

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche, Monsieur Urs Maag, pour sa disponibilité, son enthousiasme et ses conseils judicieux. Il a su doser intelligemment liberté et soutien pour faire de mon mémoire une réussite. J'exprime également mes remerciements à Madame Claire Laberge-Nadeau pour l'accès aux données utilisées au chapitre 4 et pour le support financier qu'elle m'a octroyé pour l'analyse des banques de données de la SAAQ. Je voudrais souligner la participation de Denise qui m'a encouragée et m'a offert son aide pour la gestion de grosses banques de données.

Je ne saurai oublier mes parents qui m'ont soutenue de façon remarquable tout au long de mes études. Ils m'ont inculqué le goût d'apprendre et la détermination qui m'a permis de mener ce projet à terme. Finalement, je remercie spécialement Hugues pour son aide, sa patience, sa compréhension et surtout pour faire partie de ma vie.

Table des matières

LISTE DES ABRÉVIATIONS

Liste des abréviations dans le texte:

BLUE	:	meilleur estimateur sans biais,
CRT	:	Centre de recherche sur les transports de l'Université de Montréal,
EQM	:	erreur quadratique moyenne,
GEE	:	équations d'estimation généralisée,
MCGE	:	moindres carrés généralisés empiriques,
MLG	:	modèle linéaire généralisé,
MLGM	:	modèle linéaire généralisé à effets mixtes,
PL	:	pseudo-vraisemblance,
PQL	:	quasi-vraisemblance pénalisée,
QGA	:	quadrature gaussienne adaptée,
REPL	:	pseudo-vraisemblance restreinte,
SAAQ	:	Société de l'assurance automobile du Québec.

Liste des abréviations dans les tableaux:

LOGIT	:	régression logistique,
PQL	:	méthode PQL utilisant le maximum de vraisemblance restreint pour estimer les composantes de la variance,
CPQL0	:	méthode PQL utilisant les composantes de la variance corrigées (permet d'obtenir les valeurs initiales pour les corrections d'ordre 1 et 2),
CPQL1	:	correction d'ordre 1 des estimateurs PQL,
CPQL2	:	correction d'ordre 2 des estimateurs PQL.

MPQL : méthode PQL utilisant le maximum de vraisemblance pour estimer les composantes de la variance,

MCP0 : méthode MPQL utilisant les composantes de la variance corrigées (permet d'obtenir les valeurs initiales pour les corrections d'ordre 1 et 2),

MCP1 : correction d'ordre 1 des estimateurs MPQL,

MCP2 : correction d'ordre 2 des estimateurs MPQL,

QNR : méthode QGA utilisant l'algorithme de Newton-Raphson,

QQN : méthode QGA utilisant l'algorithme de quasi-Newton.

Table des figures

Liste des tableaux

INTRODUCTION

Les études longitudinales, impliquant plusieurs réponses dans le temps pour un même individu, jouent un rôle important en sécurité routière. De telles données permettent d'examiner les changements dans le temps d'une caractéristique qui est mesurée de façon répétée pour chacun des participants de l'étude. De plus, les sujets peuvent généralement être considérés comme indépendants. L'analyse de données longitudinales requiert des méthodes statistiques spéciales puisque les observations pour un même sujet ont tendance à être dépendante. En effet, il faut tenir compte de la corrélation dans le temps pour que l'inférence statistique et l'estimation des coefficients soient valides. Notons qu'en sciences sociales et économiques, le terme études de panels est utilisé pour désigner les études de type longitudinal.

L'importance pratique de développer des méthodes appropriées à l'analyse de mesures répétées est reconnue depuis longtemps en recherche. De plus, la popularité de tels modèles s'est considérablement accrue au cours des dernières années grâce à la puissance des nouveaux ordinateurs. Notons qu'il n'y a pas eu de développements récents de la théorie statistique pour l'analyse de modèles paramétriques linéaires à mesures répétées sous l'hypothèse de normalité des données. Par contre, la méthodologie n'est pas encore consolidée pour les données longitudinales lorsque la réponse est discrète. En effet, un certain nombre d'approches ont été introduites depuis le début des années 80 (Stiralli, Laird et Ware (1984),

Anderson et Aitkin (1985), Zeger, Liang et Albert (1988),...) et sont encore en discussion dans la littérature.

Nous proposons d'étudier des extensions du modèle linéaire généralisé proposé par Nelder et Wedderburn en 1972 à l'analyse de données longitudinales de type dichotomique. Notons que contrairement au cas linéaire, les modèles pour les données discrètes comme la régression logistique peuvent mener à des interprétations distinctes selon les hypothèses posées au départ sur la source de corrélation. Il faut donc définir soigneusement les objectifs de l'analyse au moment de choisir une approche particulière. Les deux principales approches examinées sont le modèle marginal et le modèle spécifique aux sujets.

La première stratégie consiste à modéliser la moyenne marginale comme c'est le cas pour les études transversales ("cross-sectional"). Dans le cas dichotomique, l'espérance marginale dépend des variables explicatives par une fonction de lien logit. L'analyse marginale inclut des hypothèses sur la forme de la corrélation et modélise séparément la moyenne et la covariance. Ainsi, seul les deux premiers moments de la réponse sont spécifiés. Avec des données gaussiennes, ces moments déterminent complètement la vraisemblance. Par contre, ce n'est pas le cas pour les autres membres de la famille des modèles linéaires généralisés, et une approche raisonnable est alors d'utiliser la méthode des équations d'estimation généralisée (GEE) proposée par Liang et Zeger (1986). Cette méthode est basée sur la fonction de quasi-vraisemblance décrite par Wedderburn (1974).

Le modèle spécifique aux sujets est justifié lorsque l'objectif est de faire de l'inférence sur les individus plutôt que sur la moyenne. Nous supposons alors qu'il y a de l'hétérogénéité naturelle entre les sujets en ce qui a trait à leur coefficient de régression. Cette hétérogénéité peut être représentée par une distribution de probabilité. La corrélation parmi les observations pour une personne provient

donc d'une variable aléatoire non observable commune. Puisqu'un paramètre aléatoire supplémentaire est ajouté au modèle, la maximisation de la vraisemblance requiert des techniques d'intégration numérique particulières telle que la quadrature gaussienne adaptée (QGA) . Une stratégie alternative est d´éviter l'intégration en utilisant une approximation appelée quasi-vraisemblance pénalisée (PQL). Cette approche a l'avantage de permettre des modèles plus complexes, mais elle donne des estimateurs biaisés lorsque la réponse est de type dichotomique. Afin de pallier le problème, Lin et Breslow (1996a) propose une correction du biais des estimateurs PQL.

Au Laboratoire sur la sécurité des transports du Centre de recherche sur les transports (CRT) de l'Université de Montréal, on utilise la méthode du GEE lorsque le modèle marginal tient. Bien que la théorie à la base de tels modèles soit complexe et récente, il est possible de l'appliquer grâce aux nouveaux développements de la procédure GENMOD du progiciel SAS. Par contre, il existe aujourd'hui peu de logiciels statistiques qui offrent la possibilité d'analyser directement le modèle à effets aléatoires pour des réponses de type binaire. Il serait donc très intéressant pour le CRT de connaître les nouveaux développements de la méthodologie des modèles à effets aléatoires et de pouvoir appliquer ces méthodes à l'aide d'un programme approprié. Ainsi, bien que le modèle marginal soit abordé brièvement au cours du chapitre 1, la présente recherche discute essentiellement du modèle aléatoire pour la régression logistique (approche spécifique aux sujets).

Au chapitre 1, nous présentons brièvement deux extensions du modèle linéaire généralisé basées sur les modèles marginal et spécifique aux sujets. Plus particulièrement, l'interprétation et la comparaison des paramètres de ces deux approches sont examinées. Le modèle spécifique aux sujets est discuté en détail au chapitre

2 en insistant sur deux propositions permettant de maximiser une approximation de la fonction de vraisemblance : les méthodes PQL et QGA. Nous étudions également dans ce chapitre la correction proposée par Lin et Breslow (1996a) pour améliorer la performance des estimateurs PQL. Au chapitre 3, nous comparons entre eux les estimateurs PQL, PQL corrigés et QGA par des simulations selon trois modèles distincts. Le dernier chapitre est consacré à une illustration de l'approche spécifique aux sujets à l'aide de données en sécurité routière. Nous modélisons les risques d'accidents avec victimes de nouveaux conducteurs appartenant à deux banques de données. La première contient les 53 069 hommes de la banque de données analysée par Laberge-Nadeau *et al.* (1999) alors que la seconde est constituée des 3 550 conducteurs ayant eu des sanctions ou des retards de paiement de plus de 90 jours. Finalement, nous terminons ce mémoire par une conclusion qui permet d'exposer globalement les résultats et les recommandations de la recherche.

Chapitre 1

SURVOL DES MÉTHODES CONSIDÉRÉES

Lorsque les données sont de type longitudinal, le modèle linéaire généralisé (MLG) classique ne tient plus car il faut tenir compte de la corrélation dans le temps. La corrélation intra-sujet dans les problèmes de régression logistique a été l'objet d'études et de méthodes récentes. Celles-ci sont basées sur le modèle spécifique aux sujets (modèle à effets mixtes) et sur le modèle marginal (modèle de moyenne de population). Si de tels modèles sont considérés, il est important de bien choisir le plus approprié pour répondre aux questions scientifiques. En effet, dans le cas non linéaire, l'interprétation des paramètres diffère selon le modèle utilisé. Dans ce chapitre, nous discutons d'abord du MLG classique. Nous étudions entre autre l'inférence basée sur le maximum de vraisemblance et le maximum de quasi-vraisemblance. Nous exposons ensuite les approches du modèle marginal et du modèle spécifique aux sujets. Nous examinons finalement l'interprétation des paramètres de régression dans ces approches générales. Par cette démarche, nous voyons que pour des données binaires corrélées, les paramètres spécifiques aux sujets et ceux des modèles marginaux décrivent différents types d'effets des variables explicatives sur la réponse.

1.1. Inférence et modèles linéraires généralisés

Dans plusieurs situations pratiques, la variable réponse ne suit pas une distribution normale. La non-normalité est traditionnellement traitée par des transformations variées permettant d'appliquer les méthodes linéaires standards. Nelder et Wedderburn (1972) présentent une approche unifiée : le modèle linéaire généralisé (MLG). L'idée de base de cette approche est d'estimer les paramètres d'un modèle linéaire en utilisant le maximum de vraisemblance basé sur la distribution des données (McCullagh et Nelder 1989; Agresti 1990; Agresti 1996). Cette section présente le MLG classique et est essentiellement tirée des annexes de Diggle, Liang et Zeger (1994).

1.1.1. La vraisemblance maximale

En statistique, le but est souvent de modéliser un phénomène et de déterminer la statistique qui estime les paramètres inconnus du modèle. Nous discutons d'abord de l'inférence basée sur la vraisemblance maximale des paramètres pour des données observées. La fonction de densité des réponses observées, $\boldsymbol{y}$, étant donné un vecteur de paramètres inconnus, $\boldsymbol{\delta}$, est $f(\boldsymbol{y};\boldsymbol{\delta})$. Lorsque les données sont observées, la seule quantité inconnue est $\boldsymbol{\delta}$ d'où la fonction de vraisemblance de la forme

$$L(\boldsymbol{\delta}|\boldsymbol{y}) = f(\boldsymbol{y};\boldsymbol{\delta}).$$

L'estimateur du maximum de vraisemblance de $\boldsymbol{\delta}$ est la valeur $\hat{\boldsymbol{\delta}}$ qui maximise la fonction de vraisemblance ou son logarithme. Par conséquent, pour toute valeur $\boldsymbol{\delta}$,

$$L(\boldsymbol{\delta}|\boldsymbol{y}) \leq L(\hat{\boldsymbol{\delta}}|\boldsymbol{y}).$$

En pratique, $\hat{\boldsymbol{\delta}}$ est obtenu soit en maximisant directement $\log L$, soit en résolvant le système d'équations

$$\boldsymbol{S}(\boldsymbol{\delta}) = \frac{\partial \log L}{\partial \boldsymbol{\delta}} = \mathbf{0}.$$

La fonction $\boldsymbol{S}(\boldsymbol{\delta})$ est appelée fonction score de $\boldsymbol{\delta}$. L'algorithme de Newton-Raphson est parfois utilisé pour résoudre l´équation score lorsque celle-ci n'a pas de solution explicite.

L'estimateur du maximum de vraisemblance a plusieurs propriétés asymptotiques intéressantes. En particulier, sous des conditions de régularité pour f, $\hat{\boldsymbol{\delta}}$ est asymptotiquement sans biais et asymptotiquement efficace au sens où $\hat{\boldsymbol{\delta}}$ est l'estimateur non biaisé ayant la plus petite variance. La matrice de covariance asymptotique de $\hat{\boldsymbol{\delta}}$ est donnée par

$$\mathbf{V} = \left(-E\left(\frac{\partial^2 \log L}{\partial \boldsymbol{\delta}^2} \right) \right)^{-1}.$$

La matrice $\mathbf{V}^{-1}$ est appelée matrice d'information de Fisher de $\boldsymbol{\delta}$.

1.1.2. Le modèle linéaire généralisé

Les modèles linéraires généralisées (MLG) forment une vaste classe de modèles incluant la régression ordinaire, les modèles d'analyse de la variance ainsi que les modèles pour les variables réponses catégoriques. Un cas spécial important approprié lorsque les réponses sont binaires est le modèle de régression logistique. Un MLG comporte trois parties distinctes : une composante aléatoire, une composante systématique et le lien entre ces deux composantes. Il s'applique uniquement lorsque les données sont indépendantes. Appelons Y_i la variable aléatoire réponse pour un sujet i , $\boldsymbol{x}_i$ le vecteur des p variables explicatives associées à chacune des n unités expérimentales (sujets) et $\boldsymbol{\beta}$ le vecteur des paramètres de régression de dimensions $p \times 1$. Nous révisons ici les points saillants de la classe des MLG.

L'objectif principal du MLG est de décrire la dépendance entre la réponse moyenne $\mu_i = E(Y_i)$ et les variables explicatives. La réponse moyenne est reliée au vecteur de variables explicatives par une fonction de lien $h(.)$:

$$h(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta},$$

où $\boldsymbol{x}_i^T$ est le vecteur transposé du vecteur $\boldsymbol{x}_i$ et $\boldsymbol{x}_i^T \boldsymbol{\beta}$ la composante systématique.

De plus, la variance de Y_i est une fonction spécifique de la moyenne μ_i,

$$\text{Var}(Y_i) = v_i = \phi v(\mu_i).$$

Dans cette expression, la fonction $v(.)$ est connue, le paramètre d´échelle ϕ est une constante connue pour quelques membres de la famille des MLG et les autres paramètres doivent être estimés.

Chaque type de MLG correspond à un membre de la famille des distributions exponentielles (composante aléatoire) avec une fonction de vraisemblance de la forme

$$f(y_i) = exp[\{y_i\gamma_i - \psi(\gamma_i)\}/\phi + c(y_i, \phi)], \tag{1.1.1}$$

où γ_i est le paramètre naturel en relation avec μ_i par $\mu_i = \partial\psi(\gamma_i)/\partial\gamma_i$. Par exemple, la distribution binomiale est un cas spécial de la famille exponentielle avec

$$\gamma_i = \log\{\mu_i/(1-\mu_i)\},\ \psi(\gamma_i) = \log(1+e^{\gamma_i}),\ c(y_i, \phi) = \log \binom{n}{y_i},\ \phi = 1.$$

La famille exponentielle inclut les familles de distributions gaussienne, Poisson et gamma à deux paramètres. Les relations bien connues de la théorie de la vraisemblance maximale,

$$E\left(\frac{\partial f(y_i)}{\partial \gamma_i}\right) = 0 \qquad \text{et} \qquad E\left(\frac{\partial^2 f(y_i)}{\partial \gamma_i^2}\right) + E\left(\frac{\partial f(y_i)}{\partial \gamma_i}\right)^2 = 0,$$

permettent d´écrire les fonctions de moyenne et de variance suivantes

$$\mu_i = \frac{\partial \psi(\gamma_i)}{\partial \gamma_i} \qquad \text{et} \qquad \text{Var}(Y_i) = \frac{\partial^2 \psi(\gamma_i)}{\partial \gamma_i^2}\phi.$$

La régression logistique s'applique lorsque la variable expliquée est de type binaire, par exemple la présence ou l'absence d'une maladie. Le modèle logistique utilise la distribution Bernoulli pour modéliser les réponses dichotomiques. De plus, la fonction de lien est la transformation *logit*, le paramètre naturel de la distribution binomiale (lien canonique). Ainsi,

$$\log \frac{P(Y_i = 1)}{P(Y_i = 0)} = \log \frac{\mu_i}{1 - \mu_i} = \boldsymbol{x}_i^T \boldsymbol{\beta}.$$

Alors que l'estimateur de μ_i est restreint à l'intervalle $[0, 1]$, le *logit* peut prendre n'importe quelle valeur réelle. Le coefficient de régression représente le changement du logarithme de la cote de la variable réponse par changement d'unité des variables explicatives. Une autre caratéristique importante de la régression logistique est que pour des réponses binaires, la fonction de la variance est entièrement déterminée par la moyenne. Plus spécifiquement,

$$v(\mu_i) = \mu_i(1 - \mu_i).$$

Dans n'importe quel MLG, les coefficients de régression $\boldsymbol{\beta}$ sont estimés en résolvant l'équation suivante:

$$\boldsymbol{S}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^T v_i^{-1} \{Y_i - \mu_i(\boldsymbol{\beta})\} = \mathbf{0}. \tag{1.1.2}$$

Notons que $\boldsymbol{S}(\boldsymbol{\beta})$ est la dérivée du logarithme de la fonction de vraisemblance d'un membre quelconque de la famille des distributions exponentielles (équation 1.1.1). La solution $\hat{\boldsymbol{\beta}}$, qui maximise la vraisemblance, ne peut pas être obtenue directement lorsque les variables ne sont pas de distribution gaussienne. L'algorithme de Newton-Raphson est alors utilisé pour résoudre numériquement l´équation score.

Finalement, pour des échantillons de grande taille, $\hat{\boldsymbol{\beta}}$ obéit à une distribution gaussienne de moyenne $\boldsymbol{\beta}$ et de variance

$$\mathbf{V} = \left(\sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^T v_i^{-1} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^{-1}. \tag{1.1.3}$$

Cette variance peut être estimée en remplaçant $\boldsymbol{\beta}$ par $\hat{\boldsymbol{\beta}}$ dans l'expression 1.1.3.

En conclusion, le MLG est un modèle pour lequel chacune des variables expliquées est un membre de la famille des distributions exponentielles (composante aléatoire). Une relation (lien) est établie entre une fonction de la moyenne et les variables explicatives (composante systématique) pour une équation de prédiction linéaire.

1.1.3. La quasi-vraisemblance

Une caractéristique importante de la famille des MLG est que la fonction score dépend uniquement de la moyenne et de la variance de Y_i. Wedderburn (1974) est le premier dans la littérature à utiliser cette caractéristique pour définir la quasi-vraisemblance. Il souligne que l´équation 1.1.2 peut être utilisée pour estimer les coefficients de régression pour tout lien et fonction de variance peu importe le membre de la famille exponentielle à laquelle la composante aléatoire est associée. Pour définir la vraisemblance, la forme de la distribution des observations doit être précisée. Par contre, la quasi-vraisemblance est déterminée uniquement par la relation entre la moyennne et la variance des observations sans que la distribution exacte des Y_i ne soit spécifiée.

McCullagh (1983) montre que la solution $\hat{\boldsymbol{\beta}}$ de la fonction quasi-score 1.1.2 a une distribution échantillonnale qui, pour des échantillons de tailles élevées, est approximativement gaussienne avec une moyenne $\boldsymbol{\beta}$ et une variance donnée par l'équation 1.1.3. De plus, $\hat{\boldsymbol{\beta}}$ est un estimateur convergent de $\boldsymbol{\beta}$ si $h(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$.

Cette propriété robuste tient parce que l'espérance de $\boldsymbol{S}(\boldsymbol{\beta})$ est égale à zéro tant que $E(Y_i) = \mu_i(\boldsymbol{\beta})$. La matrice de la variance asymptotique de $\hat{\boldsymbol{\beta}}$ est de la forme

$$\mathbf{V}_2 = \mathbf{V}\left(\sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^T v_i^{-1} \mathrm{Var}(Y_i) v_i^{-1} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right) \mathbf{V}. \tag{1.1.4}$$

Notons que $\mathbf{V}_2$ est identique à $\mathbf{V}$ (équation 1.1.3) si et seulement si $\mathrm{Var}(Y_i) = v_i$. Lorsque cette hypothèse est douteuse, l'intervalle de confiance pour $\boldsymbol{\beta}$ peut être basé sur la matrice de la variance estimée

$$\hat{\mathbf{V}}_2 = \hat{\mathbf{V}}\left(\sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^T v_i^{-1} \{Y_i - \mu_i(\boldsymbol{\beta})\}^2 v_i^{-1} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right) \hat{\mathbf{V}},$$

évaluée en $\hat{\boldsymbol{\beta}}$. Soulignons que $\hat{\mathbf{V}}_2$ est un estimateur robuste de la variance puisque $\hat{\mathbf{V}}_2$ est convergent même si $\mathrm{Var}(Y_i)$ n'est pas spécifié correctement.

En pratique, la quasi-vraisemblance peut être un choix judicieux pour estimer les paramètres de régression. En effet, il peut être difficile de décider d'une distribution particulière des observations. Par contre, la relation entre la moyenne et la variance est souvent plus facile à postuler; la quasi-vraisemblance permet alors une analyse satisfaisante des données.

1.2. DESCRIPTION DE L'APPROCHE MARGINALE

Lorsque les variables sont de type longitudinal, le MLG ne tient plus. Une stratégie consiste alors à modéliser l'espérance marginale de la réponse en considérant la corrélation comme un paramètre de nuisance. Nous illustrons le modèle marginal pour lequel Y_{ij} est la variable de réponse aléatoire pour le sujet i $(i = 1, ..., m)$ au temps j $(j = 1, ..., t_i)$, $\boldsymbol{x}_{ij}$ le vecteur de variables explicatives pour le sujet i au temps j et $\boldsymbol{\beta}$ le vecteur des paramètres de régression de dimensions $p \times 1$. Émettons l'hypothèse que les sujets sont indépendants les uns des autres. L'ensemble des réponses répétées pour un sujet i peut être regroupé en un vecteur $\mathbf{Y}_i = (Y_{i1}, ..., Y_{it_i})^T$ de dimensions $t_i \times 1$. Le vecteur $\mathbf{Y}_i$ a une moyenne représentée par $E(\boldsymbol{Y}_i) = \boldsymbol{\mu}_i$ et une matrice de covariance de format $t_i \times t_i$ notée $\text{Var}(\boldsymbol{Y}_i)$. L'élément jk de $\text{Var}(\boldsymbol{Y}_i)$ est $\text{Cov}(Y_{ij}, Y_{ik})$, la covariance entre Y_{ij} et Y_{ik}. Pour le modèle marginal, les paramètres de régression et la corrélation intra-sujet sont modélisés séparément. Nous supposons que:

i) l'espérance marginale de la réponse, $E(Y_{ij}) = \mu_{ij}$, dépend des variables explicatives $\boldsymbol{x}_{ij}$ par $h(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$, où h est la fonction de lien du modèle et $\boldsymbol{x}_{ij}^T$ le vecteur transposé du vecteur $\boldsymbol{x}_{ij}$,

ii) la variance marginale dépend de la moyenne marginale selon l'équation $\text{Var}(Y_{ij}) = v(\mu_{ij})\phi$, où v est une fonction de variance connue et ϕ un paramètre d'échelle qu'il faut estimer dans certains cas,

iii) la corrélation entre Y_{ij} et Y_{ik} est une fonction de la moyenne marginale et peut-être d'un paramètre additionnel $\boldsymbol{\alpha}$, i.e. $\text{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \boldsymbol{\alpha})$, où $\rho(.)$ est une fonction connue.

Notons que l'espérance et la variance marginales correspondent aux résultats obtenus à la première section. En fait, le modèle marginal est différent en raison de la corrélation entre les réponses observées (item iii).

Remarquons que le modèle marginal est exprimé en terme d'espérance sur tous les membres de la population qui ont les mêmes valeurs de variables explicatives; pour cette raison, il est parfois appelé modèle de moyenne de population. Le modèle marginal peut ajuster des données provenant d'études transversales puisqu'il n'implique pas le temps comme caractéristique intrinsèque. Par conséquent, comme dans le cas des études transversales, le vecteur $\boldsymbol{\beta}$ caractérise la manière dont les variables explicatives dépendent de la distribution.

Dans le cas spécifique de la régression logistique pour des variables dichotomiques répétées, la fonction de lien est la fonction logit telle que

$$logit(\mu_{ij}) = \log \frac{\mu_{ij}}{1 - \mu_{ij}}.$$

De plus, pour les réponses binaires, $v(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$ et $\phi = 1$.

Puisque le modèle n'est pas entièrement spécifié, il n'est pas possible d'utiliser les méthodes du maximum de vraisemblance pour l'analyse sans imposer une structure plus précise. Cette structure peut être ajoutée de plusieurs façons, voir par exemple McCullagh et Nelder (1989, Sections 6.5 et 6.6). Malheureusement, toutes ces méthodes nécessitent des calculs fastidieux pour les modèles avec plus de 2 mesures répétées, pour plus de détails voir Liang, Zeger et Qaqish (1992, Section 2).

Une alternative est la méthode des moindres carrés généralisés empiriques (MCGE). Pour cette approche, l'estimation et l'inférence sont basées sur les méthodes préconisées par Grizzle, Stamer et Koch (1969) ainsi que Koch *et al.* (1977) et discutées plus récemment par Agresti (1989). La méthode MCGE n'est pas discutée ici car elle comporte deux inconvénients majeurs. Premièrement, elle nécessite des échantillons de taille plus élevée que l'approche GEE. Stokes, Davis et Koch (1995) suggèrent un échantillon efficace de 25-30 personnes pour chacune

des fonctions de la réponse. Deuxièmement, la méthode des MCGE ne peut pas être utilisée lorsque des variables explicatives sont continues.

Il n'est pas nécessaire d'utiliser la méthode du maximum de vraisemblance usuelle pour obtenir les estimateurs des paramètres du modèle marginal. Pour le reste de cette section, nous décrivons une approche alternative qui requiert seulement les spécifications du modèle marginal (2 premiers moments) et qui se nomme GEE (équation d'estimation généralisée); voir les détails dans Liang et Zeger (1986), Zeger et Liang (1986) et Prentice (1988). Cette approche semi-paramétrique, introduite par Liang et Zeger, est utilisée pour l'analyse de données longitudinales par plusieurs auteurs dans la littérature. L'intérêt majeur de l'approche GEE est qu'elle peut être utilisée pour des réponses répétées catégoriques.

Comme il n'y a pas de forme simple de la fonction de vraisemblance, $\boldsymbol{\beta}$ est estimé en résolvant un analogue multivarié de la fonction score provenant de la fonction de quasi-vraisemblance. Il faut que la formulation de la quasi-vraisemblance (Wedderburn, 1974) s'applique comme c'est le cas par exemple pour les variables de distribution normale, Poisson, binomiale et gamma. L'équation score pour $\boldsymbol{\beta}$ devient :

$$\boldsymbol{S}_{\beta}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T Var(\boldsymbol{Y}_i)^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \tag{1.2.1}$$

Cette fonction est plus compliquée que celle de Wedderburn (équation 1.1.2) car elle ne dépend pas seulement de $\boldsymbol{\beta}$, mais aussi d'un paramètre additionnel $\boldsymbol{\alpha}$ puisque $\text{Var}(\mathbf{Y}_i) = \text{Var}(\mathbf{Y}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})$. Plusieurs auteurs parlent de $\text{Var}(\mathbf{Y}_i)$ comme d'une matrice de covariances "de travail" ("working covariance matrix"). Ce terme est utilisé premièrement parce que l'estimateur de la matrice de covariances dépend du paramètre $\boldsymbol{\alpha}$. Deuxièmement, nous savons que les paramètres de régression et leurs variances convergent, et ce même lorsque la structure de la

matrice de covariance est mal spécifiée. La perte d'efficacité lorsque le choix de $\text{Var}(\mathbf{Y}_i)$ est incorrect n'a donc pas de conséquence lorsque le nombre de sujets est grand. Enfin, la structure de covariance dans le temps est traitée comme un paramètre de nuisance.

La dépendance de $\boldsymbol{\alpha}$ peut être résolue dans l'équation 1.2.1 par un estimateur $m^{1/2}$-convergent $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}})$. Liang et Zeger (1986) démontrent que la solution de l'équation résultante est asymptotiquement aussi efficace que si $\boldsymbol{\alpha}$ était connu. Pour des réponses binaires, le paramètre $\boldsymbol{\alpha}$ peut être estimé à l'aide de différentes méthodes. Une approche simple consiste à paramétriser $\text{Var}(\mathbf{Y}_i)$ en terme de corrélation et à utiliser la méthode des moments pour estimer les paramètres inconnus. Il s'agit en fait d'estimer les paramètres de corrélations en utilisant les résidus de Pearson. Cette approche est la première proposée (Liang et Zeger, 1986), et elle est utilisée dans la procédure GENMOD du progiciel SAS pour les estimateurs GEE. Dans un autre ordre d'idées, Fitzmaurice, Laird et Rotnitsky (1993) paramétrisent $\boldsymbol{\alpha}$ à l'aide des rapports de cotes conditionnelles. De plus, Lipsitz, Laird et Harrington (1991), Liang, Zeger et Qaqish (1992) et Carey, Zeger et Diggle (1993) formulent $\text{Var}(\mathbf{Y}_i)$ en fonction du rapport de cotes marginales.

Une méthode importante pour estimer les paramètres de second moment est proposée par Prentice (1988) et utilisée par Diggle, Liang et Zeger (1994). Prentice estime les paramètres associés au modèle en ajoutant un second groupe d'équations d'estimation, et il résout simultanément les équations pour $\hat{\boldsymbol{\beta}}$ et $\hat{\boldsymbol{\alpha}}$. L'équation score pour $\boldsymbol{\alpha}$ lorsque les réponses sont dichotomiques est de la forme :

$$\boldsymbol{S}_{\alpha}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}} \right)^T \mathbf{H}_i^{-1} (\boldsymbol{W}_i - \boldsymbol{\eta}_i) = \mathbf{0}, \tag{1.2.2}$$

où

$$\begin{cases} \boldsymbol{W}_i = (R_{i1}R_{i2}, R_{i1}R_{i3}, ..., R_{it_i-1}R_{it_i}), \\ \mathbf{H}_i = \text{diag}\{\text{Var}(R_{i1}R_{i2}), \text{Var}(R_{i1}R_{i3}), ..., \text{Var}(R_{it_i-1}R_{it_i})\}, \\ R_{ij} = \{Y_{ij} - \mu_{ij}\}/\{\mu_{ij}(1-\mu_{ij})\}^{1/2}, \\ \boldsymbol{\eta}_i = \text{E}(\boldsymbol{W}_i). \end{cases}$$

Pour des réponses binaires, $\boldsymbol{\eta}_i$ et H_i sont complètement déterminés par les modèles de la moyenne et de la corrélation sans hypothèse additionnelle sur les moments d'ordre supérieur. La solution $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ des équations 1.2.1 et 1.2.2 est asymptotiquement gaussienne et convergente (Liang et Zeger, 1986). La variance correspondante est estimée par :

$$\left(\sum_{i=1}^{m} \mathbf{C}_i^T \mathbf{B}_i^{-1} \mathbf{D}_i \right)^{-1} \left(\sum_{i=1}^{m} \mathbf{C}_i^T \mathbf{B}_i^{-1} \mathbf{V}_{0i} \mathbf{B}_i^{-1} \mathbf{C}_i \right) \left(\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{B}_i^{-1} \mathbf{C}_i \right)^{-1} \quad (1.2.3)$$

évaluée en $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$, où

$$\mathbf{C}_i = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}} \end{pmatrix}, \mathbf{B}_i = \begin{pmatrix} \text{Var}(\boldsymbol{Y}_i) & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_i \end{pmatrix}, \mathbf{D}_i = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\alpha}} \\ \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}} \end{pmatrix}$$

et

$$\mathbf{V}_{0i} = \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{w}_i - \boldsymbol{\eta}_i \end{pmatrix} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{w}_i - \boldsymbol{\eta}_i \end{pmatrix}^T.$$

L'équation 1.2.3 permet d'estimer de manière convergente la variance des estimateurs de la régression.

La méthode GEE possède deux avantages majeurs. Premièrement, dans plusieurs situations pratiques $\hat{\boldsymbol{\beta}}$ est presqu'aussi efficace que les estimateurs du maximum de vraisemblance de $\boldsymbol{\beta}$ lorsque $\text{Var}(\mathbf{Y}_i)$ est une approximation raisonnable (Liang et Zeger, 1986). En fait, l'équation 1.2.1 correspond à l'équation score du maximum de vraisemblance lorsque les données sont gaussiennes multivariées.

De plus, Zhao et Prentice (1990) ont identifié une classe de modèles quadratiques exponentiels pour laquelle une forme particulière du GEE est le maximum de vraisemblance et donc asymptotiquement efficace. Deuxièmement, $\hat{\boldsymbol{\beta}}$ est convergent lorsque $n \to \infty$ même si la structure de la matrice de covariances des $\mathbf{Y}_i$ n'est pas spécifiée correctement.

Sharples et Breslow (1992) présentent les résultats de simulations pour étudier les propriétés du GEE lorsque les échantillons sont de petite taille. Ils considèrent le cas pour lequel il y a peu d'observations par sujet et où les variables explicatives sont dichotomiques. Selon leurs résultats, la méthode GEE dans ce cas particulier donne de bons estimateurs des paramètres de régression; cependant l'estimation des paramètres de corrélation pose quelques problèmes. Pour les petits échantillons (≈ 100), une mauvaise structure de corrélation affecte le biais, ce qui n'est pas le cas pour des échantillons plus larges. Lipsitz *et al.* (1994) discutent des estimateurs obtenus en une étape, i.e. provenant d'une seule itération de l'approche GEE. Ils comparent leur performance pour des petits échantillons à celle des estimateurs obtenus par l'algorithme complet. Ils concluent que l'estimation en une étape est à peu près aussi performante que l'itération complète.

Pour terminer cette section, nous faisons un survol des articles qui proposent des extensions de la méthode GEE pour modéliser des réponses de différents types. Thall et Vail (1990) présentent une famille de modèles pour des données longitudinales de comptage en adoptant une approche similaire à celle de Liang et Zeger (1986). Liang, Zeger et Qaqish (1992) utilisent une extension multivariée du GEE afin d'estimer les coefficients de régression pour des réponses multivariées. Lipsitz, Kim et Zhao (1994) discutent d'une extension de la méthode de Liang et Zeger pour modéliser la corrélation entre des réponses catégoriques nominales ou ordinales. Leur méthode se réduit à celle de Liang et Zeger lorsque les réponses

sont binaires. Finalement, Fitzmaurice, Heath et Clifford (1996) discutent des manières d'analyser des données longitudinales lorsque les réponses sont binaires et que des sujets quittent l'étude en cours de route ("drop out").

1.3. DESCRIPTION DE L'APPROCHE SPÉCIFIQUE AUX SUJETS

Historiquement, le modèle avec des effets aléatoires a été développé pour tenir compte de la variabilité supplémentaire des réponses binaires en blocs. En effet, puisque la variabilité intra-blocs ne peut être expliquée seulement par une distribution binomiale, les modèles avec des effets aléatoires sont introduits pour tenir compte de cette variation "extra-binomiale". La distribution beta-binomiale (Skellam, 1948) est une des premières à être présentée. Elle est utilisée ensuite par plusieurs auteurs dans les années 70-80. Originellement, la structure beta-binomiale requiert que chaque réponse d'un même bloc ou d'un même sujet ait la même probabilité. Par contre, Rosner (1984) propose une extension de la distribution beta-binomiale afin de permettre que les variables explicatives diffèrent dans un même bloc. Malgré cette amélioration, le modèle a plusieurs limites; voir Diggle, Liang et Zeger (1994, section 9.3.2). De plus, Crouch et Spiegelman (1990) expliquent que le modèle logistique-normal est préférable au modèle beta-binomial, entre autre parce que les effets fixes et aléatoires sont ajoutés sur la même échelle, et parce que plusieurs niveaux sont facilement incorporés au modèle. Dans cette section, nous discutons des principales propriétés du modèle logistique-normal en mettant l'emphase sur les différentes propositions pour résoudre la fonction de vraisemblance.

Un modèle linéaire généralisé à effets mixtes (MLGM) est une extension du MLG qui permet d'analyser entre autre des données longitudinales en ajoutant des effets aléatoires à la fonction de prédiction linéaire. Dans le cas des études

longitudinales, le MLGM est raisonnable lorsqu'on suppose que les coefficients de la régression peuvent varier d'une personne à une autre suivant une loi F. Ce type de modèle est appelé modèle spécifique aux sujets car la corrélation des observations pour un même sujet vient de leur variable aléatoire non observable commune.

Dans le modèle spécifique aux sujets, nous supposons que, sachant $\mathbf{U}_i$, les réponses $Y_{i1}, ..., Y_{it_i}$ sont mutuellement indépendantes et suivent un modèle linéaire généralisé avec la densité de la famille exponentielle

$$g(y_{ij}|\boldsymbol{U}_i) = exp[\{y_{ij}\gamma_i - \psi(\gamma_i)\}/\phi + c(y_{ij}, \phi)].$$

Les moments conditionnels (voir l'équation 1.1.2),

$$\mu_{ij} = E(Y_{ij}|\boldsymbol{U}_i) = \frac{\partial\psi(\gamma_i)}{\partial\gamma_i} \qquad \text{et} \qquad v_{ij} = Var(Y_{ij}|\boldsymbol{U}_i) = \frac{\partial^2\psi(\gamma_i)}{\partial\gamma_i^2}\phi,$$

satisfont $h(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \mathbf{d}_{ij}^T\mathbf{U}_i$ et $v_{ij} = v(\mu_{ij})\phi$, où h et v sont respectivement les fonctions de lien et de variance. De plus, $\mathbf{d}_{ij}$ (de dimensions $c \times 1$) est un sous-ensemble de $\boldsymbol{x}_{ij}$ (de dimensions $p \times 1$) pour le sujet i au temps j. Le vecteur $\mathbf{U}_i = [\mathrm{U}_{i1}, ..., \mathrm{U}_{ic}]^T$, $i = 1, ..., m$, est de format $c \times 1$ où c est le nombre d'effets aléatoires présents dans le modèle. Les effets aléatoires $\mathbf{U}_i$, sont mutuellement indépendants et obéissent à une distribution multivariée commune F. Bien que plusieurs autres distributions des effets aléatoires puissent être utilisées, nous nous limitons à la loi gaussienne car elle est la plus communément utilisée et elle donne habituellement une approximation raisonnable. Nous supposons donc que les $\mathbf{U}_i$ sont indépendants identiquement distribués selon une loi gaussienne de moyenne $\mathbf{0}$ et de matrice de covariance $\mathbf{G}$ de dimensions $c \times c$. La matrice $\mathbf{G}$ dépend d'un paramètre inconnu $\boldsymbol{\theta} = (\theta_1, ..., \theta_c)^T$. La matrice de covariance des $\mathbf{U}_i$ est donc de

la forme

$$\text{Cov}(\boldsymbol{U}_i) = \mathbf{G}(\boldsymbol{\theta}) = \text{diag}(\theta_1, ..., \theta_c).$$

Le modèle logistique-normal est un cas particulier du modèle spécifique aux sujets. Il est utilisé pour la modélisation des réponses corrélées de type dichotomique. La composante aléatoire associée au modèle est la distribution binomiale. Ainsi, le deuxième moment conditionnel est

$$v_{ij} = Var(Y_{ij}|\boldsymbol{U}_i) = \mu_{ij}(1 - \mu_{ij}),$$

car $\phi = 1$ dans le cas du modèle logistique.

En général, pour un modèle spécifique aux sujets, la vraisemblance exprimée en fonction des paramètres inconnus $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$ est donnée par la fonction suivante :

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{m} \int \prod_{j=1}^{t_i} g(y_{ij}|\boldsymbol{U}_i) f(\boldsymbol{U}_i; \boldsymbol{\theta}) d\boldsymbol{U}_i. \tag{1.3.1}$$

Cette fonction de vraisemblance a la forme d'une intégrale de la distribution conjointe des données et des effets aléatoires, sur les effets aléatoires non observés. Dans le cas du modèle de régression ordinaire avec des effets aléatoires, l'intégrale a une expression connue et des méthodes simples existent pour maximiser la vraisemblance (par exemple, l'algorithme de Newton-Raphson). Par contre, lorsque le modèle est non linéaire, il faut souvent utiliser des techniques d'intégration numériques complexes pour évaluer le maximum de la vraisemblance.

Dans le cas du modèle logistique avec des effets aléatoires gaussiens, les difficultés de calcul sont toujours présentes. La fonction de vraisemblance pour $\boldsymbol{\beta}$ et $\mathbf{G}$ est

$$L(\boldsymbol{\beta}, \boldsymbol{G}; \boldsymbol{y}) = \prod_{i=1}^{m} \int \prod_{j=1}^{t_i} \{\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{U}_i)\}^{y_{ij}} \{1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{U}_i)\}^{1-y_{ij}} f(\boldsymbol{U}_i; \boldsymbol{\theta}) d\boldsymbol{U}_i, \tag{1.3.2}$$

où $\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{U}_i) = E(y_{ij}|\boldsymbol{U}_i;\boldsymbol{\beta}) = \exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \mathbf{d}_{ij}^T\mathbf{U}_i)/\{1+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \mathbf{d}_{ij}^T\mathbf{U}_i)\}$. Avec le lien logit et une hypothèse gaussienne sur les $\mathbf{U}_i$, l'équation 1.3.2 se réduit à

$$\prod_{i=1}^{m} \int \exp\left[\boldsymbol{\beta}^T \sum_{j=1}^{t_i} \boldsymbol{x}_{ij} y_{ij} + \boldsymbol{U}_i^T \sum_{j=1}^{t_i} \boldsymbol{d}_{ij} y_{ij} - \sum_{j=1}^{t_i} \log\{1 + \exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{d}_{ij}^T\boldsymbol{U}_i)\} \right]$$
$$\times (2\pi)^{-c/2}|\boldsymbol{G}|^{-1/2}\exp(-\mathbf{U}_i^T\mathbf{G}^{-1}\mathbf{U}_i/2)d\mathbf{U}_i. \qquad (1.3.3)$$

Le but est de trouver $\boldsymbol{\beta}$ et $\mathbf{G}$ qui maximisent cette expression, le problème principal étant l'évaluation de l'intégrale multiple. Plusieurs méthodes ont été proposées dans la littérature et un survol de ces méthodes est présenté dans cette section. Au chapitre 2, deux approches spécifiques seront décrites en détails.

L'intégrale 1.3.3 peut être calculée en utilisant la quadrature gaussienne lorsque $c \leq 2$. Diggle, Liang et Zeger (1994) conseillent d'ailleurs des techniques numériques pour estimer les coefficients du modèle logistique-normal. Crouch et Spiegelman (1990) présentent une nouvelle méthode qui est une extension de la méthode publiée par Goodwin (1949) pour résoudre l'intégrale. Leurs études numériques montrent la quadrature gaussienne (avec 20 points de quadrature) est préférable pour des petites valeurs de la réponse. Par contre, lorsque la quadrature gaussienne ne converge pas, la nouvelle méthode est une alternative intéressante. Pinheiro et Bates (1995) décrivent et comparent plusieurs approximations de l'intégrale de la vraisemblance. Ils concluent que la quadrature gaussienne adaptée (QGA) est une des meilleurs méthodes lorsque l'effet aléatoire est de petite dimension.

Pour des problèmes de dimension élevée ($c > 2$), la méthodologie de Zeger et Karim (1991) est conseillée. Il s'agit d'utiliser l'algorithme d'échantillonnage de Gibbs pour faire une simulation à partir d'une distribution a posteriori similaire à la fonction de vraisemblance. L'algorithme de Zeger et Karim (1991) est

flexible car il permet de changer facilement le nombre d'effets aléatoires et leur distribution hypothétique. Par contre, il nécessite des calculs fastidieux.

Une stratégie alternative permet d´éviter les intégrations en faisant une approximation des équations score. Plusieurs stratégies d'approximation de la procédure d'inférence sont proposées. Celles-ci incluent les approximations de Laplace de l'intégrale de la vraisemblance (Liu et Pierce 1993; Salomon et Cox 1992) et l'approche de la quasi-vraisemblance pénalisée (PQL) également basée sur les approximations de Laplace (Breslow et Clayton 1993; Schall 1991). Les approximations de Laplace sont assez complexes et d'après Breslow et Lin (1995), elles sont instables numériquement dans certaines circonstances. Par exemple, elles peuvent occasionnellement ne pas avoir de maximum local dans le voisinage de la vraie valeur du paramètre. Nous étudions donc la correction du biais des estimateurs PQL dans ce texte.

L'approche PQL est présentée par Stiralli, Laird et Ware (1984) pour l'analyse de réponses dichotomiques corrélées. L'inférence est basée sur l'estimation du maximum de vraisemblance des effets fixes et des composantes de la variance. De plus, l'estimation des effets aléatoires est effectuée par la méthode empirique bayésienne. Comme les solutions exactes n'existent pas, Stiralli, Laird et Ware proposent une approximation utilisant le mode de la distribution a posteriori des effets aléatoires. Cette approximation est effectuée au moyen de l'algorithme EM décrit par Dempster, Laird et Rubin (1977). Notons que cet algorithme a d'abord été mis au point pour l'analyse de jeux de données incomplets. Chacune des itérations comprend une étape d'évaluation de l'espérance (étape E) suivie d'une étape de maximisation (étape M). Lindstrom et Bates (1990) appliquent l'approximation de Stiralli, Laird et Ware (1984) aux modèles mixtes non linéaires.

Schall (1991) utilise la même approximation avec la modification de Harville de la méthode d'Henderson des approximations successives.

Breslow et Clayton (1993) présentent une approche unifiée en reliant les principes des modèles linéaires mixtes généralisés. Ils proposent quelques modifications à une expansion de Laplace afin de justifier des équations estimées (GEE) pouvant être résolues itérativement. Wolfinger et O'Connell (1993) raffinent la procédure de Breslow et Clayton pour l'estimation des effets. Leur méthode tient compte explicitement du paramètre de d'échelle ϕ dans l'estimation des modèles linéaires généralisés.

En fait, la procédure de Wolfinger et O'Connell est semblable à celle de Breslow et Clayton, car dans les deux cas, les équations généralisées du modèle mixte sont utilisées. La méthode de Breslow et Clayton, qu'ils appellent PQL, suppose que le paramètre d'échelle $\phi = 1$. Par contre, la procédure de Wolfinger et O'Connell, qu'ils appellent pseudo-vraisemblance (PL) ou pseudo-vraisemblance restreinte (REPL), suppose que le paramètre d´échelle est inconnu. L'estimateur de ϕ avec PL peut être vu comme un estimateur du maximum de vraisemblance alors qu'avec REPL il s'apparente plutôt à un estimateur du maximum de vraisemblance restreint. En résumé, la méthode PQL est un cas spécial de la méthode PL lorsque $\phi = 1$.

La méthode PQL est décrite en détail au chapitre suivant. Cette méthode mène à des résultats biaisés pour des réponses dichotomiques avec peu de répétitions dans le temps (Breslow et Clayton, 1993). Les expressions asymptotiques obtenues par Solomon et Cox (1992) sont donc utilisées par Breslow et Lin (1995) et Lin et Breslow (1996a) pour réduire le biais des estimateurs PQL. Nous étudions finalement l'approximation numérique qui consiste à évaluer l'intégrale par la QGA.

Pour terminer cette section, nous mentionnons deux articles intéressants pour l'analyse de données dichotomiques corrélées avec des méthodes autres que la régression logistique. Orme et Fry (1995) parlent de l'ajout d'effets aléatoires dans les modèles avec un lien probit. Dans ce cas, la vraisemblance ne requiert pas d'intégration numérique et les expressions pour l'équation score et la matrice Hessienne sont données. Conaway (1990) discute également de l'analyse des données binaires avec des effets aléatoires. Par contre, le lien choisi est le log-log et la distribution mixte log-gamma est utilisée. Dans cet article, la vraisemblance est calculée numériquement.

1.4. Comparaisons des deux approches

Suivant Diggle, Liang et Zeger (1994), Zeger et Liang (1992) et Kenward et Jones (1992), nous comparons deux approches pour l'analyse de données binaires longitudinales: le modèle marginal et le modèle spécifique aux sujets. La principale distinction entre ces deux modèles est l'interprétation des coefficients de régression. En régression linéaire, les coefficients peuvent avoir une interprétation marginale pour chacune des deux approches. Par contre, ce n'est plus cas pour les modèles non linéaires. Les coefficients du modèle marginal décrivent alors l'effet des changements des variables explicatives sur la réponse moyenne de la population alors que les coefficients du modèle spécifique aux sujets décrivent l'effet sur la réponse d'un individu. Le modèle spécifique aux sujets est donc adéquat lorsque le but est de modéliser la réponse pour un individu. Par contre, le modèle marginal est utilisé efficacement dans les études de population, par exemple en épidémiologie. Le but est alors de faire de l'inférence sur la réponse moyenne de la population entre deux groupes avec des facteurs de risque différents.

Une deuxième distinction entre les deux approches concerne les hypothèses de dépendance dans le temps. En effet, le modèle marginal décrit la covariance des réponses observées pour un même sujet tandis que le modèle spécifique aux sujets explique la source de cette covariance. Dans le modèle marginal, la seule restriction sur la matrice de covariance est qu'elle doit être définie positive. Par contre, dans le modèle spécifique aux sujets, la dépendance dans le temps vient uniquement des effets communs aux sujets ($\mathbf{U}_i$) dans l'espérance conditionnelle. La matrice de covariance est donc totalement déterminée par le choix de $h(\mu_{ij})$ et de F (la loi des effets aléatoires).

Exemple 1.4.1. *Considérons une étude clinique pour laquelle il faut tester le rôle protecteur du médicament A contre la maladie M. Un échantillon est sélectionné pour former aléatoirement un groupe d'individus qui reçoit le médicament A et un autre groupe qui reçoit un placebo. L'expérimentateur doit vérifier, à chacun des temps j ($j = 1, ..., t_i$), si l'individu i ($i = 1, ..., m$) est atteint ou non de la maladie M. Les variables de l´étude sont donc les suivantes:*

$$y_{ij} = \begin{cases} 1 & \textit{si l'individu i est atteint de la maladie M au temps j,} \\ 0 & \textit{sinon.} \end{cases}$$

$$x_{ij} = \begin{cases} 1 & \textit{si l'individu i est traité avec le médicament A,} \\ 0 & \textit{si l'individu i reçoit le placebo.} \end{cases}$$

Le groupe de référence est formé des individus ayant pris le placebo puisque $x_{ij} = 0$ *pour ce groupe.*

L'approche marginale du modèle de régression logistique (lien *logit*) pour l'exemple 1.4.1 consiste à poser

i) $logit(\mu_{ij}) = \log \dfrac{P(Y_{ij}=1)}{P(Y_{ij}=0)} = \log \dfrac{\mu_{ij}}{1-\mu_{ij}} = \beta_0 + \beta_1 x_{ij}$, où $E(Y_{ij}) = \mu_{ij}$,

ii) $\mathrm{Var}(Y_{ij}) = \mu_{ij}(1-\mu_{ij})$,

iii) $\mathrm{Corr}(Y_{ij}, Y_{ik}) = \alpha$.

Ce modèle ignore la différence entre les individus en ce qui a trait à leur prédisposition naturelle à contracter la maladie M. Dans ce contexte, $\exp(\beta_0)$ est le ratio de la fréquence des malades versus les non-malades parmi la sous-population du groupe de référence. Le terme $\exp(\beta_0)$ représente donc la cote des malades parmi le groupe de référence. Par conséquent, le taux de maladie M parmi les individus du groupe de référence est $\exp\{\beta_0\}/\exp\{1+\exp(\beta_0)\}$. De plus, $\exp(\beta_1)$ représente le risque approximatif de contracter la maladie M pour les individus ayant reçu le médicament A divisé par le même risque pour les sujets du groupe placebo. En d'autres termes, $\exp(\beta_1)$ est le ratio approximatif de la prévalence de la maladie M des deux sous-populations d'individus lorsque la prévalence est petite. En effet, nous savons que le rapport de cotes est approximativement égal au risque relatif lorsque la proportion de succès est proche de zéro pour les deux groupes. En résumé, les coefficients de régression du modèle marginal sont interprétés en terme de moyenne de population parce qu'ils comparent les cotes des malades de la population avec et sans le facteur d'exposition (médicament A).

En ce qui concerne l'analyse par l'approche spécifique aux sujets pour l'exemple 1.4.1, le modèle logistique-normal peut être utilisé comme suit:

i) $logit(E(Y_{ij}|U_i)) = \log \dfrac{P(Y_{ij}=1|U_i)}{P(Y_{ij}=0|U_i)} = \beta_0^* + d_{ij}U_i + \beta_1^* x_{ij} = \beta_0^* + U_i + \beta_1^* x_{ij},$

ii) $\mathrm{Var}(Y_{ij}|U_i) = E(Y_{ij}|U_i)\{1 - E(Y_{ij}|U_i)\},$

iii) $U_i \sim \mathrm{N}(0, \theta).$

Notons que dans ce cas, U_i et θ ne sont pas des vecteurs puisque le modèle n'a qu'un seul effet aléatoire ($c = 1$). Cet effet aléatoire a la forme d'une ordonnée à l'origine puisque $d_{ij} = 1 \ \forall \ i, j$. Nous supposons donc que la probabilité de contracter la maladie M varie parmi les sujets, reflétant ainsi leur prédisposition naturelle à avoir cette maladie et l'influence non mesurable des facteurs environnementaux. L'effet du médicament A sur la probabilité d'infection est la même pour chacun des sujets. La variance θ de la distribution gaussienne représente donc le degré d'hétérogénéité parmi ces mêmes sujets.

Le rapport des cotes de la maladie M pour un sujet qui aurait reçu le médicament A versus le même sujet qui aurait reçu le placebo est $\exp(\beta_1^*)$. Nous constatons que ce rapport de cotes est le même pour tous les individus appartenant au groupe d'étude. Par contre, chacun de ces indivividus a son propre risque de référence, soit $\exp(\beta_0^* + U_i)/\{1 + \exp(\beta_0^* + U_i)\}$. Notons que le paramètre $\exp(\beta_0^*)$ représente la cote d'infection pour une personne typique lorsque l'effet aléatoire $U_i = 0$.

En somme, les paramètres des modèles marginal et spécifique aux sujets diffèrent pour le cas logistique. Le premier décrit le ratio des prévalences de la

population, tandis que le second décrit le ratio des cotes individuelles. Comme l'interprétation des coefficients est différente, il en est de même pour leur magnitude.

Posons un modèle spécifique aux sujets avec un seul effet aléatoire U_i avec $d_{ij} = 1$. Neuhaus, Kalbfleisch et Hauck (1991) montrent que pour toute distribution F avec $\text{Var}(U_i) > 0$, les éléments des vecteurs de régression marginale ($\boldsymbol{\beta}$) et spécifique aux sujets ($\boldsymbol{\beta}^*$) satisfont les énoncés suivants :

(1) $|\beta_k| \leq |\beta_k^*|,\ k = 1, ..., p,$

(2) l'égalité tient si et seulement si $\beta_k^* = 0,$

(3) la différence entre β_k et β_k^* augmente avec $\text{Var}(U_i)$.

En particulier, Zeger, Liang et Albert (1988) montrent que si F est une distribution gaussienne de moyenne 0 et de variance θ, alors

$$logitE(Y_{ij}) \approx (g^2\theta^2 + 1)^{-1/2}\boldsymbol{x}_{ij}^T\boldsymbol{\beta}^*,$$

où $g = 16\sqrt{3}/(15\pi)$. Il suit que

$$\boldsymbol{\beta} \approx (g^2\theta^2 + 1)^{-1/2}\boldsymbol{\beta}^*, \tag{1.4.1}$$

avec $g^2 \approx 0,346$. Notons que l´équation 1.4.1 vérifie les trois propriétés démontrées par Neuhaus, Kalbfleisch et Hauck (1991).

Chapitre 2

MÉTHODE PQL ET QUADRATURE GAUSSIENNE

Ce chapitre est consacré entièrement à l'approche spécifique aux sujets. Nous discutons de deux méthodes pour résoudre le maximum de la vraisemblance du modèle logistique-normal. Il s'agit premièrement de décrire l'approche PQL. Pour ce faire, une notation matricielle correspondant à la notation de la section 1.3 est introduite. Deuxièmement, puisque la méthode PQL est biaisée pour les réponses binaires, nous examinons la correction du biais des estimateurs proposée par Lin et Breslow (1996a). Nous terminons le chapitre en étudiant une méthode sans biais basée sur la quadrature gaussienne.

2.1. NOTATION

Le modèle spécifique aux sujets qui a comme cas spécial le modèle logistique-normal est décrit à la section 1.3. Nous avons postulé le modèle suivant pour l'individu $i = 1, ..., m$ au temps $j = 1, ..., t_i$:

$$h(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \mathbf{d}_{ij}^T\mathbf{U}_i \qquad \text{où} \qquad \boldsymbol{U}_i = [U_{i1}, ..., U_{ic}]^T. \tag{2.1.1}$$

Dans le cas du modèle logistique-normal, les moments conditionnels sont

$$\mu_{ij} = E(Y_{ij}|\boldsymbol{U}_i) \qquad \text{et} \qquad v_{ij} = Var(Y_{ij}|\boldsymbol{U}_i) = v(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij}).$$

Nous supposons également que les $\mathbf{U}_i$ sont indépendants identiquement distribués selon une loi gaussienne de moyenne 0 et de matrice de covariance $\mathbf{G}$ de dimensions $c \times c$. La matrice de covariance des $\mathbf{U}_i$ est donc de la forme

$$\mathrm{Cov}(\boldsymbol{U}_i) = \mathbf{G}(\boldsymbol{\theta}) = \mathrm{diag}(\theta_1, ..., \theta_c).$$

Dans le présent chapitre, une notation matricielle est utilisée afin de simplifier les développements théoriques. Définissons donc $\mathbf{b}$ comme le vecteur partitionné de dimensions $cm \times 1$ telle que

$$\mathbf{b} = [\mathbf{b}_1^T, ..., \mathbf{b}_c^T]^T,$$

où

$$\mathbf{b}_k = [U_{1k}, ..., U_{mk}]^T, \quad k = 1, ..., c.$$

Notons que $\mathbf{b}_k$ est un vecteur de m éléments qui représente le $k^{\text{ième}}$ effet aléatoire pour tous les individus de l'échantillon. Posons d_{ijk} la variable explicative associée à l'effet aléatoire k au temps j pour l'individu i, et $n = \sum_{i=1}^{m} t_i$ le nombre total d'observations de l'étude. Ceci permet de définir les variables suivantes :

$$\begin{cases} \mathbf{d}_{i.k} = [d_{i1k}, ..., d_{it_ik}]^T, & \text{de dimensions} \quad t_i \times 1, \\ \mathbf{Z}_k = \mathrm{diag}\{\mathbf{d}_{1i.k}, ..., \mathbf{d}_{m.k}\}^T, & \text{de dimensions} \quad n \times m, \\ \mathbf{Z} = \{\mathbf{Z}_1, ..., \mathbf{Z}_c\}^T, & \text{de dimensions} \quad n \times cm, \\ \mathbf{X} = [\boldsymbol{x}_{11}, ..., \boldsymbol{x}_{1t_1}, \boldsymbol{x}_{21}, ..., \boldsymbol{x}_{2t_2}, ..., \boldsymbol{x}_{m1}, ..., \boldsymbol{x}_{mt_m}]^T, & \text{de dimensions} \quad n \times p, \\ \mathbf{Y} = [Y_{11}, ..., Y_{1t_1}, Y_{21}, ..., Y_{2t_2}, ..., Y_{m1}, ..., Y_{mt_m}]^T, & \text{de dimensions} \quad n \times 1, \\ \boldsymbol{\mu} = [\mu_{11}, ..., \mu_{1t_1}, \mu_{21}, ..., \mu_{2t_2}, ..., \mu_{m1}, ..., \mu_{mt_m}]^T, & \text{de dimensions} \quad n \times 1. \end{cases}$$

Notons que $\mathbf{Y}$ est le vecteur des n observations de Bernoulli indépendantes et $\mathbf{Z}_k$ la matrice des variables explicatives associées au $k^{\text{ième}}$ effet aléatoire.

Exemple 2.1.1. *Considérons l'exemple 1.4.1 dans le cas particulier où* $m = 4$, $t_i = t = 2\ \forall i$, $n = \sum_{i=1}^{m} t_i = m \times 2 = 8$ *et* $c = 1$. *Les vecteurs* $\mathbf{b}$ *et* $\boldsymbol{Z}$ *sont alors définis comme suit :*

$$\mathbf{b} = (\mathbf{b}_1^T)^T = (U_{11}, ..., U_{41})^T$$

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

D'après la nouvelle notation, l'équation 2.1.1 peut s'écrire comme

$$\begin{aligned} h(\boldsymbol{\mu}^{\mathbf{b}}) &= \boldsymbol{X\beta} + \mathbf{Z}_1\mathbf{b}_1 + ... + \mathbf{Z}_c\mathbf{b}_c \\ &= \boldsymbol{X\beta} + \mathbf{Zb}, \end{aligned}$$

avec

$$\boldsymbol{\mu}^{\mathbf{b}} = E(\mathbf{Y}|\mathbf{b}).$$

Dans le cas particulier du modèle logistique-normal,

$$\begin{aligned}
&\mathrm{E}(Y_{ij}|\mathbf{b}) = \mu_{ij}^{\mathbf{b}} = \mathrm{P}(Y_{ij} = 1|\mathbf{b}),\\
&\mathrm{Var}(Y_{ij}|\mathbf{b}) = v_(\mu_{ij}^{\mathbf{b}}) = \mu_{ij}^{\mathbf{b}}(1 - \mu_{ij}^{\mathbf{b}}),\\
&h(\mu_{ij}^{\mathbf{b}}) = logit(\mu_{ij}^{\mathbf{b}}) = \log \frac{\mu_{ij}^{\mathbf{b}}}{1 - \mu_{ij}^{\mathbf{b}}}.
\end{aligned}$$

Ce modèle implique que les effets aléatoires $\mathbf{b}_k$ sont de distribution gaussienne de moyenne zéro et de covariance $\mathrm{cov}(\mathbf{b}_k) = \theta_k \mathbf{I}_m$, où $\mathbf{I}_m$ est la matrice identité d'ordre m. Ainsi, avec $\boldsymbol{\theta} = (\theta_1, ..., \theta_c)^T$, la matrice de covariance de $\mathbf{b}$ est $\mathrm{cov}(\mathbf{b}) = \mathbf{D}(\boldsymbol{\theta}) = \mathrm{diag}(\theta_1 \mathbf{I}_m, ..., \theta_c \mathbf{I}_m)$.

2.2. MÉTHODE PQL

Nous savons que la fonction de vraisemblance pour le modèle logistique-normal ne permet pas d'obtenir des solutions explicites des estimateurs. Tel que mentionné à la section 1.3, une stratégie consiste alors à utiliser la quasi-vraisemblance pénalisée (PQL) afin d´éviter les intégrations.

2.2.1. Motivation

Nous avons déjà défini la fonction de vraisemblance du modèle logistique-normal (équations 1.3.2 et 1.3.3). L'approche PQL repose essentiellement sur le fait qu'il soit possible d´écrire la fonction de vraisemblance de la façon suivante:

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\theta}) &= e^{l(\boldsymbol{\beta}, \boldsymbol{\theta})}\\
&\propto |\mathbf{D}|^{-1/2} \int \exp\left\{ \sum_{i=1}^{m} \sum_{j=1}^{t_i} l_{ij}(\boldsymbol{\beta}; \mathbf{b}) - \frac{1}{2}\mathbf{b}^T \mathbf{D}^{-1} \mathbf{b} \right\} d\mathbf{b}, \quad (2.2.1)
\end{aligned}$$

où,

$$l_{ij}(\boldsymbol{\beta}; \mathbf{b}) = \left[y_{ij} \log \left(\frac{\mu_{ij}^{\mathbf{b}}}{1 - \mu_{ij}^{\mathbf{b}}} \right) + \log(1 - \mu_{ij}^{\mathbf{b}}) \right]. \tag{2.2.2}$$

La procédure d'estimation PQL peut être obtenue à partir de plusieurs arguments. Breslow et Clayton (1993) utilisent une modification de l'aproximation de Laplace. Il s'agit en fait d'écrire l'équation 2.2.1 sous la forme $c|\mathbf{D}|^{-1/2} \int e^{-\kappa(\mathbf{b})} d\mathbf{b}$ et d'appliquer la méthode de Laplace pour l'approximation de l'intégrale (Barndorff-Nielsen et Cox 1989, sec. 3.3).

Supposons que $\boldsymbol{\kappa}'$ et $\boldsymbol{\kappa}''$ représentent respectivement les matrices des dérivées partielles de premier et de second ordre de κ par rapport à $\mathbf{b}$. En ignorant la constante de multiplication c, nous obtenons

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \log |\boldsymbol{\kappa}''(\tilde{\mathbf{b}})| - \kappa(\tilde{\mathbf{b}}), \tag{2.2.3}$$

où $\tilde{\mathbf{b}} = \tilde{\mathbf{b}}(\boldsymbol{\beta}, \boldsymbol{\theta})$ est le vecteur qui minimise $\kappa(\mathbf{b})$. $\tilde{\mathbf{b}}$ est donc la solution du système d´équations

$$\boldsymbol{\kappa}'(\mathbf{b}) = -\sum_{i=1}^{m} \sum_{j=1}^{t_i} (y_{ij} - \mu_{ij}^{\mathbf{b}}) \mathbf{Z}_{ij} + \mathbf{D}^{-1}\mathbf{b} = \mathbf{0},$$

où $\mathbf{Z}_{ij}$ dénote la $(ij)^{\text{e}}$ ligne de $\mathbf{Z}$. En dérivant $\kappa(\mathbf{b})$ une seconde fois par rapport à $\mathbf{b}$, nous calculons

$$\begin{aligned} \boldsymbol{\kappa}''(\mathbf{b}) &= \sum_{i=1}^{m} \sum_{j=1}^{t_i} \mathbf{Z}_{ij} \mathbf{Z}_{ij}^T \mu_{ij}^{\mathbf{b}} (1 - \mu_{ij}^{\mathbf{b}}) + \mathbf{D}^{-1} \\ &= \mathbf{Z}^T \mathbf{W} \mathbf{Z} + \mathbf{D}^{-1}, \end{aligned} \tag{2.2.4}$$

où $\mathbf{W}$ est une matrice diagonale $n \times n$ dont les éléments de la diagonale sont $w_{ij} = v(\mu_{ij}^{\mathbf{b}}) = \mu_{ij}^{\mathbf{b}}(1 - \mu_{ij}^{\mathbf{b}})$ (en énumérant les t_i temps du sujet 1 au sujet m). La

combinaison des équations 2.2.1 à 2.2.4 donne

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2}\log|\mathbf{I} + \mathbf{Z}^T\mathbf{W}\mathbf{Z}\mathbf{D}| + \sum_{i=1}^{m}\sum_{j=1}^{t_i} l_{ij}(\boldsymbol{\beta}; \tilde{\mathbf{b}}) - \frac{1}{2}\tilde{\mathbf{b}}^T\mathbf{D}^{-1}\tilde{\mathbf{b}}. \qquad (2.2.5)$$

Breslow et Clayton (1993) supposent que les éléments de la diagonale de $\mathbf{W}$ varient lentement comme une fonction de la moyenne et ignorent le premier terme de l'expression 2.2.5. Les estimateurs $(\hat{\boldsymbol{\beta}}_P, \hat{\mathbf{b}}) = (\hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta}), \hat{\mathbf{b}}(\boldsymbol{\theta}))$ où $\hat{\mathbf{b}}(\boldsymbol{\theta}) = \tilde{\mathbf{b}}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))$, sont donc obtenus conjointement en maximisant le logarithme de la fonction de quasi-vraisemblance pénalisée (PQL)

$$\sum_{i=1}^{m}\sum_{j=1}^{t_i} l_{ij}(\boldsymbol{\beta}; \mathbf{b}) - \frac{1}{2}\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b},$$

pour $\boldsymbol{\theta}$ fixé. De manière équivalente, l'estimateur PQL $\hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta})$ maximise

$$l_P(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^{m}\sum_{j=1}^{t_i} l_{ij}(\boldsymbol{\beta}; \tilde{\mathbf{b}}) - \frac{1}{2}\tilde{\mathbf{b}}^T\mathbf{D}^{-1}\tilde{\mathbf{b}}, \qquad (2.2.6)$$

pour $\boldsymbol{\theta}$ fixé. De plus, $\tilde{\mathbf{b}}$ résout

$$\mathbf{Z}^T\mathbf{r}_{\mathbf{b}} - \mathbf{D}^{-1}\mathbf{b} = \mathbf{0}$$

et $\mathbf{r}_{\mathbf{b}}$ est un vecteur de résidus $y_{ij} - \mu_{ij}^{\mathbf{b}}$ de dimension $n \times 1$ (en énumérant les t_i temps du sujet 1 au sujet m). Ceci nous amène à écrire les équations scores pour $\boldsymbol{\beta}$ et $\mathbf{b}$:

$$\sum_{i=1}^{m}\sum_{j=1}^{t_i}(Y_{ij} - \mu_{ij}^{\tilde{\mathbf{b}}})\boldsymbol{x}_{ij} = \mathbf{0}$$

et

$$\sum_{i=1}^{m}\sum_{j=1}^{t_i}(Y_{ij} - \mu_{ij}^{\mathbf{b}})\mathbf{Z}_{ij} = \mathbf{D}^{-1}\mathbf{b}.$$

Stiralli, Laird et Ware (1984) et Schall (1991) obtiennent ces mêmes équations à l'aide de la théorie bayésienne.

2.2.2. Solution des estimateurs PQL

2.2.2.1. *Estimation des coefficients de la régression*

Harville (1977) propose un système d'équations permettant d'obtenir le meilleur estimateur sans biais (BLUE) de $\boldsymbol{\beta}$ et $\mathbf{b}$ pour un modèle mixte avec des variables réponses normales. L'équation du modèle de Harville est de la forme $\mathbf{Y} = \boldsymbol{X\beta} + \mathbf{Zb} + \boldsymbol{\epsilon}$, où $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{W}^{-1})$, $\mathbf{b} \sim N(\mathbf{0}, \mathbf{D})$ et $\boldsymbol{\epsilon}$ et $\mathbf{b}$ sont indépendants. Les estimateurs obtenus par la méthode des scores de Fisher peuvent être exprimés comme la solution du système d'équations suivant (Harville 1977, théorème 2):

$$\begin{bmatrix} \mathbf{X}^T\mathbf{WX} & \mathbf{X}^T\mathbf{WZD} \\ \mathbf{Z}^T\mathbf{WX} & \mathbf{I} + \mathbf{Z}^T\mathbf{WZD} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{WY} \\ \mathbf{Z}^T\mathbf{WY} \end{bmatrix}, \tag{2.2.7}$$

où $\mathbf{b} = \boldsymbol{D\nu}$.

Une caractéristique importante des estimateurs PQL est qu'ils peuvent être estimés en ajustant itérativement un modèle linéaire à effets mixtes avec un vecteur de variables réponses "de travail". Il s'agit de définir un vecteur $\mathbf{Y}^*$ de dimensions $n \times 1$ formé des composantes $Y_{ij}^* = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\tilde{\mathbf{b}} + (y_{ij} - \mu_{ij}^{\tilde{\mathbf{b}}})[v(\mu_{ij}^{\tilde{\mathbf{b}}})]^{-1}$ (en énumérant les t_i temps du sujet 1 au sujet m). Il est alors possible d'exploiter les développements théoriques de Harville (1977) puisque $\mathbf{Y}^*$ est normalement distribué avec une moyenne $\boldsymbol{X\beta}$ et une matrice de covariance $\mathbf{V} = \tilde{\mathbf{W}}^{-1} + \mathbf{ZDZ}^T$, où $\tilde{\mathbf{W}}$ est $\mathbf{W}$ évalué en $\tilde{\mathbf{b}}$.

Les estimateurs PQL de $\boldsymbol{\beta}$ et $\mathbf{b}$ sont obtenus en résolvant le système 2.2.7 pour $\mathbf{Y}^*$. De manière équivalente, nous pouvons premièrement résoudre pour $\boldsymbol{\beta}$ par

$$(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}^*, \tag{2.2.8}$$

où $\mathbf{V} = \tilde{\mathbf{W}}^{-1} + \mathbf{Z}\mathbf{D}\mathbf{Z}^T = \tilde{\mathbf{W}}^{-1} + \sum_{k=1}^{c} \theta_k \mathbf{Z}_k \mathbf{Z}_k^T$. Ceci suggère que la covariance approximative de $\hat{\boldsymbol{\beta}}_{\boldsymbol{P}}$ est $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$. Cette dernière valeur est l'inverse de l'information de Fisher pour le modèle linéaire ordinaire lorsque $\boldsymbol{\theta}$ est connu. Deuxièmement, nous posons

$$\hat{\mathbf{b}} = \boldsymbol{D}\hat{\boldsymbol{\nu}} = \mathbf{D}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P). \tag{2.2.9}$$

Les écarts types estimés pour $\hat{\mathbf{b}}$ peuvent être calculés à partir de l'équation 2.2.9. Mentionnons que les équations 2.2.8 et 2.2.9 ne tiennent pas compte de la variabilité additionnelle provenant de l'estimation de $\boldsymbol{\theta}$.

2.2.2.2. *Estimation de la variance*

Les estimateurs PQL des composantes de la variance sont également obtenus à l'aide des relations connues de la théorie de la loi normale. Par contre, la dépendance de $\tilde{\mathbf{W}}$ par rappport à $\boldsymbol{\theta}$ est ignorée lors du calcul de $\partial\mathbf{V}/\partial\theta_k$ $(k = 1, ..., c)$. L'équation du maximum de vraisemblance pour l'estimation des composantes de la variance θ_k, déduite d'après Harville (1977, sec. 4), est approximativement égale à

$$-\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P)^T\mathbf{V}^{-1}(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P),$$

où $\hat{\boldsymbol{\beta}}_P = \hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta})$. Les équations d'estimation correspondantes sont donc définies comme suit :

$$\frac{1}{2}\left[(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P)^T\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_k}\mathbf{V}^{-1}(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P) - \mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_k} \right) \right] = \mathbf{0}, \tag{2.2.10}$$

où $\hat{\boldsymbol{\beta}}_P = \hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta})$.

Breslow et Clayton (1993) notent que l'approche du maximum de vraisemblance pour estimer θ_k ne tient pas compte de la perte de degrés de liberté associée à l'estimation de $\boldsymbol{\beta}$ dans l'équation 2.2.10. Pour cette raison, les estimateurs du

maximum de vraisemblance sont généralement biaisés lorsque le nombre de degrés de liberté est petit. Ce problème est éliminé en utilisant l'approche du maximum de vraisemblance restreint. L'estimateur PQL de la variance est alors la valeur de $\boldsymbol{\theta}$ qui maximise

$$-\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P)^T\mathbf{V}^{-1}(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P),$$

où $\hat{\boldsymbol{\beta}}_P = \hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta})$.

Les dérivées partielles de l'équation précédente par rapport aux composantes de $\boldsymbol{\theta}$ permettent d'obtenir les équations d'estimation PQL suivantes pour les paramètres de la variance:

$$\frac{1}{2}\left[(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P)^T\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}\mathbf{V}^{-1}(\mathbf{Y}^* - \boldsymbol{X}\hat{\boldsymbol{\beta}}_P) - \mathrm{tr}\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_k} \right) \right] = \mathbf{0}, \quad (2.2.11)$$

où $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ (Harville, 1977) et $\hat{\boldsymbol{\beta}}_P = \hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta})$. La matrice d'information de Fisher correspondante est formé des composantes

$$-\frac{1}{2}\mathrm{tr}\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_k}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_l} \right).$$

L'approche du maximum de vraisemblance restreint est justifiée lorsque $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$ sont des paramètres orthogonaux au sens de Cox et Reid (1987) et que la matrice d'information de Fisher pour $\hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta})$ est $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$. Notons qu'aucune de ces deux spécifications n'est exacte pour le modèle logistique-normal.

2.2.3. Algorithme

Un des principaux avantages de l'approche PQL est qu'elle peut être programmée en ajustant itérativement un modèle linéraire mixte à la variable réponse "de travail" $\mathbf{Y}^*$. Wolfinger (1993a) a d'ailleurs développé une macro SAS nommée GLIMMIX qui permet d'obtenir aisément les estimateurs PQL.

Cette macro calcule les estimateurs basés sur la pseudo-vraisemblance (PL) et la pseudo-vraisemblance restreinte (REPL) d'après l'article de Wolfinger et O'Connell (1993). Tel que mentionné à la section 1.3, l'approche PQL est un cas spécial de la méthode PL lorsque le paramètre d'échelle est égal à 1 ($\phi = 1$). Voici un bref aperçu des étapes de l'algorithme PQL de la macro GLIMMIX pour des réponses binaires.

1) Calculer l'estimateur initial de $\boldsymbol{\mu}^{\tilde{\mathbf{b}}}$, noté $\hat{\boldsymbol{\mu}}^{\tilde{\mathbf{b}}}$, en utilisant un ajustement qui permet d'utiliser la fonction de lien. Par exemple, une donnée binaire y_{ij} pourrait être remplacée par $(y_{ij} + 0.5)/2$;
2) Calculer le vecteur $\mathbf{Y}^*$ formé des éléments $h(\hat{\mu}_{ij}^{\mathbf{b}}) + (y_{ij} - \hat{\mu}_{ij}^{\mathbf{b}})[v(\hat{\mu}_{ij}^{\mathbf{b}})]^{-1}$ (en énumérant les t_i temps du sujet 1 au sujet m);
3) Utiliser une procédure d'estimation du maximum de vraisemblance restreint pour ajuster un modèle linéaire mixte à la variable réponse $\mathbf{Y}^*$; ce qui revient à résoudre les équations d'estimation 2.2.10. Cette procédure permet d'obtenir l'estimateur PQL des composantes de la variance noté $\hat{\boldsymbol{\theta}}_P$;
4) Comparer les anciens estimateurs $\hat{\boldsymbol{\theta}}_P$ avec les nouveaux. Si la différence est suffisamment petite, arrêter, sinon aller à l'étape suivante;
5) Résoudre les équations du modèle mixte pour $\hat{\boldsymbol{\beta}}_P$ et $\hat{\mathbf{b}}$ où $\hat{\mathbf{b}}(\boldsymbol{\theta}) = \tilde{\mathbf{b}}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))$. Les estimateurs correspondent à ceux donnés par les équations 2.2.8 et 2.2.9;
6) Calculer le nouveaux estimateurs des $\boldsymbol{\mu}_{ij}^{\tilde{\mathbf{b}}}$ en substituant les nouveaux $\hat{\boldsymbol{\beta}}_P$ et $\hat{\mathbf{b}}$ dans l'expression $\hat{\boldsymbol{\mu}}^{\mathbf{b}} = h^{-1}(\boldsymbol{X}\hat{\boldsymbol{\beta}}_P + \mathbf{Z}\hat{\mathbf{b}})$. Retourner à l'étape 2.

Les étapes 3 à 5 sont effectués à l'aide de la procédure PROC MIXED de SAS (Littell *et al.*, 1996). Notons enfin que dans l´étape 3, l'algorithme de Newton-Raphson est préférable à l'algorithme EM pour les raisons discutées par Lindstrom et Bates (1988).

2.2.4. Convergence

Comme mentionné précédemment, les estimateurs PQL $\hat{\boldsymbol{\beta}}_P$ et $\hat{\boldsymbol{\theta}}_P$ résolvent conjointement les équations 2.2.8 et 2.2.10. Ces équations d'estimation nécessitent plusieurs approximations pour lesquelles aucune justification analytique n'a été donnée. Généralement, les estimateurs PQL convergent lorsque les Y_{ij} sont de loi normale. Par conséquent, les estimateurs sont sérieusement biaisés lorsque les données sont de type binaire. Pour une bonne discussion de ces points, voir Breslow et Clayton (1993, sec. 2.5).

Breslow et Lin (1995) étudient le biais asymptotique des estimateurs PQL pour des données appariées avec des réponses binaires. Leurs études numériques montrent que l'estimateur des composantes de la variance est sérieusement biaisé. Lin et Breslow (1996a) obtiennent des formules générales du biais asymptotique des estimateurs PQL dans le cas des MLGM avec une fonction de lien canonique et plusieurs effets aléatoires indépendants.

2.3. Correction du biais des estimateurs PQL

Il y a deux sources possibles du biais de $\hat{\boldsymbol{\beta}}_P$. Premièrement, $\hat{\boldsymbol{\beta}}_P$ dépend de la matrice $\boldsymbol{\theta}$ qui doit être estimée. Deuxièmement, les équations d'estimation PQL pour $\hat{\boldsymbol{\beta}}_P$ lorsqu'on suppose $\boldsymbol{\theta}$ connu sont des approximations qui peuvent occasionner un biais. Il est possible de diminuer le biais de la première source en utilisant les composantes corrigées de $\hat{\boldsymbol{\theta}}_P$ dans l'équation 2.2.8. Le biais provenant de la seconde source peut être réduit en écrivant $\hat{\boldsymbol{\beta}}_P$ comme une fonction polynômiale de $\boldsymbol{\theta}$ pour des petites valeurs de $\boldsymbol{\theta}$.

2.3.1. Biais des estimateurs des coefficients de régression

Solomon et Cox (1992) font une approximation de l'intégrale de la vraisemblance (2.2.1) à l'aide d'un développement de Taylor autour de $\mathbf{b} = \mathbf{0}$ dans le cas spécifique où $c = 1$. Leur approximation est à la base des formules de correction du biais des estimateurs PQL présentées dans cette section. Lin et Breslow (1996a) généralisent les approximations proposées par Solomon et Cox (1992) au MLGM avec plusieurs composantes de dispersion ($c > 1$). Ils en déduisent un développement quadratique de $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ autour de $\boldsymbol{\theta} = \mathbf{0}$:

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \sum_{i=1}^{m} \sum_{j=1}^{t_i} l_{ij}(\boldsymbol{\beta}; \mathbf{0}) + (\boldsymbol{\ell}_{11} + \boldsymbol{\ell}_{12})^T \boldsymbol{\theta} \\ &\quad + \frac{1}{2} \boldsymbol{\theta}^T (\boldsymbol{\ell}_{21} + \boldsymbol{\ell}_{22} + \boldsymbol{\ell}_{23} + \boldsymbol{\ell}_{24}) \boldsymbol{\theta} + o(\|\boldsymbol{\theta}\|^2), \end{aligned} \tag{2.3.1}$$

où $\boldsymbol{\ell}_{11}$ et $\boldsymbol{\ell}_{12}$ sont des vecteurs de dimensions $c \times 1$ et $\boldsymbol{\ell}_{21}$, $\boldsymbol{\ell}_{22}$, $\boldsymbol{\ell}_{23}$ et $\boldsymbol{\ell}_{24}$ des matrices de dimensions $c \times c$ avec les composantes

$$\begin{aligned} \boldsymbol{\ell}_{11}(k) &= \frac{1}{2} \mathbf{r}_0^T \mathbf{Z}_k \mathbf{Z}_k^T \mathbf{r}_0 \\ \boldsymbol{\ell}_{12}(k) &= -\frac{1}{2} \mathrm{tr}(\mathbf{Z}_k^T \mathbf{W}_0 \mathbf{Z}_k) \\ \boldsymbol{\ell}_{21}(k,l) &= -\mathbf{r}_0^T \mathbf{Z}_k \mathbf{Z}_k^T \mathbf{W}_0 \mathbf{Z}_l \mathbf{Z}_l^T \mathbf{r}_0 \\ \boldsymbol{\ell}_{22}(k,l) &= \frac{1}{2} \mathrm{tr}(\mathbf{Z}_k^T \mathbf{W}_0 \mathbf{Z}_l \mathbf{Z}_l^T \mathbf{W}_0 \mathbf{Z}_k) = \frac{1}{2} \mathbf{1}_m^T (\mathbf{Z}_k^T \mathbf{W}_0 \mathbf{Z}_l)^{(2)} \mathbf{1}_m \\ \boldsymbol{\ell}_{23}(k,l) &= -\frac{1}{2} \left[\mathbf{1}_m^T \mathbf{Z}_l^{(2)T} \mathbf{W}_1 \mathbf{Z}_k \mathbf{Z}_k^T \mathbf{r}_0 + \mathbf{1}_m^T \mathbf{Z}_k^{(2)T} \mathbf{W}_1 \mathbf{Z}_l \mathbf{Z}_l^T \mathbf{r}_0 \right] \\ \boldsymbol{\ell}_{24}(k,l) &= -\frac{1}{4} \mathbf{1}_m^T \mathbf{Z}_k^{(2)T} \mathbf{W}_2 \mathbf{Z}_l^{(2)} \mathbf{1}_m. \end{aligned}$$

Notons que $\mathbf{W}_0 = \mathbf{W}|_{\mathbf{b}=\mathbf{0}}$, $\mathbf{r}_0 = \mathbf{r}_\mathbf{b}|_{\mathbf{b}=\mathbf{0}}$ et $\mu_{ij}^0 = \mu_{ij}^\mathbf{b}|_{\mathbf{b}=\mathbf{0}}$. De plus, $\mathbf{W}_1$ et $\mathbf{W}_2$ sont des matrices diagonales avec les éléments $v(\mu_{ij}^0)(1 - 2\mu_{ij}^0)$ et $v(\mu_{ij}^0)(1 - 6v(\mu_{ij}^0))$.

Enfin, la notation suivante est utilisée : $\mathbf{H}^{(2)} = \{h_{ij}^2\}$ pour toute matrice $\mathbf{H}$ et $\mathbf{1}_s$ est un vecteur de 1 de dimensions $s \times 1$.

En utilisant les équations 2.2.6 et 2.3.1, nous déduisons la relation suivante entre l_P (équation 2.2.6) et l (équation 2.3.1) pour de petites valeurs de $\boldsymbol{\theta}$:

$$l_P = l - \boldsymbol{\ell}_{12}^T\boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^T(\boldsymbol{\ell}_{22} + \boldsymbol{\ell}_{23} + \boldsymbol{\ell}_{24})\boldsymbol{\theta} + o(\|\boldsymbol{\theta}\|^2). \tag{2.3.2}$$

D'après Breslow et Lin (1995), le développement linéaire de l'équation score correspondante autour de $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_P$ est

$$\left.\frac{\partial l}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \left.\frac{\partial l}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} + \left.\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_P) = \mathbf{0},$$

où $\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_P\| \leq \|\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_P\|$. L'estimateur du maximum de vraisemblance $\hat{\boldsymbol{\beta}}$ et son approximation $\hat{\boldsymbol{\beta}}_P$ (l'estimateur PQL) sont donc reliés par l'équation

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_P - \left(\left.\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} \right)^{-1} \left.\frac{\partial l}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} + O(\|\hat{\boldsymbol{\beta}}_P - \hat{\boldsymbol{\beta}}\|^2). \tag{2.3.3}$$

2.3.1.1. *Correction d'ordre 1*

Pour obtenir la correction d'ordre 1, nous évaluons le développement linéaire du second terme du côté droit de l'équation 2.3.3 autour de $\boldsymbol{\theta} = \mathbf{0}$. Il s'agit de développer les deux dérivées partielles séparément autour de $\boldsymbol{\theta} = \mathbf{0}$. Premièrement, de l'équation 2.3.1, il suit que

$$\begin{aligned} \left.\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} &= \sum_{i=1}^{m}\sum_{j=1}^{t_i} \left.\frac{\partial l_{ij}(\boldsymbol{\beta};\mathbf{0})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} + O(\|\boldsymbol{\theta}\|) \\ &= -\mathbf{X}^T\mathbf{W}_0\mathbf{X}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} + O(\|\boldsymbol{\theta}\|) \end{aligned} \tag{2.3.4}$$

Deuxièmement, de l'équation 2.3.2, nous déduisons que l et l_P sont en relation d'après reliés par

$$l = l_P + \boldsymbol{\ell}_{12}^T\boldsymbol{\theta} + o(\|\boldsymbol{\theta}\|).$$

En dérivant de chacun des cotés de l'égalité par rapport à $\boldsymbol{\beta}$ et en évaluant l'expression au point $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_P$, nous obtenons

$$\begin{aligned} \left.\frac{\partial l}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} &= \left.\frac{\partial l_P}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} + \left(\left.\frac{\partial \boldsymbol{\ell}_{12}^T}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P} \right) \boldsymbol{\theta} + o(\|\boldsymbol{\theta}\|) \\ &= -\frac{1}{2}(\mathbf{X}^T\mathbf{W}_1\mathbf{Z}^{(2)}\mathbf{J}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P})\boldsymbol{\theta} + o(\|\boldsymbol{\theta}\|), \end{aligned} \tag{2.3.5}$$

où $\mathbf{J} = \text{diag}(\mathbf{1}_m, ..., \mathbf{1}_m)$ est une matrice diagonale en blocs de dimensions $cm \times c$.

La substitution de 2.3.4 et 2.3.5 dans l´équation 2.3.3 mène à

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_P - \frac{1}{2}\left[(\mathbf{X}^T\mathbf{W}_0\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}_1\mathbf{Z}^{(2)}\mathbf{J}\right]\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P}\boldsymbol{\theta} + o(\|\boldsymbol{\theta}\|).$$

Puisque $\hat{\boldsymbol{\beta}}$ converge vers $\boldsymbol{\beta}$ lorsque $m \to \infty$, le biais asymptotique de $\hat{\boldsymbol{\beta}}_P$ est d'ordre $\boldsymbol{\theta}$. L'estimateur PQL d'ordre 1 corrigé est donc

$$\hat{\boldsymbol{\beta}}_{CP1} = \hat{\boldsymbol{\beta}}_P - \frac{1}{2}\left[(\mathbf{X}^T\mathbf{W}_0\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}_1\mathbf{Z}^{(2)}\mathbf{J}\right]\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P}\boldsymbol{\theta}. \tag{2.3.6}$$

2.3.1.2. *Correction d'ordre 2*

Les résultats numériques pour des données appariées (Breslow et Lin, 1995) et les études de simulation pour des données binaires (Lin et Breslow, 1996b) suggèrent que $\hat{\boldsymbol{\beta}}_{CP1}$ surestime le biais réel, spécialement pour des petites valeurs de $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$. Lin et Breslow (1996a) proposent donc une deuxième correction de l'estimateur PQL des coefficients.

Cette correction est obtenue à l'aide du développement en série de Taylor d'ordre 2 du second terme de l'équation 2.3.3 autour de $\boldsymbol{\theta} = \mathbf{0}$. Lin et Breslow (1996a) définissent l'estimateur corrigé de deuxième ordre de la façon suivante:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{CP2} &= \hat{\boldsymbol{\beta}}_{CP1} + \frac{1}{2}(\mathbf{X}^T\mathbf{W}_0\mathbf{X})^{-1} \\ &\quad \left\{\sum_{k=1}^{c}\sum_{l=1}^{c}\mathbf{X}^T\mathbf{W}_1[(\mathbf{Z}_k\mathbf{Z}_k^T\mathbf{W}_0\mathbf{Z}_l)\cdot\mathbf{Z}_l]\mathbf{1}_m\theta_k\theta_l\right\}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_P}, \end{aligned} \tag{2.3.7}$$

où l'opérateur $\cdot$ représente la multiplication des matrices élément par élément. Le vecteur $\hat{\boldsymbol{\beta}}_{CP2}$ tel que défini en 2.3.7 est en fait une correction "partielle" de deuxième ordre, car seul le terme quadratique principal de 2.3.3 ($\boldsymbol{\theta}^T \boldsymbol{\ell}_{22} \boldsymbol{\theta}$) est pris en considération. Dans ce cas, l'équation 2.3.3 peut être récrite comme

$$l = l_P + \boldsymbol{\ell}_{12}^T \boldsymbol{\theta} + \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\ell}_{22} \boldsymbol{\theta} + O(\|\boldsymbol{\theta}\|^2).$$

D'après Lin et Breslow (1996a), le fait d'ignorer les termes quadratiques $\boldsymbol{\ell}_{23}$ et $\boldsymbol{\ell}_{24}$ dans 2.3.3 simplifie grandement la correction d'ordre 2. De plus, les résultats numériques montrent que cette correction produit des résultats satisfaisants pour des réponses binaires.

2.3.1.3. *Covariance des estimateurs*

L'approximation de la covariance de $\hat{\boldsymbol{\beta}}_{CP1}$ et $\hat{\boldsymbol{\beta}}_{CP2}$ est évaluée par

$$\begin{aligned} \operatorname{cov}(\hat{\boldsymbol{\beta}}_{CP1}) &\approx \operatorname{cov}(\hat{\boldsymbol{\beta}}_{CP2}) && (2.3.8) \\ &\approx \operatorname{cov}(\hat{\boldsymbol{\beta}}_{CP0}) + (\mathbf{X}^T \mathbf{W}_0 \mathbf{X})^{-1} \mathbf{B} \operatorname{cov}(\hat{\boldsymbol{\theta}}_{CP}) \mathbf{B}^T (\mathbf{X}^T \mathbf{W}_0 \mathbf{X})^{-1}, && (2.3.9) \end{aligned}$$

où $\mathbf{B} = (1/2)\mathbf{X}^T \mathbf{W}_1 \mathbf{Z}^{(2)} \mathbf{J}$. Notons que les matrices $\operatorname{cov}(\hat{\boldsymbol{\beta}}_{CP0})$ et $\operatorname{cov}(\hat{\boldsymbol{\theta}}_{CP})$ sont décrites à la section suivante. L'approximation de la covariance est obtenue en ignorant la corrélation asymptotique entre $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$ ainsi que la variabilité venant de $(\mathbf{X}^T \mathbf{W}_0 \mathbf{X})^{-1} \mathbf{B}$. De plus, le calcul de la covariance de $\hat{\boldsymbol{\beta}}_{CP2}$ ne tient pas compte de la variabilité provenant de $(\mathbf{X}^T \mathbf{W}_0 \mathbf{X})^{-1} \mathbf{A}$ où

$$\mathbf{A} = \frac{1}{2} \Big\{ \sum_{k=1}^{c} \sum_{l=1}^{c} \mathbf{X}^T \mathbf{W}_1 [(\mathbf{Z}_k \mathbf{Z}_k^T \mathbf{W}_0 \mathbf{Z}_l) \cdot \mathbf{Z}_l] \mathbf{1}_m \theta_k \theta_l \Big\} \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_P}.$$

En effet, la variabilité de $\mathbf{A}$ est moins importante pour des petites valeurs de $\boldsymbol{\theta}$ puisque $\mathbf{A}$ est une fonction quadratique de $\boldsymbol{\theta}$. Les simulations pour des réponses binaires présentées par Lin et Breslow (1996b) montrent que l'approximation de la covariance des estimateurs corrigés donne des résultats satisfaisants.

2.3.2. Biais des estimateurs des composantes de la variance

L'équation générale du biais asymptotique des estimateurs PQL de $\boldsymbol{\theta}$ est obtenue selon un principe semblable à celui utilisé pour l'étude du biais asymptotique de $\hat{\boldsymbol{\beta}}_P$. Le biais asymptotique est donc calculé à l'aide du développement en série de Taylor et de la vraisemblance pénalisée (équation 2.2.6).

Posons $l^\sharp(\boldsymbol{\theta}) = \log L\{\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}\}$ le logarithme de la vraisemblance de profil et $\hat{\boldsymbol{\theta}}$ l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta}$. Puisque nous nous intéressons au biais asymptotique, il n'est plus utile de tenir compte de la perte de degrés de liberté associée à l'estimation de $\boldsymbol{\beta}$. Par conséquent, les équations d'estimation de $\boldsymbol{\theta}$ notées $\partial l_P^\sharp / \partial \boldsymbol{\theta}$ sont définies selon l'équation 2.2.10 et l'estimateur PQL correspondant est noté $\hat{\boldsymbol{\theta}}_P$.

La relation entre l'estimateur du maximum de vraisemblance $\hat{\boldsymbol{\theta}}$ et l'estimateur PQL $\hat{\boldsymbol{\theta}}_P$ est étudiée en prenant le développement en série de Taylor d'ordre 1 de $\partial l^\sharp / \partial \boldsymbol{\theta}$ et de son approximation $\partial l_P^\sharp / \partial \boldsymbol{\theta}$, autour de $\boldsymbol{\theta} = \mathbf{0}$. Par ce raisonnement, Breslow et Lin (1995, sec. 5) déduisent que

$$\hat{\boldsymbol{\theta}} = \left(\frac{\partial^2 l^\sharp}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right)^{-1} \left(\frac{\partial^2 l_P^\sharp}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_P^*} \right) \hat{\boldsymbol{\theta}}_P \tag{2.3.10}$$

où $\|\boldsymbol{\theta}^*\| \leq \|\hat{\boldsymbol{\theta}}\|$, $\|\boldsymbol{\theta}_P^*\| \leq \|\hat{\boldsymbol{\theta}}_P\|$. D'après l'équation 2.3.10, le biais asymptotique de $\hat{\boldsymbol{\theta}}_P$ est d'ordre $\boldsymbol{\theta}$ lorsque $\boldsymbol{\theta} \downarrow 0$.

L'estimateur PQL corrigé est obtenu à l'aide d'une approximation des dérivées partielles d'ordre 2 de l´équation 2.3.10 sous l'hypothèse que $\boldsymbol{\theta} = \mathbf{0}$. Nous calculons d'abord

$$\frac{\partial^2 l^\sharp}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\bigg|_{\boldsymbol{\theta}=\mathbf{0}} = (-\boldsymbol{\ell}_{22} + \boldsymbol{\ell}_{24} + \mathbf{B}^T(\mathbf{X}^T\mathbf{W}_0\mathbf{X})^{-1}\mathbf{B})\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_T} + o_P(n), \tag{2.3.11}$$

où $\boldsymbol{\beta}_T$ est la vraie valeur de $\boldsymbol{\beta}$. Par la suite, sous des conditions de régularité, $\partial\hat{\boldsymbol{\beta}}_P(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_k|_{\boldsymbol{\theta}=\mathbf{0}}$ converge vers 0 en probabilité et par conséquent,

$$\left.\frac{\partial^2 l_P^\sharp}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=0} = (-\boldsymbol{\ell}_{22})\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_T} + o_P(n). \tag{2.3.12}$$

Voir Lin et Breslow (1996a, annexe B) pour les détails.

La combinaison des équations 2.3.11 et 2.3.12 dans 2.3.10 mène à l'estimateur PQL corrigé de la variance:

$$\hat{\boldsymbol{\theta}}_{CP} = \mathbf{C}^{-1}\mathbf{C}_P\hat{\boldsymbol{\theta}}_P. \tag{2.3.13}$$

$\mathbf{C}$ et $\mathbf{C}_P$ sont évalués à la vraie valeur de $\boldsymbol{\beta}$ et définis par

$$\begin{aligned}\mathbf{C} &= \frac{1}{2}\mathbf{J}^T(\mathbf{Z}^T\mathbf{W}_0\mathbf{Z})^{(2)}\mathbf{J} + \frac{1}{4}\mathbf{J}^T\mathbf{Z}^{(2)T}\mathbf{W}_2\mathbf{Z}^{(2)}\mathbf{J} \\ &\quad -\mathbf{B}^T(\mathbf{X}^T\mathbf{W}_0\mathbf{X})^{-1}\mathbf{B}\end{aligned}$$

et

$$\mathbf{C}_P = \frac{1}{2}\mathbf{J}^T(\mathbf{Z}^T\mathbf{W}_0\mathbf{Z})^{(2)}\mathbf{J},$$

En pratique, $\mathbf{C}$ et $\mathbf{C}_P$ sont évalués à la valeur de l'estimateur PQL $\hat{\boldsymbol{\beta}}_P$. De plus, afin d'améliorer la performance pour les petits échantillons, $\hat{\boldsymbol{\theta}}_P$ est l'estimateur du maximum de vraisemblance restreint donné par les équations d'estimation 2.2.10. Enfin, il est intéressant de remarquer que la matrice de correction est réduite à l'identité lorsque les réponses sont gaussiennes.

Une approximation de la covariance de $\hat{\boldsymbol{\theta}}_{CP}$ est donnée par

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}_{CP}) \approx \mathbf{C}^{-1}\mathbf{C}_P \operatorname{cov}(\hat{\boldsymbol{\theta}}_P)\, \mathbf{C}_P\mathbf{C}^{-1}. \tag{2.3.14}$$

Ce résultat ne tient pas compte de la variabilité supplémentaire provenant de la dépendance entre $\mathbf{C}^{-1}\mathbf{C}_P$ et $\hat{\boldsymbol{\beta}}_P$; la corrélation entre $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$ est également

ignorée. L'approximation est valide lorsque les éléments de la diagonale de $\mathbf{W}_0$ et leur dérivée varient lentement avec $\boldsymbol{\beta}$.

2.3.3. Algorithme

Lin et Breslow (1996a) proposent un algortihme en 4 étapes pour le calcul des estimateurs PQL corrigés:

1) Évaluer les estimateurs PQL $\hat{\boldsymbol{\beta}}_P$ et $\hat{\boldsymbol{\theta}}_P$ ainsi que leur covariance respective $\text{cov}(\hat{\boldsymbol{\beta}}_P)$ et $\text{cov}(\hat{\boldsymbol{\theta}}_P)$. Cette étape peut être réalisée à l'aide de la macro GLIMMIX de SAS décrite à la section 2.2.3.
2) Corriger $\hat{\boldsymbol{\theta}}_P$ par $\hat{\boldsymbol{\theta}}_{CP}$ (équation 2.3.13). Évaluer la covariance de $\hat{\boldsymbol{\theta}}_{CP}$ par son approximation donnée par 2.3.14.
3) Utiliser $\hat{\boldsymbol{\theta}}_{CP}$ pour réestimer $\boldsymbol{\beta}$ à l'aide de la macro GLIMMIX. Cette étape mène aux estimateurs de départ $\hat{\boldsymbol{\beta}}_{CP0}$ et $\text{cov}(\hat{\boldsymbol{\beta}}_{CP0})$.
4) Corriger $\hat{\boldsymbol{\beta}}_{CP0}$ par $\hat{\boldsymbol{\beta}}_{CP1}$ et $\hat{\boldsymbol{\beta}}_{CP2}$. Les deux corrections sont évaluées au point $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{CP0}$. L'approximation des covariances de $\hat{\boldsymbol{\beta}}_{CP1}$ et $\hat{\boldsymbol{\beta}}_{CP2}$ est donnée par l´équation 2.3.8.

En somme, $\hat{\boldsymbol{\beta}}_{CP1}$, $\hat{\boldsymbol{\beta}}_{CP2}$, $\text{cov}(\hat{\boldsymbol{\beta}}_{CP1})$ et $\text{cov}(\hat{\boldsymbol{\beta}}_{CP2})$ donnent les corrections de premier et de deuxième ordre des estimateurs PQL des coefficients de la régression ainsi que les estimateurs de leur covariance. La variable $\hat{\boldsymbol{\theta}}_{CP}$ est l'estimateur PQL corrigé des composantes de la variance et $\text{cov}(\hat{\boldsymbol{\theta}}_{CP})$ la matrice de leurs covariances estimées.

2.4. QUADRATURE GAUSSIENNE ADAPTÉE

Puisque la méthode PQL peut produire des résultats biaisés pour les données binaires, nous discutons d'une approche de quadrature numérique. Cette approche consiste à maximiser la fonction de vraisemblance sans faire d'approximation préalable afin d'obtenir des estimateurs sans biais. Davidian et Gallant (1992) proposent l'approche de la quadrature dans le cadre d'une étude pharmacodynamique de population. Pinheiro et Bates (1995) décrivent et comparent plusieurs techniques de calcul de l'intégrale de la vraisemblance. Selon eux, la quadrature gaussienne adaptée (QGA) est une des meilleures méthodes.

2.4.1. Description de l'approche QGA

La quadrature est une méthode d'approximation qui consiste à évaluer l'intégrale par une combinaison linéaire telle que

$$\int_a^b f(x)dx \approx w_1 f(z_1^*) + w_2 f(z_2^*) + ... + w_N f(z_N^*),$$
$$-\infty \leq a \leq b \leq +\infty,$$

où $z_1^*, ..., z_N^*$ sont des points prédéfinis nommés abscisses et $w_1, ..., w_N$ les N poids qui accompagnent ces points. Voir Davis et Rabinowitz (1984) pour plus de détails.

2.4.1.1. *Quadrature gaussienne*

Dans le cas particulier de la quadrature gaussienne, z_j^* et w_j, $j = 1, ..., N_{QG}$ représentent respectivement les abscisses et les poids de la fonction de Gauss-Hermite (en une dimension). Ces valeurs sont disponibles dans des tables standards pour $N_{QG} \leq 20$ (Abramowitz et Stegun, 1964 table 25.10). Pour $N_{QG} > 20$, des formules de calcul sont données par Golub et Welsch (1969).

La notation introduite au chapitre 1 sera utilisée pour le reste de cette section. Revenons maintenant à l'intégrale du maximum de vraisemblance présentée à la section 1.3 pour le modèle logistique-normal. Selon les équations 1.3.1 à 1.3.3, la fonction de vraisemblance pour $\boldsymbol{\beta}$ et $\boldsymbol{G}$ est :

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{G}; \boldsymbol{y}) &= \prod_{i=1}^{m} \int \prod_{j=1}^{t_i} g(y_{ij}|\boldsymbol{U}_i) f(\boldsymbol{U}_i; \boldsymbol{\theta}) d\boldsymbol{U}_i \\ &= \prod_{i=1}^{m} \int \exp \left[\sum_{j=1}^{t_i} y_{ij} \ln \left(\frac{\mu_{ij}}{1-\mu_{ij}} \right) + \sum_{j=1}^{t_i} \ln(1-\mu_{ij}) \right] f(\mathbf{U}_i) d\mathbf{U}_i. \end{aligned}$$

Puisque $\mathbf{U}_i \sim N(\mathbf{0}, \mathbf{G})$ où $\mathbf{G}$ est de dimensions $c \times c$, nous obtenons

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{m} \int (2\pi)^{-c/2} |\mathbf{G}|^{-1/2} \exp[\, l_i(\boldsymbol{\beta}, \mathbf{U}_i)\,] \exp(-\mathbf{U}_i^T \mathbf{G}^{-1} \mathbf{U}_i / 2) d\mathbf{U}_i$$

$$\text{où } l_i(\boldsymbol{\beta}, \mathbf{U}_i) = \sum_{j=1}^{t_i} y_{ij} \ln \left(\frac{\mu_{ij}}{1-\mu_{ij}} \right) + \sum_{j=1}^{t_i} \ln(1-\mu_{ij})$$

L'approximation de l'intégrale par la quadrature gaussienne est faite à l'aide d'une somme pondérée des abscisses prédéfinies sur l'effet aléatoire.

Les règles de la quadrature gaussienne pour les intégrales multiples sont complexes. En utilisant la structure de l'intégrale du modèle logistique-normal, il est possible de transformer le problème en une application successive de quadratures gaussiennes à une dimension. Ainsi, lorsque la dimension de $\mathbf{U}_i$ est petite (c=1 ou 2) la quadrature gaussienne peut être appliquée à l'aide d'un logiciel approprié.

Posons $\mathbf{z}^*$ un vecteur de N_{QG} abscisses d'un noyau de distribution normale centrée réduite. Par un changement de variable, la fonction de vraisemblance du

modèle logistique-normal peut être récrite de la façon suivante:

$$\prod_{i=1}^{m} \int (2\pi)^{-c/2} \exp[\, l_i(\boldsymbol{\beta}, \mathbf{G}^{T/2}\mathbf{z}^*)\,] \exp(-\|\mathbf{z}^*\|^2/2) d\mathbf{z}^*$$

$$\approx \sum_{j_1=1}^{N_{QG}} \cdots \sum_{j_c=1}^{N_{QG}} \exp[\, l_i(\boldsymbol{\beta}, \mathbf{G}^{T/2}\mathbf{z}^*_{j_1,\ldots,j_c})\,] \prod_{k=1}^{c} w_{jk},$$

où $\mathbf{z}^*_{j_1,\ldots,j_c} = (\mathbf{z}^*_{j_1}, \ldots, \mathbf{z}^*_{j_c})^T$. L'approximation correspondante du logarithme de cette vraisemblance est

$$l_{QG}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = -n\log(2\pi)/2 + \sum_{i=1}^{m} \log \left\{ \sum_{\mathbf{j}}^{N_{QG}} \exp[\, l_i(\boldsymbol{\beta}, \mathbf{G}^{T/2}\mathbf{z}^*_{j_1,\ldots,j_c})\,] \prod_{k=1}^{c} w_{jk} \right\},$$

où $\mathbf{j} = (j_1, \ldots, j_c)^T$ et $n = \sum_{i=1}^{m} t_i$.

Dans ce cas, la quadrature gaussienne peut être vue comme une version modifiée de l'intégration de Monte Carlo pour laquelle les échantillons aléatoires des $\mathbf{U}_i$ sont générés d'une distribution $N(\mathbf{0}, \mathbf{G})$. La seule distinction est que dans notre cas, les z_j^* et les poids w_j sont fixés au départ tandis qu'ils sont choisis aléatoirement pour la méthode de Monte Carlo. D'après Pinheiro et Bates (1995), l'échantillonnage d'importance ("importance sampling") est une approche plus efficace que l'intégration de Monte Carlo. Pour cette raison, ils considèrent l'approche QGA qui est l´équivalent de l´échantillonnage d'importance dans le contexte de la quadrature gaussienne.

2.4.1.2. *Quadrature gaussienne adaptée*

Il s'agit d'abord de centrer l'intégrale autour de l'estimateur bayésien empirique de $\mathbf{U}_i$ plutôt qu'en zéro. L'estimateur $\hat{\mathbf{U}}_i$ est donc le mode conditionnel de $\mathbf{U}_i$ obtenu en maximisant

$$-\kappa_i(\mathbf{U}_i) = l_i(\boldsymbol{\beta}, \mathbf{U}_i) - \frac{1}{2}\mathbf{U}_i^T \mathbf{G}^{-1} \mathbf{U}_i.$$

De la même manière, à la section 2.2.1 nous avons posé $\tilde{\mathbf{b}}$ comme le vecteur qui minimise $\kappa(\mathbf{b})$. Ensuite, nous calculons la matrice hessienne notée $\boldsymbol{\Gamma}(\mathbf{U}_i)$ provenant de la maximisation de $-\kappa_i(\mathbf{U}_i)$. Elle est donnée par

$$-\boldsymbol{\kappa}_i''(\mathbf{U}_i) = \boldsymbol{\Gamma}(\mathbf{U}_i) = \sum_{j=1}^{t_i} \mathbf{d}_{ij}\mathbf{d}_{ij}^T(1-\mu_{ij}) + \mathbf{G}^{-1}.$$

Nous supposons que $\boldsymbol{\Gamma}(\mathbf{U}_i)$ est une matrice d´échelle appropriée pour les abscisses de la quadrature.

L'approximation de la fonction du maximum de vraisemblance par la QGA est donc

$$\prod_{i=1}^{m} \int (2\pi)^{-c/2} |\boldsymbol{\Gamma}(\mathbf{U}_i)\mathbf{G}|^{-1/2} \exp\left[-\boldsymbol{\kappa}_i(\hat{\mathbf{U}}_i + \boldsymbol{\Gamma}(\mathbf{U}_i)^{-1/2}\mathbf{z}^*) + \|\mathbf{z}^*\|^2/2 \right] \times \exp(-\|\mathbf{z}^*\|^2/2) d\mathbf{z}^*$$

$$\approx \sum_{j_1=1}^{N_{QG}} \cdots \sum_{j_c=1}^{N_{QG}} \exp\left[-\kappa_i(\hat{\mathbf{U}}_i + \boldsymbol{\Gamma}(\mathbf{U}_i)^{-1/2}\mathbf{z}^*_{j_1,\ldots,j_c}) + \|\mathbf{z}^*_{j_1,\ldots,j_c}\|^2/2 \right] \prod_{k=1}^{c} w_{jk}.$$

L'approximation correspondante du logarithme de la vraisemblance est

$$l_{QGA}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = -\frac{1}{2}\left\{ n\log(2\pi)/2 + m\log|\mathbf{G}| + \sum_{i=1}^{m} \log|\boldsymbol{\Gamma}(\mathbf{U}_i)| \right\}$$
$$+ \sum_{i=1}^{m} \log\left\{ \sum_{\mathbf{j}}^{N_{QG}} \exp\left[-\kappa_i(\hat{\mathbf{U}}_i + \boldsymbol{\Gamma}(\mathbf{U}_i)^{-1/2}\mathbf{z}^*_{\mathbf{j}}) + \|\mathbf{z}^*_{\mathbf{j}}\|^2/2 \right] \prod_{k=1}^{c} w_{jk} \right\}.$$

Il est possible d'ajuster le nombre de points de quadrature N_{QG} afin d'obtenir différents niveaux d'exactitude. Dans le cas où $N_{QG} = 1$, l'approximation QGA correspond à l'approximation modifiée de Laplace décrite à la section 2.2.1 et utilisée par Wolfinger (1993b).

D'après Crouch et Spiegelman (1990), $N_{QG} = 20$ est généralement adéquat pour l'application du modèle logistique-normal à l'aide de la quadrature gaussienne. Par contre, selon Pinheiro et Bates (1995), l'approximation de l'intégrale

par la quadrature gaussienne donne des résultats adéquats lorsque le nombre d'abscisses est grand (> 100). L'avantage de cette approximation est qu'elle ne requiert pas l'estimation des modes a posteriori des effets aléatoires à chacune des itérations. Par contre, ceci ne compense pas l'inefficacité des calculs et le manque de précision de la quadrature gaussienne; l'approche QGA (quadrature gaussienne adaptée) est donc à prévilégier.

2.4.2. Algorithme

La nouvelle procédure PROC NLMIXED de la version 8 de SAS (disponible seulement depuis mars 2000) permet d'ajuster un modèle non linéaire en maximisant une approximation de l'intégrale de la vraisemblance des effets aléatoires. Différentes approximations de l'intégrale sont disponibles dont la QGA décrite précédemment. De plus, la procédure permet d'utiliser une grande variété de techniques d'optimisation.

Par défaut, PROC NLMIXED utilise l'algorithme de quasi-Newton lorsque l'approximation QGA est spécifiée. Cet algorithme est idéal pour les échantillons de taille moyenne, car contrairement à l'algorithme de Newton-Raphson, il ne requiert que les dérivées de premier ordre. En effet, la matrice des dérivées de second ordre, en l'occurrence la matrice de covariance des estimateurs, est calculée à l'aide des approximations par des différences finies (" finite difference approximations "). Il s'agit grosso modo d'évaluer numériquement les dérivées de second ordre. Comme la fonction objective et le gradient sont plus rapides à calculer que la matrice hessienne, l'algorithme de quasi-Newton est plus rapide que l'algorithme de Newton-Raphson. En fait, l'algorithme de quasi-Newton requiert plus d'itérations que celui de Newton-Raphson, mais chacune des itérations est calculée beaucoup plus rapidement. Par défaut, la technique de quasi-Newton exploite

la double mise à jour de Broyden, Fletcher, Goldfard et Shanna du facteur de Cholesky de la matrice hessiennne approximative. Les détails de cette technique sont donnés par Fletcher (1987).

Tel que mentionné précédemment, l'approche GQA pour l'approximation de l'intégrale permet d'ajuster le nombre de points de quadrature N_{QG} afin d'obtenir un niveau d'exactitude suffisant. La procédure PROC NLMIXED propose une technique pour déterminer le nombre de ces points : lorsque la différence relative entre deux calculs successifs de la vraisemblance est plus petite que $r = 1 \times 10^{-4}$, la recherche se termine et le plus petit nombre de points est utilisé pour l'optimisation subséquente. La variable r représente donc la tolérance pour sélectionner le nombre de points de quadrature. Pour la méthode QGA, la séquence de recherche pour N_{QG} est $1, 3, 5, 7, 11, 11 + s, 11 + 2s, \ldots$ où la valeur par défaut de s est 10. Mentionnons finalement que le nombre de points maximum permis par défaut est égal à 31.

Chapitre 3

COMPARAISON DES MÉTHODES PQL ET QGA PAR DES SIMULATIONS

Ce chapitre a pour but de comparer les estimateurs PQL, PQL corrigés et QGA à l'aide de simulations. Pour ce faire, nous utilisons trois modèles différents. Le premier est un modèle fréquemment étudié dans la littérature mais avec une modification du nombre d'observations dans le temps. Le deuxième correspond exactement au modèle de la littérature tandis que le dernier est semblable aux modèles analysés en sécurité routière. Pour le premier modèle, nous choisissons de faire varier le paramètre de la variance ainsi que le nombre d'individus. Par contre, seul le nombre d'individus varie pour le modèle de sécurité routière.

Pour débuter ce chapitre, nous expliquons et justifions le cadre de l'étude. Nous présentons ensuite les résultats obtenus en séparant l'analyse du premier modèle pour chacune des valeurs du paramètre de la variance. Nous examinons plus spécifiquement les diagrammes en boîtes, les moyennes, les écarts types simulés et estimés ainsi que les erreurs quadratiques moyennes des paramètres estimés. En terminant, nous discutons de manière globale des résultats obtenus pour en arriver à une recommandation générale quant à l'utilisation des différentes approches.

3.1. Cadre d'étude

Afin de comparer la performance des estimateurs PQL et QGA, nous considérons trois modèles distincts.

3.1.1. Premier modèle étudié

Nous étudions d'abord le modèle suivant :

$$logit(\mu_{ij}) = \beta_0 + \beta_1 h_j + \beta_2 x_i + \beta_3 x_i h_j + U_i,$$
$$i = 1, \ldots, m, \ j = 1, 2, 3, \tag{3.1.1}$$

où

$$h_j = \begin{cases} -3 & \text{si } j = 1, \\ 0 & \text{si } j = 2, \\ 3 & \text{si } j = 3. \end{cases}$$

$$x_i = \begin{cases} 1 & \text{avec une probabilité de } 1/2 \\ 0 & \text{avec une probabilité de } 1/2. \end{cases}$$

$$\boldsymbol{\beta}^T = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{pmatrix} = \begin{pmatrix} -2,5 & 1,0 & -1,0 & 0,5 \end{pmatrix},$$

$$U_i \sim \mathrm{N}(0, \theta).$$

D'après le modèle postulé, nous avons $t_i = t = 3$ et $c = 1$. Les coefficients des effets $\beta_0, \ldots, \beta_3$ sont posés de telle façon que la probabilité d'une réponse positive varie de 0,0041 à 0,62 lorsque $x_i = 0$, et de 0,0003 à 0,731 lorsque $x_i = 1$ avec $U_i = 0$ pour les deux valeurs de x_i. Au tableau 3.1.1, nous présentons l'effet de l'ajout de $\pm 2\sqrt{\theta}$ à la probabilité de succès pour quelques valeurs θ.

Dans un premier temps, nous générons les variables explicatives x_i d'après une loi de probabilité Bernoulli($\frac{1}{2}$). Les valeurs obtenues sont considérées fixes

TAB. 3.1.1. *Variation de la probabilité d'une réponse positive pour le modèle 3.1.1 lors de l'ajout* $\pm 2\sqrt{\theta}$ *à l'équation* $(U_i = 0)$.

θ	x_i	Variation		
0,5	0	0,0010	à	0,87
	1	0,0001	à	0,92
1,0	0	0,0006	à	0,92
	1	0,00005	à	0,95
1,5	0	0,0004	à	0,95
	1	0,00003	à	0,97

pour le reste de l'expérience. Dans un deuxième temps, nous générons les effets aléatoires U_i pour chacun des m individus à l'aide d'une loi de probabilité $\mathrm{N}(0, \theta)$. Nous pouvons ainsi calculer les valeurs de $logit(\mu_{ij})$ à l'aide des vraies valeurs des effets $\beta_0, \ldots, \beta_3$. Nous déduisons ensuite la probabilité de succès μ_{ij} en appliquant la transformation inverse. Finalement, les variables réponses y_{ij} sont obtenus en générant une loi Bernoulli(μ_{ij}).

Nous générons 200 jeux de données de $n = 3 \times m$ observations chacun. Nous croyons que 200 répétitions nous assurent une précision suffisante puisque Lin et Breslow (1996b) obtiennent une valeur maximale d'écart type de $0,5$ pour un cadre d´étude similaire mais avec $t = 7$ répétitions. De plus, nous évaluons que les simulations prendrons approximativement 40 heures lorsque $m = 1000$. Ce délai nous semble raisonnable étant donné le nombre de simulations que nous désirons effectuer.

L'expérience entière est reproduite pour $m = 250,\ 500,\ 750,\ 1000$ respectivement. De plus, pour chacune de ces simulations, nous faisons varier la valeur de θ : $\theta = (0,5\ 1,0\ 1,5)$. Nous analysons donc un total de 12 modèles, chacun d'eux étant répété 200 fois, selon trois approches différentes.

La première approche consiste à ajuster une régression logistique à l'ensemble des 200 jeux de données pour chacun des 12 modèles. Il s'agit d'une approche "naïve", car elle ne tient pas compte de l'extra-variabilité introduite par la loi

normale. Le critère relatif de convergence requiert qu'à l'itération i,

$$\frac{\mathbf{g}_i^T \mathbf{H}_i \mathbf{g}_i}{|l_i| + 1\text{E}{-}6} < valeur$$

où l_i est la valeur du logarithme de la fonction de vraisemblance, $\mathbf{g}_i$ le gradient et $\mathbf{H}_i$ l'espérance de la matrice hessienne inverse. Nous utilisons la *valeur* par défaut de la procédure PROC LOGISTIC de SAS, soit 1E−8 (notation SAS pour 1×10^{-8}).

La deuxième approche consiste à analyser les 200 jeux de données à l'aide des méthodes PQL et PQL corrigée. Pour chacune de ces méthodes, nous utilisons d'abord le maximum de vraisemblance et ensuite le maximum de vraisemblance restreint pour estimer la variance (voir section 2.2.2.2). Le paramètre initial pour l'estimation de la variance est fixé à 1 et le critère relatif de convergence pour chacun des modèles linéaires (étape 3 de la section 2.2.3) est 1E−8 (1×10^{-8}).

Pour la troisième approche, nous analysons chacun des 200 jeux de données avec la méthode QGA et ce, pour les 12 modèles. Les valeurs initiales des coefficients de la régression sont celles de l'approche naïve et la variance initiale est fixée à 1 (valeur par défaut de PROC NLMIXED). Le critère relatif de convergence est la valeur par défaut de la procédure PROC NLMIXED de SAS, soit 1E−8 (1×10^{-8}). Il s'agit une fois de plus d'un critère standard calculé en utilisant une forme quadratique du gradient et de la matrice hessienne inverse. Toujours par défaut, la borne supérieure de la longueur de l´étape initiale pour la recherche linéaire des 5 premières itérations est fixée à 1. La méthode QGA nécessite l´utilisation des techniques d'optimisation. Nous choisissons l'algorithme de quasi-Newton décrit succintement à la section 2.4.2 et l'algorithme bien connu de Newton-Raphson. Notons que l'ajustement du nombre de points de quadrature est effectué selon la technique présentée à la même section.

3.1.2. Deuxième modèle étudié

Le deuxième modèle considéré est :

$$logit(\mu_{ij}) = \beta_0 + \beta_1 h_j + \beta_2 x_i + \beta_3 x_i h_j + U_i,$$
$$i = 1, \ldots, 100,\ j = 1, \ldots, 7, \tag{3.1.2}$$

où

$$h_j = j - 4$$

$$x_i = \begin{cases} 1 & \text{avec une probabilité de } 1/2 \\ 0 & \text{avec une probabilité de } 1/2. \end{cases}$$

$$\boldsymbol{\beta}^T = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{pmatrix} = \begin{pmatrix} -2,5 & 1,0 & -1,0 & 0,5 \end{pmatrix},$$

$$U_i \sim \mathrm{N}(0, 1).$$

Comme pour le modèle 3.1.1, la probabilité de succès varie de 0,0041 à 0,62 lorsque $x_i = 0$, et de 0,0003 à 0,731 lorsque $x_i = 1$ avec $U_i = 0$ pour les deux valeurs de x_i. Nous générons 200 jeux de données de 700 observations chacun. Nous analysons le modèle 3.1.2 selon les approches logistique, PQL, PQL corrigée et QGA. Afin de limiter le nombre de comparaisons, la variance de la méthode PQL est estimée à l'aide du maximum de vraisemblance restreint seulement. De plus, seul l'algorithme de Newton-Raphson est utilisé pour la méthode QGA.

3.1.3. Troisième modèle étudié

Le dernier modèle étudié est de la forme suivante :

$$logit(\mu_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 h_j + U_i,$$
$$i = 1, \ldots m,\ j = 1, \ldots, 3, \tag{3.1.3}$$

où

$h_j = j$

$$x_{1i} = \begin{cases} 1 & \text{avec une probabilité de 0,3} \\ 0 & \text{avec une probabilité de 0,7.} \end{cases}$$

$$x_{2i} = \begin{cases} 1 & \text{avec une probabilité de 0,25} \\ 0 & \text{avec une probabilité de 0,75.} \end{cases}$$

$$x_{3i} = \begin{cases} 1 & \text{avec une probabilité de 0,20} \\ 0 & \text{avec une probabilité de 0,80.} \end{cases}$$

$$\boldsymbol{\beta}^T = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix} = \begin{pmatrix} -1,96 & -0,86 & -0,17 & 0,04 & -0,44 \end{pmatrix},$$

$U_i \sim \mathrm{N}(0, 1/2)$.

Pour ce modèle, la probabilité de succès varie de 0,01 à 0,09 lorsque $U_i = 0$. De la même manière que pour le modèle 3.1.1, nous générons 200 jeux de données de $n = 3 \times m$ observations chacun. L'expérience est répétée pour $m = 1000,\ 1500,\ 2000$. Notons en terminant que les jeux de données obtenus sont analysés selon les mêmes approches que pour le modèle 3.1.1.

3.2. Motivation

Le modèle 3.1.1 est analogue à celui utilisé par Zeger et Karim (1991) dans le cadre de l'échantillonnage de Gibbs pour $t = 7$. Il est également repris par Breslow et Clayton (1993) pour l'étude des estimateurs PQL, avec $\beta_3 = -0,5$, ce qui, selon eux, a un effet négligeable sur la comparaison des résultats. Enfin, Lin et Breslow (1996b) examinent la performance des estimateurs PQL corrigés en suivant le plan d'expérience de Zeger et Karim (1991). Ils fixent $\theta = 1$ et

$m = 100, 200$ pour des données binaires et génèrent leurs jeux de données 200 fois.

Selon les résultats obtenus par Lin et Breslow (1996b), les estimateurs PQL, en particulier les estimateurs de la variance, sont fortement biaisés négativement lorsque les données sont binaires. Ces résultats sont en accord avec ceux de Breslow et Clayton (1993). La correction proposée par Lin et Breslow (1996a) réduit efficacement le biais et l'erreur quadratique moyenne (EQM) de la variance. Elle fonctionne mieux lorsque les composantes de la variance sont petites. La correction augmente les écarts types estimés de chacun des coefficients estimés. Lorsque la variance associée à la loi normale est petite et que l'échantillon est de petite taille, il est possible que l'inflation des écarts types domine. En ce qui a trait aux coefficients de la régression, les deux corrections ne fonctionnent pas bien lorsque la variance est élevée. Au-delà des limites de l'étude, les résultats présentés par Lin et Breslow (1996b) suggèrent que la correction du biais augmente significativement la performance des estimateurs PQL pour des échantillons de grande taille. Ceci est spécialement vrai pour les problèmes qui impliquent seulement quelques observations binaires à chacun des niveaux de l'effet aléatoire.

Le pertinence de la correction des estimateurs PQL pour chaque problème particulier dépend de la taille de l'échantillon et de la forme de la distribution conditionnelle de la réponse. En sécurité routière, par exemple, les banques de données contiennent généralement de 3 à 5 observations dans le temps. Pour cette raison, et contrairement aux modèles présentés dans la littérature, le modèle 3.1.1 est posé avec $t = 3$. Nous choisissons également de faire varier θ et m : aucun des articles mentionnés précédemment n'examine le comportement des estimateurs PQL et PQL corrigés lorsque θ varie. Enfin, les estimateurs QGA sont calculés à l'aide de la nouvelle procédure PROC NLMIXED de la version 8 de SAS. À

notre connaissance, la performance de la quadrature gaussienne adaptée (QGA) dans le cadre du modèle logistique-normal n'a jamais été étudiée ni comparée à la performance des estimateurs PQL. Comme l'approche QGA ne permet pas l'utilisation du maximum de vraisemblance restreint, nous calculons, à des fins de comparaison, les estimateurs PQL et PQL corrigés en utilisant le maximum de vraisemblance pour estimer la variance.

Le modèle 3.1.2, postulé avec $t = 7$, correspond au modèle étudié par Lin et Breslow (1996b). Comme eux, nous fixons $\theta = 1$, $m = 100$ et nous générons 200 jeux de données. Les résulats pour les estimateurs PQL et PQL corrigés doivent donc logiquement être semblables à ceux de Lin et Breslow (1996b). Par contre, notre étude permet de comparer la performance de ces estimateurs avec celle des estimateurs QGA. Notons que la taille d'échantillon du modèle 3.1.2 est de 700 (7×100), ce qui correspond à peu près à la taille d'échantillon du modèle 3.1.2 lorsque $m = 250$ ($3 \times 250 = 750$).

En pratique, il est rare que les modèles 3.1.2 et 3.1.2 soient observés en sécurité routière, car les probabilité de succès sont assez élevées et dispersées. Pour cette raison, nous étudions un modèle qui n'est pas présent dans la littérature: le modèle 3.1.3. La probabilité de succès étant moins élevé, nous devons générer des échantillons de plus grandes tailles. Nous ajustons donc le modèle 3.1.3 pour des échantillons de taille $m = 1000,\ 1500,\ 2000$.

En résumé notre plan d'expérience permet de répondre aux questions suivantes :

1) Est-ce que les résultats des simulations de Lin et Breslow (1996b) tiennent encore lorsque $t = 3$?
2) Est-ce que la correction des estimateurs PQL est aussi performante que la méthode QGA?
3) Quelle est l'influence de la variance associée à la loi normale dans l'estimation des coefficients à l'aide des méthodes PQL, PQL corrigée et QGA?
4) Est-ce que les algorithmes de Newton-Raphson et de quasi-Newton donnent des résultats semblables pour les tailles d'échantillon choisies?
5) Quelle est la performance de l'approche QGA lorsque $t = 7$?
6) Quelle approche devrait-on privilégier pour la modélisation de données en sécurité routière?

3.3. RÉSULTATS DES SIMULATIONS POUR LE MODÈLE 3.1.1

Afin d'alléger la présentation des graphiques et des tableaux de résultats, les abréviations suivantes désigneront les méthodes d'estimation utilisées:

LOGIT : régression logistique,

PQL : méthode PQL utilisant le maximum de vraisemblance restreint pour estimer les composantes de la variance,

CPQL0 : méthode PQL utilisant les composantes de la variance corrigées (permet d'obtenir les valeurs initiales pour les corrections d'ordre 1 et 2),

CPQL1 : correction d'ordre 1 des estimateurs PQL,

CPQL2 : correction d'ordre 2 des estimateurs PQL,

MPQL : méthode PQL utilisant le maximum de vraisemblance pour estimer les composantes de la variance,

MCP0 : méthode MPQL utilisant les composantes de la variance corrigées (permet d'obtenir les valeurs initiales pour les corrections d'ordre 1 et 2),

MCP1 : correction d'ordre 1 des estimateurs MPQL,

MCP2 : correction d'ordre 2 des estimateurs MPQL,

QNR : méthode QGA utilisant l'algorithme de Newton-Raphson,

QQN : méthode QGA utilisant l'algorithme de quasi-Newton.

Le tableau 3.3.1 présente le nombre de simulations parmi les 200 pour lesquelles la méthode utilisée converge. Notons que la non-convergence d'un jeu de données implique que le critère de convergence ne peut être atteint par l'algorithme. Il est évident d'après ce tableau que les simulations avec une vraie valeur de $\theta = 0,5$ convergent moins souvent que les autres. Toutefois, le nombre de jeux de données pour lesquels les méthodes convergent augmente avec m. Il faut donc être très prudent dans l'interprétation des résultats des simulations lorsque

$m < 1000$ et $\theta = 0,5$. En observant les colonnes pour les autres valeurs de θ, nous constatons que presque tous les jeux de données convergent lorsque $m > 250$.

TAB. 3.3.1. *Nombre de simulations parmi les 200 pour lesquelles la méthode utilisée converge (modèle 3.1.1).*

m	Méthode	$\theta = 0,5$	$\theta = 1$	$\theta = 1,5$
250	LOGIT	200	200	200
	PQL	178	190	199
	MQLP	169	187	195
	QNR	166	187	195
	QQN	178	193	197
500	LOGIT	200	200	200
	PQL	180	200	200
	MQLP	175	200	200
	QNR	169	199	199
	QQN	185	200	200
750	LOGIT	200	200	200
	PQL	187	200	200
	MQLP	183	200	200
	QNR	180	199	200
	QQN	187	199	200
1 000	LOGIT	200	200	200
	PQL	193	200	200
	MQLP	191	200	200
	QNR	190	200	199
	QQN	194	200	200

Nous présentons maintenant les résultats des simulations effectuées à l'aide du modèle 3.1.1. Afin de décrire les 200 jeux de données générés, nous examinons les graphiques en boîtes des coefficients de la régression et des composantes de la variance pour m = 250, 500, 750, 1000. Nous calculons ensuite les valeurs moyennes des paramètres estimés selon chacune des méthodes ainsi que les écarts types simulés et estimés. Finalement, les tableaux des EQM permettent d´évaluer la performance globale des différentes approches. Cette démarche en trois étapes est répétée pour les trois valeurs de θ choisies.

3.3.1. Résultats pour $\theta = 1$

Les graphiques en boîtes pour les échantillons des 200 paramètres estimés selon différentes méthodes avec une vraie valeur $\theta = 1$ sont présentés en annexe A. Alors que nous ne remarquons rien de spécial pour les distributions des paramètres de la régression, il semble que les estimateurs QGA (méthodes QNR et QQN) du paramètre θ soient très éparpillés lorsque $m = 250$ (voir figure A.0.1). En fait, ils varient de 0,02 jusqu'à 4,34. De plus, nous distinguons plusieurs valeurs aberrantes sur les diagrammes en boîtes des valeurs estimées de θ obtenus par les méthodes QNR et QQN. Notons toutefois que la dispersion de ces estimateurs diminue lorsque m augmente : l'étendue n'est que de 2,18 lorsque $m = 1000$ (voir figure A.0.4).

Nous étudions de manière plus précise plusieurs exemples de jeux de données simulés pour lesquels des valeurs aberrantes de l'estimateur QGA de θ sont obtenues. Il est important de mentionner que le critère de convergence est vérifié pour toutes ces simulations et que tous les éléments du gradient projeté sont plus petits que 1E−3 1×10^{-3}. Il n'y a donc aucune irrégularité de la convergence des estimateurs QGA. Il est peu problable que la valeur initiale du paramètre θ cause problème puisque celle-ci est égale à 1 par défaut. En terminant, l'estimateur de θ est toujours loin de la vraie valeur peu importe les techniques d'optimisation ou de mise à jour choisies.

Le tableau 3.3.2 présente les valeurs moyennes des paramètres estimés lorsque la vraie valeur de θ est égale à 1. Les estimateurs PQL et MPQL de θ sont fortement biaisés négativement, et ce peu importe la taille de l'échantillon. Par contre, la correction proposée par Lin et Breslow (1996a) améliore considérablement la moyenne de ces estimateurs. Les estimateurs PQL et MQPL des coefficients de la régression ont également un biais négatif, mais beaucoup moins prononcé que

celui des estimateurs de θ. Les résultats pour les méthodes PQL corrigés d'ordre 1 ou 2 sont à peu près semblables. De plus, il semble que les corrections d'ordre 1 et 2 donnent des résultats légèrement moins biaisés, sauf pour le paramètre négatif β_2 qui est estimé plus adéquatement par la méthode CPQL0. En général, peu importe la taille de l'échantillon, le maximum de vraisemblance restreint donne des résultats plus proches des vraies valeurs que le maximum de vraisemblance. Il est probable que la différence entre les deux approches diminue avec le nombre d'observations dans le temps.

En examinant les résultats moyens pour la méthode QGA (QNR et QQN), nous remarquons que l'estimateur de θ est fortement biaisé positivement lorsque $m = 250$. Ceci reflète probablement le fait que les estimateurs QGA de θ ne sont pas stables lorsque l'échantillon est de petite taille. Par contre, à partir de $m = 500$, les moyennes de tous les estimateurs QGA sont raisonnables puisque qu'elles ne diffèrent des vraies valeurs qu'à la deuxième décimale. Mentionnons enfin que la technique d'optimisation n'a pas vraiment d'effet sur les estimateurs QGA, car les résultats sont semblables pour les algorithmes de Newton-Raphson et quasi-Newton.

TAB. 3.3.2. *Valeurs moyennes des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour une vraie valeur de* $\theta = 1$.

m	Méthode	θ	β_0	β_1	β_2	β_3
Vraie valeur		1,00	-2,50	1,00	-1,00	0,50
250	LOGIT	—	-2,23	0,88	-0,99	0,46
	PQL	0,35	-2,24	0,89	-0,99	0,46
	CPQL0	0,93	-2,30	0,91	-0,99	0,46
	CPQL1	0,93	-2,63	0,84	-0,85	0,52
	CPQL2	0,93	-2,51	0,83	-0,90	0,51
	MPQL	0,32	-2,23	0,89	-1,10	0,46
	MCP0	0,85	-2,29	0,91	-1,00	0,47
	MCP1	0,85	-2,59	0,84	-0,87	0,52
	MCP2	0,85	-2,49	0,83	-0,92	0,51
	QNR	1,29	-2,63	1,05	-1,12	0,53
	QQN	1,25	-2,62	1,05	-1,10	0,53
500	LOGIT	—	-2,21	0,88	-0,94	0,46
	PQL	0,31	-2,22	0,88	-0,94	0,46
	CPQL0	0,81	-2,27	0,90	-0,94	0,46
	CPQL1	0,81	-2,56	0,84	-0,82	0,51
	CPQL2	0,81	-2,47	0,83	-0,85	0,51
	MQLP	0,29	-2,22	0,88	-0,94	0,46
	MCP0	0,76	-2,27	0,90	-0,94	0,46
	MCP1	0,76	-2,54	0,84	-0,82	0,51
	MCP2	0,76	-2,46	0,83	-0,86	0,51
	QNR	1,07	-2,56	1,02	-1,04	0,52
	QQN	1,07	-2,56	1,02	-1,04	0,52
750	LOGIT	—	-2,20	0,88	-0,91	0,44
	PQL	0,31	-2,21	0,88	-0,91	0,44
	CPQL0	0,81	-2,26	0,90	-0,91	0,45
	CPQL1	0,81	-2,54	0,84	-0,78	0,50
	CPQL2	0,81	-2,46	0,83	-0,82	0,49
	MPQL	0,30	-2,21	0,88	-0,91	0,44
	MCP0	0,78	-2,25	0,90	-0,91	0,45
	MCP1	0,78	-2,53	0,84	-0,79	0,49
	MCP2	0,78	-2,45	0,83	-0,82	0,49
	QNR	1,08	-2,55	1,02	-1,01	0,51
	QQN	1,08	-2,55	1,02	-1,01	0,51
1000	LOGIT	—	-2,18	0,87	-0,90	0,44
	PQL	0,31	-2,19	0,87	-0,90	0,44
	CPQL0	0,81	-2,23	0,89	-0,90	0,44
	CPQL1	0,81	-2,52	0,82	-0,78	0,49
	CPQL2	0,91	-2,43	0,82	-0,81	0,49
	MPQL	0,31	-2,19	0,87	-0,90	0,44
	MCP0	0,78	-2,23	0,89	-0,90	0,44
	MCP1	0,78	-2,51	0,82	-0,78	0,49
	MCP2	0,78	-2,43	0,82	-0,82	0,49
	QNR	1,07	-2,52	1,01	-1,00	0,50
	QQN	1,07	-2,52	1,01	-1,00	0,50

Nous présentons les valeurs des écarts types estimés et simulés au tableau 3.3.3 pour $m = 250, 500$ et au tableau 3.3.4 pour $m = 750, 1000$. Notons d'abord que les plus grandes valeurs d´écarts types sont observées pour le paramètre β_2. Nous observons également que les écarts types estimés des paramètres fixes concordent raisonnablement avec les écarts types simulés pour la méthode PQL et ses variantes. Par contre, cette affirmation ne tient plus pour θ: les écarts types estimés sont considérablement plus élevés que les écarts types simulées. Ce phénomène, bien que de plus grande amplitude, est en accord avec la tendance décelée par les simulations de Lin et Breslow (1996b). Dans notre cas, il est possible que la différence importante entre les écarts types simulés et estimés de θ soit attribuable au petit nombre d'observations répétées par individu ($t = 3$). L'écart entre les deux valeurs diminue lentement lorsque m augmente, montrant ainsi que l'écart type estimé du paramètre de la variance est asymptotiquement adéquat. Nous croyons également que l´écart entre les différentes probabilités de succès associée au modèle 3.1.1 puisse influencer l'estimation.

Analysons maintenant les écarts types obtenus par les méthodes QGA. Nous remarquons que les écarts types estimés et simulés de θ sont très élevés pour les méthodes QNR et QQN lorsque $m = 250$. En fait, ils sont de l'ordre de l'estimateur (≈ 1). Tel que mentionné précédemment, la dispersion des estimateurs de la variance diminue lorsque m augmente. Enfin, nous ne constatons aucune différence majeure entre les écarts types estimés et simulés des coefficients de la régression. Notons toutefois que les valeurs les plus élevées sont une fois de plus observées pour le coefficient β_2.

TAB. 3.3.3. *Comparaison des écarts types estimés et simulés du modèle 3.1.1 avec 200 répétitions pour* $m = 250, 500$ *et une vraie valeur de* $\theta = 1$.

m	Méthode		θ	β_0	β_1	β_2	β_3
250	LOGIT	Sim.[a]	—	0,29	0,10	0,64	0,22
		Est.[b]	—	0,27	0,10	0,54	0,20
	PQL	Sim.	0,17	0,28	0,10	0,63	0,22
		Est.	0,34	0,27	0,10	0,54	0,20
	CPQL0	Sim.	0,46	0,29	0,11	0,63	0,22
		Est.	0,90	0,28	0,10	0,56	0,20
	CPQL1	Sim.	0,46	0,40	0,13	0,64	0,24
		Est.	0,90	0,43	0,13	0,58	0,21
	CPQL2	Sim.	0,46	0,34	0,13	0,64	0,24
		Est.	0,90	0,43	0,13	0,58	0,21
	MPQL	Sim.	0,16	0,28	0,10	0,63	0,22
		Est.	0,33	0,27	0,10	0,55	0,20
	MCP0	Sim.	0,43	0,29	0,11	0,63	0,22
		Est.	0,89	0,28	0,10	0,56	0,20
	MCP1	Sim.	0,43	0,39	0,13	0,65	0,23
		Est.	0,89	0,43	0,13	0,57	0,21
	MCP2	Sim.	0,43	0,34	0,13	0,64	0,23
		Est.	0,89	0,43	0,13	0,57	0,21
	QNR	Sim.	0,89	0,44	0,17	0,70	0,25
		Est.	0,96	0,41	0,16	0,60	0,22
	QQN	Sim.	0,91	0,44	0,17	0,70	0,25
		Est.	0,95	0,41	0,16	0,59	0,22
500	LOGIT	Sim.	—	0,20	0,07	0,38	0,14
		Est.	—	0,18	0,07	0,36	0,13
	PQL	Sim.	0,13	0,20	0,07	0,38	0,14
		Est.	0,24	0,19	0,07	0,37	0,13
	CPQL0	Sim.	0,33	0,21	0,07	0,38	0,14
		Est.	0,62	0,20	0,07	0,38	0,14
	CPQL1	Sim.	0,33	0,28	0,08	0,40	0,15
		Est.	0,62	0,30	0,09	0,39	0,14
	CPQL2	Sim.	0,33	0,25	0,09	0,39	0,15
		Est.	0,62	0,30	0,09	0,39	0,14
	MPQL	Sim.	0,13	0,20	0,07	0,38	0,14
		Est.	0,23	0,19	0,07	0,37	0,13
	MCP0	Sim.	0,33	0,21	0,07	0,38	0,14
		Est.	0,62	0,19	0,07	0,37	0,14
	MCP1	Sim.	0,33	0,28	0,08	0,40	0,15
		Est.	0,62	0,29	0,09	0,39	0,14
	MCP2	Sim.	0,33	0,24	0,08	0,39	0,15
		Est.	0,62	0,29	0,09	0,39	0,14
	QNR	Sim.	0,57	0,30	0,11	0,41	0,16
		Est.	0,61	0,28	0,11	0,40	0,15
	QQN	Sim.	0,57	0,30	0,11	0,41	0,16
		Est.	0,61	0,28	0,11	0,40	0,15

[a] écart type simulé : écart type estimé par les simulations.

[b] écart type estimé : moyenne des écarts types estimés.

TAB. 3.3.4. *Comparaison des écarts types estimés et simulés du modèle 3.1.1 avec 200 répétitions pour* $m = 750, 1000$ *et une vraie valeur de* $\theta = 1$.

m	Méthode		θ	β_0	β_1	β_2	β_3
750	LOGIT	Sim.[a]	—	0,15	0,06	0,29	0,11
		Est.[b]	—	0,15	0,06	0,29	0,11
	PQL	Sim.	0,11	0,15	0,06	0,29	0,11
		Est.	0,19	0,15	0,06	0,29	0,11
	CPQL0	Sim.	0,28	0,16	0,06	0,29	0,11
		Est.	0,50	0,16	0,06	0,30	0,11
	CPQL1	Sim.	0,28	0,22	0,07	0,30	0,12
		Est.	0,50	0,24	0,07	0,31	0,11
	CPQL2	Sim.	0,28	0,19	0,07	0,30	0,12
		Est.	0,50	0,24	0,07	0,31	0,11
	MPQL	Sim.	0,10	0,15	0,06	0,29	0,11
		Est.	0,19	0,15	0,06	0,29	0,11
	MCP0	Sim.	0,27	0,16	0,06	0,29	0,11
		Est.	0,49	0,16	0,06	0,30	0,11
	MCP1	Sim.	0,27	0,22	0,07	0,30	0,12
		Est.	0,49	0,24	0,07	0,31	0,11
	MCP2	Sim.	0,27	0,19	0,07	0,30	0,12
		Est.	0,49	0,24	0,07	0,31	0,11
	QNR	Sim.	0,46	0,24	0,09	0,32	0,12
		Est.	0,50	0,23	0,09	0,32	0,12
	QQN	Sim.	0,46	0,24	0,09	0,32	0,12
		Est.	0,50	0,23	0,09	0,32	0,12
1 000	LOGIT	Sim.	—	0,14	0,05	0,23	0,08
		Est.	—	0,13	0,05	0,25	0,09
	PQL	Sim.	0,09	0,13	0,05	0,23	0,08
		Est.	0,16	0,13	0,05	0,25	0,09
	CPQL0	Sim.	0,22	0,14	0,05	0,23	0,08
		Est.	0,42	0,14	0,05	0,25	0,10
	CPQL1	Sim.	0,22	0,18	0,06	0,23	0,09
		Est.	0,42	0,20	0,06	0,26	0,10
	CPQL2	Sim.	0,22	0,16	0,06	0,23	0,09
		Est.	0,42	0,20	0,06	0,26	0,10
	MPQL	Sim.	0,09	0,13	0,05	0,23	0,08
		Est.	0,16	0,13	0,05	0,25	0,09
	MCP0	Sim.	0,22	0,14	0,05	0,23	0,08
		Est.	0,42	0,13	0,05	0,25	0,09
	MCP1	Sim.	0,24	0,17	0,06	0,23	0,09
		Est.	0,42	0,20	0,06	0,26	0,10
	MCP2	Sim.	0,24	0,16	0,06	0,23	0,09
		Est.	0,42	0,20	0,06	0,26	0,10
	QNR	Sim.	0,38	0,19	0,08	0,26	0,10
		Est.	0,42	0,19	0,07	0,27	0,10
	QQN	Sim.	0,38	0,19	0,07	0,26	0,10
		Est.	0,42	0,19	0,07	0,27	0,10

[a] écart type simulé : écart type estimé par les simulations.

[b] écart type estimé : moyenne des écarts types estimés.

Puisque la correction du biais proposée a été développée à partir de la théorie asymptotique, nous nous attendons à ce que la performance soit meilleure lorsque le nombre d'observations augmente. Comme prévu, les EQM des estimateurs PQL et de ses variantes diminuent lorsque m augmente (voir tableau 3.3.5). Par exemple, l'EQM de θ pour la méthode CPQL0 diminue de 64% lorsque m passe de 250 à 500. Par contre, pour un m fixé, les EQM correspondant aux estimateurs des paramètres de la régression ne diminuent pas vraiment lorsqu'on applique la correction.

La performance des estimateurs QGA de θ est faible lorsque $m = 250$, car les EQM sont très élevés. Toutefois, nous constatons que les EQM diminuent lorsque $m = 500$. Cette mauvaise performance est probablement attribuable aux écarts types élevés des estimateurs de θ. En fait, peu importe la taille de l'échantillon, la performance des estimateurs PQL corrigés de θ est meilleure que celle des estimateurs QGA. Notons toutefois que la différence entre les performances des deux approches diminue losrque m augmente. En terminant, nous constatons que les valeurs des EQM pour le paramètre β_2 sont très grandes lorsque $m = 250$, mais que la performance rejoint celle des autres estimateurs lorsque m augmente.

TAB. 3.3.5. *EQM des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour une vraie valeur de* $\theta = 1$.

m	Méthode	θ	β_0	β_1	β_2	β_3
250	LOGIT	—	0,19	0,03	0,68	0,08
	PQL	0,45	0,18	0,03	0,66	0,08
	CPQL0	0,35	0,17	0,02	0,66	0,08
	CPQL1	0,35	0,28	0,05	0,71	0,09
	CPQL2	0,35	0,19	0,05	0,69	0,09
	MPQL	0,49	0,18	0,03	0,65	0,08
	MCP0	0,33	0,18	0,03	0,65	0,08
	MCP1	0,33	0,26	0,05	0,71	0,09
	MCP2	0,33	0,19	0,05	0,68	0,09
	QNR	1,25	0,32	0,05	0,79	0,10
	QQN	1,29	0,31	0,05	0,79	0,10
500	LOGIT	—	0,14	0,02	0,24	0,03
	PQL	0,50	0,13	0,02	0,24	0,03
	CPQL0	0,21	0,11	0,02	0,24	0,03
	CPQL1	0,21	0,13	0,04	0,28	0,04
	CPQL2	0,21	0,10	0,04	0,26	0,04
	MPQL	0,52	0,13	0,02	0,24	0,03
	MCP0	0,22	0,12	0,02	0,24	0,03
	MCP1	0,22	0,12	0,03	0,25	0,04
	MCP2	0,22	0,10	0,04	0,26	0,04
	QNR	0,53	0,14	0,02	0,27	0,04
	QQN	0,53	0,14	0,02	0,27	0,04
750	LOGIT	—	0,12	0,02	0,15	0,02
	PQL	0,48	0,11	0,02	0,15	0,02
	CPQL0	0,15	0,09	0,01	0,15	0,02
	CPQL1	0,15	0,08	0,03	0,18	0,02
	CPQL2	0,15	0,06	0,03	0,17	0,02
	MPQL	0,50	0,11	0,02	0,15	0,02
	MCP0	0,15	0,09	0,01	0,15	0,02
	MCP1	0,15	0,08	0,03	0,18	0,02
	MCP2	0,15	0,06	0,03	0,17	0,02
	QNR	0,35	0,09	0,01	0,17	0,02
	QQN	0,35	0,09	0,01	0,17	0,02
1 000	LOGIT	—	0,12	0,02	0,09	0,01
	PQL	0,48	0,12	0,02	0,09	0,01
	CPQL0	0,11	0,09	0,01	0,09	0,01
	CPQL1	0,11	0,05	0,04	0,12	0,01
	CPQL2	0,11	0,05	0,04	0,12	0,01
	MPQL	0,49	0,12	0,02	0,09	0,01
	MCP0	0,11	0,10	0,02	0,09	0,01
	MCP1	0,11	0,05	0,04	0,12	0,01
	MCP2	0,11	0,04	0,04	0,11	0,01
	QNR	0,24	0,05	0,01	0,11	0,02
	QQN	0,24	0,05	0,01	0,11	0,02

3.3.2. Résultats pour $\theta = 1{,}5$

En annexe B, nous présentons les graphiques en boîtes des échantillons des paramètres θ estimés selon différentes méthodes avec une vraie valeur de $\theta = 1,5$. Les mêmes remarques que pour les graphiques de l'annexe A s'appliquent. Il semble toutefois que la dispersion des estimateurs QNR et QQN de θ soit plus éparse dans le cas où $\theta = 1,5$. En effet, l'étendue des estimateurs est de 7,34 lorsque $m = 250$ et de 2,93 lorsque $m = 1000$.

Afin d'alléger le texte, seul le tableau des EQM des estimateurs est présenté dans cette section. Les tableaux des moyennes et des écarts types simulés et estimés peuvent être consultés à l'annexe C. En général, le comportement des résultats est à peu près le même que pour $\theta = 1$. Par contre, en examinant les moyennes estimées des tableaux C.0.1 et C.0.2 de l'annexe C, nous remarquons que la correction des estimateurs PQL fonctionne moins bien que dans la section précédente. Ceci s'explique par le fait que les estimateurs PQL corrigés sont obtenus en faisant un développement linéaire autour de $\theta = 0$. Il est donc normal que la correction soit moins adéquate lorsque θ s'éloigne de 0. Les tableaux C.0.3 et C.0.4 montrent que les écarts types des estimateurs de θ obtenus par les méthodes QNR et QQN sont encore plus élevés lorsque la vraie valeur de θ est fixée à 1,5. Ceci confirme le phénomène observé sur les diagrammes en boîtes.

Le tableau 3.3.6 permet d'analyser la performance des paramètres estimés selon les différentes méthodologies. Les EQM des estimateurs de θ sont sensiblement plus élevées que celles du tableau 3.3.5. Nous en concluons que la performance de ces estimateurs diminue lorsque la valeur de θ augmente, et ce pour toutes les méthodes étudiées. Plus spécifiquement, nous pouvons affirmer que ce phénomène est attribuable d'une part au biais élevé des estimateurs PQL et PQL

corrigés de θ lorsque $\theta = 1,5$, et d'autre part, à l'augmentation des écarts types des estimateurs de θ par les méthodes QGA lorsque la vraie valeur de θ augmente.

TAB. 3.3.6. *EQM des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour une vraie valeur de* $\theta = 1,5$.

m	Méthode	θ	β_0	β_1	β_2	β_3
250	LOGIT	—	0,23	0,04	0,43	0,06
	PQL	1,13	0,21	0,03	0,43	0,06
	CPQL0	0,47	0,18	0,03	0,43	0,06
	CPQL1	0,47	0,20	0,06	0,51	0,07
	CPQL2	0,47	0,14	0,07	0,47	0,07
	MPQL	1,18	0,22	0,03	0,44	0,06
	MCP0	0,47	0,18	0,03	0,44	0,06
	MCP1	0,47	0,18	0,06	0,51	0,07
	MCP2	0,47	0,14	0,07	0,47	0,07
	QNR	1,72	0,25	0,04	0,55	0,08
	QQN	1,74	0,25	0,04	0,54	0,08
500	LOGIT	—	0,21	0,04	0,16	0,02
	PQL	1,16	0,20	0,03	0,16	0,02
	CPQL0	0,31	0,15	0,02	0,16	0,02
	CPQL1	0,31	0,11	0,06	0,22	0,02
	CPQL2	0,31	0,09	0,06	0,19	0,02
	MPQL	1,20	0,20	0,03	0,16	0,02
	MCP0	0,34	0,16	0,03	0,16	0,02
	MCP1	0,34	0,10	0,06	0,22	0,02
	MCP2	0,34	0,09	0,06	0,19	0,02
	QNR	0,56	0,12	0,02	0,20	0,03
	QQN	0,56	0,12	0,02	0,20	0,03
750	LOGIT	—	0,22	0,03	0,16	0,02
	PQL	1,14	0,20	0,03	0,16	0,02
	CPQL0	0,27	0,15	0,02	0,16	0,02
	CPQL1	0,27	0,07	0,06	0,21	0,02
	CPQL2	0,27	0,06	0,06	0,19	0,02
	MPQL	1,17	0,20	0,03	0,16	0,02
	MCP0	0,29	0,16	0,02	0,16	0,02
	MCP1	0,29	0,07	0,06	0,21	0,02
	MCP2	0,29	0,06	0,06	0,19	0,02
	QNR	0,43	0,08	0,01	0,19	0,02
	QQN	0,43	0,08	0,01	0,19	0,02
1 000	LOGIT	—	0,20	0,03	0,41	0,02
	PQL	1,15	0,19	0,03	0,11	0,02
	CPQL0	0,26	0,14	0,02	0,11	0,02
	CPQL1	0,26	0,05	0,06	0,17	0,01
	CPQL2	0,26	0,05	0,06	0,14	0,01
	MPQL	1,17	0,19	0,03	0,11	0,02
	MCP0	0,28	0,14	0,02	0,11	0,02
	MCP1	0,28	0,05	0,06	0,16	0,01
	MCP2	0,28	0,05	0,06	0,14	0,01
	QNR	0,35	0,06	0,01	0,11	0,01
	QQN	0,35	0,06	0,01	0,12	0,01

3.3.3. Résultats pour $\theta = 0{,}5$

Les diagrammes en boîtes des estimateurs de θ lorsque la vraie valeur de $\theta = 0,5$ sont présentés en annexe B. Une fois de plus, les mêmes remarques que pour $\theta = 1$ s'appliquent (voir section 3.3.1). Nous constatons également que l'étendue des estimateurs QGA de θ est de 4,33 lorsque $m = 250$ contre 1,73 lorsque $m = 1000$.

Comme mentionné précédemment, nous observons un nombre non négligeable de jeux de données pour lesquels la convergence n'est pas atteinte lorsque $\theta = 0,5$ et m est petit. Pour cette raison, nous ne discutons ici que du cas où $m = 1000$, l'analyse des résultats pour $m < 1000$ étant laissée à la discrétion du lecteur. Les tableaux des moyennes des estimateurs et des écarts types estimés et simulés sont disponibles en annexe C. Nous constatons que pour $m = 1000$, l'estimateur de θ a un biais positif important pour les méthodes QNR et QQN alors qu'il n'est que légèrement biaisé pour la méthode PQL corrigée. Les estimateurs des coefficients de la régression obtenus par les méthodes CPQL2 et MCP2 sont généralement plus proches des vraies valeurs. La seule exception est l'estimateur de β_1 qui est moins biaisé en appliquant l'une ou l'autre des méthodes QGA.

Les écarts types estimés pour $m = 1000$ et $\theta = 0,5$ sont présentés au tableau C.0.8. Les mêmes remarques que pour $\theta = 1$ s'appliquent. Toutefois, nous remarquons que les écarts types simulés et estimés de θ sont légèrement plus petits que ceux du tableau 3.3.4. Finalement, les EQM sont présentées au tableau 3.3.7. La performance des estimateurs de θ est meilleure pour les estimateurs PQL ou MPQL corrigés, et ce même lorsque $m = 1000$.

TAB. 3.3.7. *EQM des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour une vraie valeur de* $\theta = 0,5$.

m	Méthode	θ	β_0	β_1	β_2	β_3
250	LOGIT	—	0,15	0,02	0,56	0,07
	PQL	0,09	0,15	0,02	0,54	0,06
	CPQL0	0,39	0,15	0,02	0,54	0,06
	CPQL1	0,39	0,28	0,03	0,57	0,07
	CPQL2	0,39	0,20	0,04	0,56	0,08
	MQLP	0,09	0,14	0,02	0,54	0,06
	MCP0	0,33	0,15	0,02	0,54	0,06
	MCP1	0,33	0,26	0,03	0,57	0,07
	MCP2	0,33	0,19	0,03	0,56	0,07
	QNR	1,18	0,31	0,05	0,60	0,08
	QQN	1,19	0,30	0,05	0,62	0,08
500	LOGIT	—	0,08	0,01	0,26	0,03
	PQL	0,11	0,08	0,01	0,26	0,04
	CPQL0	0,20	0,08	0,01	0,26	0,04
	CPQL1	0,20	0,13	0,02	0,28	0,04
	CPQL2	0,20	0,10	0,02	0,27	0,04
	MQLP	0,11	0,08	0,01	0,26	0,03
	MCP0	0,19	0,08	0,02	0,26	0,03
	MCP1	0,19	0,12	0,02	0,27	0,04
	MCP2	0,19	0,10	0,02	0,27	0,04
	QNR	0,44	0,13	0,02	0,29	0,04
	QQN	0,46	0,13	0,02	0,28	0,04
750	LOGIT	—	0,07	0,01	0,17	0,02
	PQL	0,11	0,07	0,01	0,16	0,02
	CPQL0	0,11	0,06	0,01	0,16	0,02
	CPQL1	0,11	0,07	0,01	0,18	0,02
	CPQL2	0,11	0,06	0,01	0,17	0,02
	MQLP	0,12	0,07	0,01	0,16	0,02
	MCP0	0,10	0,06	0,01	0,16	0,02
	MCP1	0,10	0,07	0,01	0,18	0,02
	MCP2	0,10	0,06	0,01	0,17	0,02
	QNR	0,22	0,08	0,01	0,17	0,02
	QQN	0,23	0,08	0,01	0,17	0,02
1000	LOGIT	—	0,05	0,01	0,13	0,02
	PQL	0,11	0,05	0,01	0,13	0,02
	CPQL0	0,09	0,05	0,01	0,13	0,02
	CPQL1	0,09	0,06	0,01	0,14	0,02
	CPQL2	0,09	0,05	0,02	0,14	0,02
	MQLP	0,12	0,05	0,01	0,13	0,02
	MCP0	0,09	0,05	0,01	0,13	0,02
	MCP1	0,09	0,06	0,01	0,14	0,02
	MCP2	0,09	0,05	0,01	0,14	0,02
	QNR	0,19	0,07	0,01	0,14	0,02
	QQN	0,19	0,07	0,01	0,13	0,02

3.4. RÉSULTATS DES SIMULATIONS POUR LE MODÈLE 3.1.2

Nous présentons les résultats des simulations effectuées d'après le modèle 3.1.2. Rappelons que pour ce modèle, la vraie valeur de θ est égale à 1 et $n = m \times 7 = 700$. Le tableau 3.4.1 présente le nombre de simulations parmi les 200 pour lesquelles la méthode utilisée converge. Nous observons une seule simulation que ne converge pas (méthode QNR). En examinant la figure 3.4.1, nous constatons que les estimateurs du paramètre θ ne sont pas très éparpillés, et ce peu importe l'approche choisie. L'estimateur de θ par la méthode QNR par exemple varie de 0,2 à 3,1.

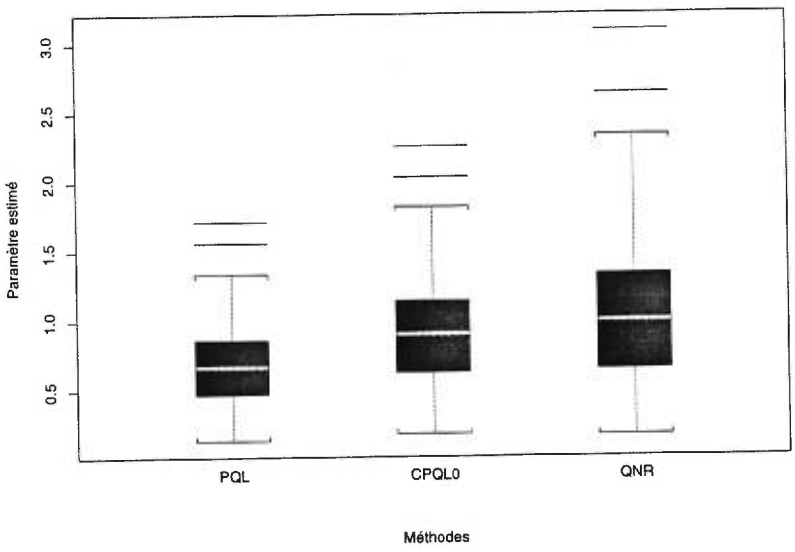

FIG. 3.4.1. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.2 avec $m = 100$ individus (vraie valeur = 1,0).*

TAB. 3.4.1. *Nombre de simulations parmi les 200 pour lesquelles la méthode utilisée converge (modèle 3.1.2).*

m	Méthode	Nombre
100	LOGIT	200
	PQL	200
	QNR	199

Le tableau 3.4.2 présente les valeurs moyennes des paramètres estimés pour le modèle 3.1.2. Nous remarquons que l'estimateur PQL de θ est fortement biaisé négativement. Par contre, la correction améliore considérablement cet estimateur. Nous constatons également que les corrections d'ordre 1 et 2 (CPQL1 et CPQL2) diminuent le biais des estimateurs β_0, β_1 et β_3, les meilleurs résultats étant obtenu à l'aide de la méthode CPQL2. Notons toutefois que la méthode CPQL0 donne l'estimateur de type PQL le moins biaisé pour β_2. En général, la méthode QNR donne les estimateurs les moins biaisés au tableau 3.4.2.

TAB. 3.4.2. *Valeurs moyennes des paramètres estimés du modèle 3.1.2 avec 200 répétitions.*

m	Méthode	θ	β_0	β_1	β_2	β_3
Vraie valeur		1,00	-2,50	1,00	-1,00	0,50
100	LOGIT	—	-2,20	0,89	-0,89	0,44
	PQL	0,68	-2,29	0,92	-0,91	0,45
	CPQL0	0,91	-2,32	0,94	-0,92	0,46
	CPQL1	0,91	-2,67	0,80	-0,78	0,53
	CPQL2	0,91	-2,39	0,83	-0,90	0,51
	QNR	1,03	-2,54	1,02	-1,03	0,51

Le tableau 3.4.3 permet de comparer les valeurs des écarts types estimés et simulés pour le modèle 3.1.2. Comme pour la section précédente, les plus grandes valeurs d'écarts types sont observées pour β_2. Par contre, nous ne remarquons ici aucune différence marquée entre les écarts types estimés et simulés. De plus, contrairement aux résultats du tableau 3.3.3 pour $m = 250$, les écarts types pour θ ne sont pas trop élevés.

Finalement, nous présentons les EQM des paramètres estimés au tableau 3.4.4. Nous constatons que la performance des estimateurs QNR est moins bonne que celle des estimateurs CPQL2. En effet, les valeurs des EQM sont plus élevées

TAB. 3.4.3. *Comparaison des écarts types estimés et simulés du modèle 3.1.2 avec 200 répétitions.*

m	Méthode		θ	β_0	β_1	β_2	β_3
100	LOGIT	Sim.[a]	—	0,28	0,11	0,50	0,20
		Est.[b]	—	0,24	0,12	0,45	0,21
	PQL	Sim.	0,29	0,29	0,11	0,50	0,20
		Est.	0,31	0,28	0,12	0,49	0,22
	CPQL0	Sim.	0,40	0,30	0,12	0,51	0,21
		Est.	0,42	0,29	0,12	0,51	0,22
	CPQL1	Sim.	0,40	0,41	0,16	0,51	0,25
		Est.	0,42	0,33	0,14	0,51	0,22
	CPQL2	Sim.	0,40	0,33	0,15	0,52	0,23
		Est.	0,42	0,33	0,14	0,51	0,22
	QNR	Sim.	0,54	0,36	0,14	0,58	0,24
		Est.	0,50	0,33	0,14	0,54	0,24

[a] écart type simulé : écart type estimé par les simulations.
[b] écart type estimé : moyenne des écarts types estimés.

pour la méthode QNR. Tel que prévu, les valeurs sont moins élevées que celle du tableau 3.3.5 pour $m = 250$.

TAB. 3.4.4. *EQM des paramètres estimés du modèle 3.1.2 avec 200 répétitions.*

m	Méthode	θ	β_0	β_1	β_2	β_3
100	LOGIT	—	0,20	0,03	0,41	0,07
	PQL	0,23	0,17	0,02	0,42	0,07
	CPQL0	0,27	0,17	0,02	0,43	0,07
	CPQL1	0,27	0,28	0,07	0,46	0,10
	CPQL2	0,27	0,18	0,06	0,45	0,09
	QNR	0,46	0,21	0,03	0,53	0,09

3.5. RÉSULTATS DES SIMULATIONS POUR LE MODÈLE 3.1.3

Nous étudions les résultats des simulations pour le modèle 3.1.3. Au tableau 3.5.1, nous remarquons que le nombre de simulations parmi les 200 pour lesquelles la méthode QNR converge est moins élevé que pour les autres méthodes. En fait, nous observons entre 14% et 21% de non-convergence pour la méthode QNR.

TAB. 3.5.1. *Nombre de simulations parmi les 200 pour lesquelles la méthode utilisée converge (modèle 3.1.3).*

m	Méthode	Nombre
1 000	LOGIT	200
	PQL	187
	MPQL	186
	QNR	158
	QQN	188
1 500	LOGIT	200
	PQL	189
	MPQL	186
	QNR	161
	QQN	191
2 000	LOGIT	200
	PQL	194
	MQLP	193
	QNR	172
	QQN	196

Les diagrammes en boîtes pour les échantillons des 200 paramètres estimés selon différentes méthodes sont présentés en annexe D. Nous remarquons que les estimateurs ne sont pas trop éparpillés peu importe l'approche utilisée.

Les valeurs moyennes des paramètres estimés sont disponibles au tableau 3.5.2. Comme pour les autres modèles, nous constatons que l'estimateur PQL de θ est inadéquat, car il est fortement biaisé négativement. De plus, bien que la correction améliore considérablement le biais de l'estimateur de θ, l'estimateur le moins biaisé est obtenu par la méthode QQN. Notons enfin que les estimateurs de $\beta_0, \ldots, \beta_4$ sont à peu près semblables peu importe l'approche choisie.

Les tableaux 3.5.3 et 3.5.4 présentent les écarts types simulés et estimés pour $m = 1000, 1500$ et $m = 2000$ respectivement. Pour la méthode PQL et ses variantes, nous remarquons un écart entre les écarts types estimés et simulés pour θ lorsque $m = 1000$. Il semble toutefois que l'écart diminue lorsque m augmente. En fait, les écarts types des différentes méthodes sont à peu près semblables pour tous les paramètres lorsque $m = 2000$.

TAB. 3.5.2. *Valeurs moyennes des paramètres estimés du modèle 3.1.3 avec 200 répétitions.*

m	Méthode	θ	β_0	β_1	β_2	β_3	β_4
Vraie valeur		0,5	-1,96	-0,86	-0,17	0,04	-0,44
1 000	LOGIT	—	-1,79	-0,85	-0,18	0,02	-0,43
	PQL	0,31	-1,79	-0,85	-0,18	0,02	-0,43
	CPQL0	0,48	-1,80	-0,85	-0,18	0,02	-0,43
	CPQL1	0,48	-1,99	-0,87	-0,19	0,02	-0,44
	CPQL2	0,48	-1,97	-0,88	-0,19	0,02	-0,44
	MPQL	0,29	-1,79	-0,85	-0,18	0,02	-0,43
	MCP0	0,46	-1,80	-0,85	-0,18	0,02	-0,46
	MCP1	0,46	-1,98	-0,87	-0,19	0,02	-0,44
	MCP2	0,46	-1,96	-0,88	-0,19	0,02	-0,44
	QNR	0,59	-2,00	-0,90	-0,20	0,03	-0,44
	QQN	0,52	-1,97	-0,88	-0,19	0,02	-0,44
1 500	LOGIT	—	-1,79	-0,84	-0,17	0,05	-0,43
	PQL	0,30	-7,80	-0,84	-0,17	0,04	-0,43
	CPQL0	0,47	-1,81	-0,83	-0,17	0,04	-0,43
	CPQL1	0,47	-1,99	-0,85	-0,17	0,04	-0,44
	CPQL2	0,47	-1,97	-0,86	-0,17	0,04	-0,44
	MPQL	0,29	-1,80	-0,84	-0,17	0,04	-0,43
	MCP0	0,46	-1,81	-0,83	-0,17	0,04	-0,43
	MCP1	0,46	-1,98	-0,85	-0,17	0,04	-0,44
	MCP2	0,46	-1,96	-0,86	-0,17	0,04	-0,44
	QNR	0,57	-2,00	-0,87	-0,17	0,04	-0,44
	QQN	0,51	-1,97	-0,87	-0,17	0,05	-0,44
2 000	LOGIT	—	-1,78	-0,84	-0,19	0,02	-0,43
	PQL	0,29	-1,79	-0,83	-0,18	0,02	-0,43
	CPQL0	0,45	-1,80	-0,83	-0,18	0,02	-0,43
	CPQL1	0,45	-1,97	-0,85	-0,19	0,02	-0,44
	CPQL2	0,45	-1,95	-0,86	-0,19	0,02	0,44
	MQLP	0,28	-1,79	-0,84	-0,19	0,02	-0,43
	MCP0	0,44	-1,79	-0,83	-0,18	0,02	-0,42
	MCP1	0,44	-1,96	-0,85	-0,19	0,02	-0,44
	MCP2	0,44	-1,95	-0,86	-0,19	0,02	-0,44
	QNR	0,53	-1,96	-0,87	-0,19	0,02	-0,44
	QQN	0,49	-1,95	-0,86	-0,19	0,02	-0,44

Nous remarquons que les paramètres β_2 et β_3 ne sont jamais significativement différents de 0 au seuil 5% (test de Student), même avec $m = 2000$. Par contre, le paramètre de θ est presque significatif lorsque m est grand.

TAB. 3.5.3. *Comparaison des écarts types estimés et simulés du modèle 3.1.3 avec 200 répétitions,* $m = 1000, 1500$.

m	Méthode		θ	β_0	β_1	β_2	β_3	β_4
1 000	LOGIT	Sim.[a]	—	0,20	0,21	0,19	0,22	0,10
		Est.[b]	—	0,21	0,21	0,19	0,20	0,10
	PQL	Sim.	0,14	0,20	0,21	0,19	0,22	0,10
		Est.	0,24	0,21	0,22	0,20	0,21	0,10
	CPQL0	Sim.	0,22	0,21	0,20	0,19	0,22	0,10
		Est.	0,38	0,21	0,22	0,20	0,21	0,10
	CPQL1	Sim.	0,22	0,24	0,21	0,20	0,22	0,11
		Est.	0,38	0,26	0,22	0,20	0,21	0,10
	CPQL2	Sim.	0,22	0,24	0,21	0,20	0,23	0,11
		Est.	0,38	0,26	0,22	0,20	0,21	0,10
	MPQL	Sim.	0,14	0,20	0,21	0,19	0,22	0,10
		Est.	0,24	0,21	0,22	0,20	0,21	0,10
	MCP0	Sim.	0,22	0,21	0,20	0,19	0,22	0,10
		Est.	0,38	0,21	0,22	0,20	0,21	0,10
	MCP1	Sim.	0,22	0,24	0,21	0,20	0,22	0,11
		Est.	0,38	0,26	0,22	0,20	0,21	0,10
	MCP2	Sim.	0,22	0,24	0,21	0,20	0,23	0,11
		Est.	0,38	0,26	0,22	0,20	0,21	0,10
	QNR	Sim.	0,23	0,24	0,21	0,21	0,23	0,11
		Est.	0,37	0,25	0,23	0,21	0,22	0,10
	QQN	Sim.	0,28	0,24	0,21	0,20	0,23	0,11
		Est.	0,37	0,25	0,23	0,21	0,21	0,10
1 500	LOGIT	Sim.	—	0,17	0,17	0,17	0,19	0,08
		Est.	—	0,17	0,17	0,16	0,16	0,08
	PQL	Sim.	0,15	0,17	0,17	0,17	0,19	0,08
		Est.	0,19	0,17	0,18	0,16	0,17	0,08
	CPQL0	Sim.	0,24	0,17	0,17	0,17	0,19	0,08
		Est.	0,31	0,17	0,18	0,16	0,17	0,08
	CPQL1	Sim.	0,24	0,21	0,17	0,17	0,19	0,08
		Est.	0,31	0,21	0,18	0,16	0,17	0,08
	CPQL2	Sim.	0,24	0,20	0,17	0,17	0,20	0,08
		Est.	0,31	0,21	0,18	0,16	0,17	0,08
	MPQL	Sim.	0,15	0,17	0,17	0,17	0,19	0,08
		Est.	0,19	0,17	0,18	0,16	0,17	0,08
	MCP0	Sim.	0,24	0,17	0,17	0,17	0,19	0,08
		Est.	0,31	0,17	0,18	0,16	0,17	0,08
	MCP1	Sim.	0,24	0,21	0,17	0,17	0,19	0,08
		Est.	0,31	0,21	0,18	0,16	0,17	0,08
	MCP2	Sim.	0,24	0,20	0,18	0,17	0,20	0,08
		Est.	0,31	0,21	0,18	0,16	0,17	0,08
	QNR	Sim.	0,29	0,20	0,18	0,18	0,19	0,08
		Est.	0,30	0,20	0,18	0,17	0,17	0,08
	QQN	Sim.	0,31	0,21	0,18	0,17	0,20	0,08
		Est.	0,29	0,20	0,18	0,17	0,17	0,08

[a] écart type simulé : écart type estimé par les simulations.

[b] écart type estimé : moyenne des écarts types estimés.

TAB. 3.5.4. *Comparaison des écarts types estimés et simulés du modèle 3.1.3 avec 200 répétitions, $m = 2000$.*

m	Méthode		θ	β_0	β_1	β_2	β_3	β_4
2 000	LOGIT	Sim.[a]	—	0,15	0,14	0,13	0,14	0,07
		Est.[b]	—	0,15	0,15	0,14	0,14	0,07
	PQL	Sim.	0,13	0,15	0,14	0,13	0,14	0,07
		Est.	0,17	0,15	0,15	0,14	0,14	0,07
	CPQL0	Sim.	0,20	0,15	0,14	0,13	0,14	0,07
		Est.	0,27	0,15	0,15	0,14	0,15	0,07
	CPQL1	Sim.	0,20	0,18	0,14	0,14	0,15	0,08
		Est.	0,27	0,18	0,15	0,14	0,15	0,07
	CPQL2	Sim.	0,20	0,17	0,14	0,14	0,15	0,08
		Est.	0,27	0,18	0,15	0,14	0,15	0,07
	MPQL	Sim.	0,13	0,15	0,14	0,13	0,14	0,07
		Est.	0,17	0,15	0,15	0,14	0,14	0,07
	MCP0	Sim.	0,20	0,15	0,14	0,13	0,14	0,08
		Est.	0,27	0,15	0,15	0,14	0,15	0,07
	MCP1	Sim.	0,20	0,18	0,14	0,14	0,15	0,08
		Est.	0,27	0,18	0,15	0,14	0,15	0,07
	MCP2	Sim.	0,20	0,18	0,14	0,14	0,15	0,08
		Est.	0,27	0,18	0,15	0,14	0,15	0,07
	QNR	Sim.	0,23	0,18	0,15	0,14	0,15	0,08
		Est.	0,26	0,17	0,16	0,14	0,15	0,07
	QGA	Sim.	0,26	0,18	0,15	0,14	0,15	0,08
		Est.	0,25	0,17	0,16	0,14	0,15	0,07

[a] écart type simulé : écart type estimé par les simulations.
[b] écart type estimé : moyenne des écarts types estimés.

Le tableau 3.5.5 permet d'évaluer la performance des estimateurs pour le modèle 3.1.3. En général, nous constatons que les EQM des paramètres estimés sont raisonnables et à peu près égales pour toutes les méthodes.

TAB. 3.5.5. *EQM des paramètres estimés du modèle 3.1.3 avec 200 répétitions.*

m	Méthode	θ	β_0	β_1	β_2	β_3	β_4
1 000	LOGIT	—	0,09	0,07	0,06	0,08	0,02
	PQL	0,06	0,09	0,07	0,06	0,08	0,02
	CPQL0	0,08	0,08	0,07	0,06	0,08	0,02
	CPQL1	0,08	0,10	0,07	0,06	0,08	0,02
	CPQL2	0,08	0,09	0,07	0,06	0,09	0,02
	MPQL	0,07	0,09	0,07	0,06	0,08	0,02
	MCP0	0,08	0,08	0,07	0,06	0,08	0,02
	MCP1	0,08	0,10	0,07	0,06	0,08	0,02
	MCP2	0,08	0,09	0,07	0,06	0,09	0,02
	QNR	0,09	0,09	0,07	0,07	0,09	0,02
	QQN	0,13	0,10	0,07	0,07	0,08	0,02
1 500	LOGIT	—	0,07	0,05	0,05	0,05	0,01
	PQL	0,07	0,07	0,05	0,05	0,05	0,01
	CPQL0	0,10	0,06	0,05	0,05	0,06	0,01
	CPQL1	0,10	0,07	0,05	0,05	0,06	0,01
	CPQL2	0,10	0,07	0,05	0,05	0,06	0,01
	MPQL	0,07	0,07	0,05	0,05	0,06	0,01
	MCP0	0,09	0,07	0,05	0,05	0,06	0,01
	MCP1	0,09	0,07	0,05	0,05	0,06	0,01
	MCP2	0,09	0,07	0,05	0,05	0,06	0,01
	QNR	0,14	0,07	0,05	0,05	0,06	0,01
	QQN	0,15	0,07	0,05	0,05	0,06	0,01
2 000	LOGIT	—	0,06	0,03	0,03	0,03	0,01
	PQL	0,06	0,06	0,03	0,03	0,03	0,01
	CPQL0	0,07	0,06	0,03	0,03	0,03	0,01
	CPQL1	0,07	0,05	0,03	0,03	0,03	0,01
	CPQL2	0,07	0,05	0,03	0,03	0,04	0,01
	MQLP	0,07	0,06	0,03	0,03	0,03	0,01
	MCP0	0,07	0,06	0,03	0,03	0,03	0,01
	MCP1	0,07	0,05	0,03	0,03	0,03	0,01
	MCP2	0,07	0,05	0,03	0,03	0,04	0,01
	QNR	0,09	0,05	0,03	0,03	0,04	0,01
	QQN	0,11	0,05	0,03	0,03	0,04	0,01

3.6. Discussion et recommandations

Dans le cadre de notre étude, les simulations sont effectuées à partir de trois modèles. Le modèle 3.1.1 reprend un modèle présent dans la littérature en modifiant le nombre d'observations dans le temps ($t = 3$). Le modèle 3.1.2 correspond exactement au modèle étudié par Lin et Breslow (1996b). Nous utilisons toutefois une approche supplémentaire afin de comparer la performance des estimateurs QGA et PQL corrigés. En pratique, il est rare qu'un tel modèle soit observé en sécurité routière, car ses probabilités de succès sont assez élevées et dispersées. Pour cette raison, nous étudions un troisième modèle qui représente plus adéquatement le type d'analyses effectuées en sécurité routière.

Notre première remarque concerne la convergence des différents algorithmes. En effet, il faut que la taille de l'échantillon soit assez élevée pour détecter des petites valeurs de θ, sinon l'algorithme a tendance à ne pas converger et les résultats sont douteux. Par exemple, dans le cadre de l'étude du modèle 3.1.1, le nombre de simulations qui convergent est adéquat seulement pour $m = 1000$ lorsque $\theta = 0,5$.

D'après nos résultats, la correction des estimateurs PQL réduit efficacement le biais pour θ. Par contre, cette correction fonctionne de moins en moins bien à mesure que la vraie valeur de θ augmente. Une autre faille de la correction de Lin et Breslow (1996a) lorsque $t = 3$ est que les écarts types estimés pour θ sont considérablement plus élevés que les écarts types simulés. Notons toutefois que la différence entre les deux valeurs est beaucoup moins grande lorsque $t = 7$ (voir tableau 3.4.3). Le biais des estimateurs PQL des paramètres $\beta_0, \beta_1, \beta_3$ pour les modèles 3.1.1 et 3.1.2 diminue lorsque nous appliquons les corrections d'ordre 1 ou 2. Toutefois, le meilleur estimateur du coefficient négatif β_2 est obtenu par l'approche CPQL0 (ou MCP0). Finalement, il est préférable d'utiliser le maximum

de vraisemblance restreint pour estimer la variance, car les résultats obtenus, bien que similaires, sont moins biaisés qu'avec le maximum de vraisemblance.

Les estimateurs QGA de θ sont très dispersés, en particulier pour m petit et θ élevée lorsque $t = 3$. Ainsi, même si le biais des estimateurs QGA est négligeable, leur performance est relativement faible dans le cadre de l'étude du modèle 3.1.1 pour $m < 500$. Par contre, la dispersion des estimateurs de θ d'apès le modèle 3.1.2 ($m = 100$ et $t = 7$) n'est pas problématique.

Les estimateurs PQL (ou MPQL) corrigés du modèle 3.1.1 ont des EQM moins grandes que celles des estimateurs QGA, et ce même pour $m = 1000$. De plus, malgré le biais négligeable des estimateurs QNR du modèle 3.1.2, la performance des estimateurs PQL corrigés est meilleure en raison de l'inflation des écarts types. Notons enfin qu'il n'y a pas de différence marquée entre les deux algorithmes d'optimisation pour les tailles d'échantillons choisies. L'algorithme de quasi-Newton est donc à privilégier, car il est plus rapide et converge plus souvent que celui de Newton-Raphson.

En sécurité routière, les banques de données comportent généralement beaucoup d'individus, peu d'observations dans le temps et des probabilités de succès petites (événements rares). Le modèle 3.1.3 reflète bien cette réalité. Les simulations de la section 3.5 montrent que les différentes approches peuvent donner des résultats différents lorsque l´échantillon est de taille moyenne. Par contre, les performances sont à peu près semblables lorsque m augmente.

À la lumière des résultats obtenus, nous sommes à même de faire des recommandations générales pour le choix d'une approche. Nos résultats montrent l'importance du nombre d'observations dans le temps lorsque l'approche spécifique aux sujets est choisie. La prudence est donc de mise lors de l'application de

l'approche QGA pour des petits échantillons surtout lorsque t est petit, car l'estimateur de la variance peut être loin de la vraie valeur de θ. Une alternative est alors l'utilisation des estimateurs PQL corrigés. Contrairement aux estimateurs PQL, les estimateurs QGA ne sont pas biaisés asymptotiquement; ils sont donc à privilégier lorsque m est grand.

Chapitre 4

ILLUSTRATION DU MODÈLE LOGISTIQUE-NORMAL

Ce chapitre est entièrement consacré à une illustration du modèle spécifique aux sujets dans le cadre d'une étude en sécurité routière. Après un bref résumé de la problématique, nous présentons la population et les variables étudiées. Nous effectuons également une analyse descriptive de ces variables. Par la suite, nous examinons un même modèle pour deux populations distinctes de conducteurs québécois. Nous comparons les résultats obtenus selon chacune des approches décrites au chapitre 2. Nous terminons en présentant l'interprétation des résultats à l'aide des risques relatifs.

4.1. PROBLÉMATIQUE

La plupart des pays industrialisés ont une législation se rapportant aux examens exigés pour avoir le privilège de conduire un véhicule. L'objectif est de s'assurer que les nouveaux conducteurs possèdent les connaissances et les habilités minimales nécessaires à la conduite automobile. Plusieurs études ont d'ailleurs été menées afin de comprendre le comportement des nouveaux conducteurs. L'équipe du Laboratoire sur la sécurité des transports du Centre de recherche sur les transports (CRT) a publié récemment une étude sur le lien entre la performance aux

examens (théorique et pratique) pour l'obtention d'un permis et le taux d'accidents (Laberge-Nadeau *et al.*, 1999). Une banque de données de la Société de l'assurance automobile du Québec (SAAQ) comprenant les accidents de nouveaux conducteurs enregistrés pendant les 3 années suivant l'obtention du permis a alors été analysée.

Nous savons que la plupart des études sur les nouveaux conducteurs d'autres pays industrialisés traitent des accidents avec victimes, i.e. les accidents avec blessés graves ou légers et les accidents avec décès. De plus, les accidents ayant donné lieu à des constats amiables ne sont pas inclus dans la banque de données de la SAAQ. Pour ces raisons, nous effectuons des analyses supplémentaires afin d'exploiter plus en détail la richesse de cette banque.

Nous examinons particulièrement l'existence d'un lien entre les résultats aux examens permettant d'obtenir un permis de conduire et l'implication ultérieure des individus dans des accidents de la route avec victimes. Il s'agit d'une étude longitudinale puisque que la réponse (accident ou non) est observée de façon répétée pour les 3 années d'observation. Nous considérons donc la corrélation dans le temps en appliquant le modèle logistique-normal.

4.2. Populations étudiées

La première banque de données déjà constituée contient 111 533 nouveaux titulaires (53 069 hommes et 58 464 femmes) ayant tous entrepris leurs démarches d'obtention d'un premier permis de classe 5 au cours de la période du 1er mars 1991 au 28 février 1993. De plus, ces nouveaux titulaires sont des aspirants conducteurs ayant obtenu le permis régulier ou probatoire de classe 5 avant le 1er janvier 1994, qui ne détiennent pas de permis d'autres classes, qui ont un dossier actif 3

ans après l'obtention du permis et qui ont moins de 90 jours d'interruption totale de leur permis de conduire.

Nous tenons également compte dans nos analyses de la plupart des 4 911 sujets (1 322 femmes et 3 589 hommes) ayant des interruptions totales supérieures à 90 jours. Parmi ces personnes, 2 536 ont eu plus de 90 jours de sanction et 2 287 ont retardé leur renouvellement de permis pour une période totale de plus de 90 jours. Les 88 autres individus dépassent 90 jours d'interruption en tenant compte à la fois des sanctions et des retards de paiement. Nous choisissons d'exclure les 66 personnes qui ont plus de 540 jours d'interruption pour retards de renouvellement de permis. En effet, ces conducteurs et conductrices ont peu d'accidents avec victimes pendant l'interruption; ils n'ont donc probablement pas conduit pendant plus de la moitié des trois années d'observation. La deuxième banque de données contient donc 4 845 nouveaux titulaires (1 295 femmes et 3 550 hommes).

En sécurité routière, les hommes ont généralement des taux d'accidents plus élevés que les femmes. Pour cette raison, les conducteurs et les conductrices sont habituellement analysés séparément. À des fins d'illustrations, nous choisissons dans ce chapitre d'examiner deux populations distinctes de conducteurs masculins. En premier lieu, nous présentons un modèle spécifique aux sujets avec interaction pour les 53 069 hommes du fichier des 111 533 nouveaux titulaires. Par la suite, nous proposons d'ajuster le même modèle pour les 3 550 conducteurs ayant eu des sanctions ou des retards de paiement de plus de 90 jours. En effet, il est intéressant d'étudier le comportement en termes d'accidents corporels ou mortels des 3 550 sujets ayant des interruptions totales supérieures à 90 jours, surtout si ces conducteurs " marginaux " ont des taux d'accidents avec victimes plus élevés que les 53 069 hommes.

Plusieurs autres modèles de ces mêmes populations pour les hommes et les femmes seront présentés sous peu dans un nouveau rapport du CRT.

4.3. Description des variables à l'étude

Nous subdivisons les trois années d'observations des accidents suivant l'obtention du permis en 3 périodes de 365 jours. La variable réponse (variable dépendante) est donc de la forme suivante :

$$y_{ij} = \begin{cases} 1 & \text{s'il y a présence d'au moins un accident avec victime(s) au cours de la } j^{\text{ième}} \\ & \text{année suivant la date d'obtention du permis de conduire de l'individu } i, \\ 0 & \text{sinon (présence d'accidents avec dommages matériels seulement} \\ & \text{ou d'aucun accident),} \end{cases}$$

où $j = 1, 2, 3$. La plupart des variables explicatives du modèle sont de type catégorique. En fait, seule la variable " expérience ", introduite pour tenir compte du nombre d'années accumulées comme nouveau conducteur est de type continue. Pour cette variable l'unité de mesure utilisée est l'année.

Voici une brève description des variables explicatives présentes dans notre modèle :

- L'âge en années révolues de la personne au moment où elle obtient son permis de conduire :

16 ans,
17 ans,
18 à 19 ans,
20 à 24 ans,
25 ans et plus;

- La date d'entrée dans le processus d'accès à la conduite :

01/03/91 au 29/02/92,
01/03/92 au 28/02/93;

- L'année d'obtention du permis de conduire :

1991,
1992,
1993;

- La durée de l'apprentissage correspondant au nombre de jours entre la date d'obtention du permis d'apprenti et la date d'obtention du permis probatoire :

  90 à 180 jours,
  181 à 270 jours,
  Plus de 270 jours;

- Le nombre de tentatives pour réussir les examens théorique et pratique :

  1 si le conducteur a réussi les deux examens dès la première tentative,
  0 sinon;

- L'expérience du conducteur en années, i.e. le nombre d'années accumulées comme nouveau conducteur.

4.4. ANALYSE DESCRIPTIVE

Comme mentionné précédemment, les trois années d'observation des accidents suivant la date d'obtention du permis sont subdivisées en 3 périodes de 365 jours. Pour chacune de ces périodes, nous calculons un taux moyen d'accidents corporels ou mortels par nouveau conducteur. Les figures 4.4.1 et 4.4.2 montrent que les taux d'accidents corporels ou mortels des nouveaux conducteurs des deux populations à l'étude diminuent lorsque l'expérience des conducteurs augmente.

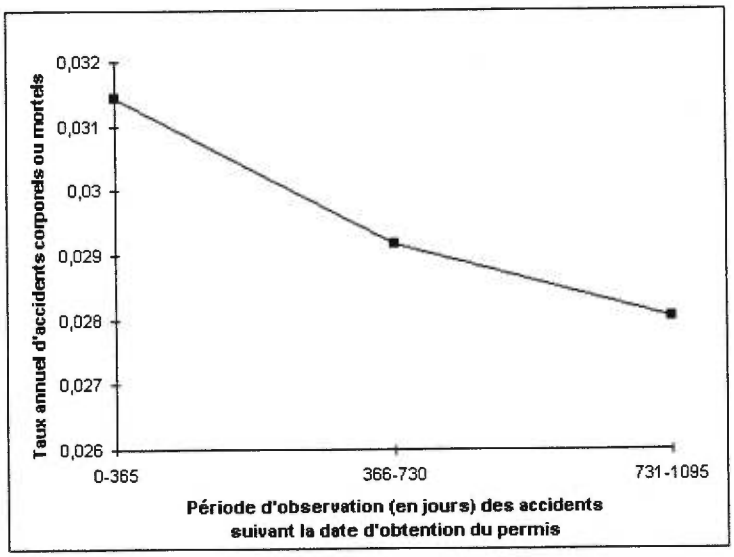

FIG. 4.4.1. *Taux annuel d'accidents avec victimes par nouveau conducteur appartenant au fichier des 53 069 conducteurs*

À la figure 4.4.2, le calcul des taux est fait en considérant que les sujets conduisent pendant leurs sanctions : il s'agit donc d'une sous-estimation des taux réels. Cette approche est justifiée, car les personnes qui ont plus de 90 jours de sanction ont en moyenne 3 sanctions pour conduite pendant sanction. De plus, pour ces personnes, les taux d'accidents avec victimes annualisés pendant et hors période de sanction sont respectivement de 0,016 et 0,087. En ce qui concerne les

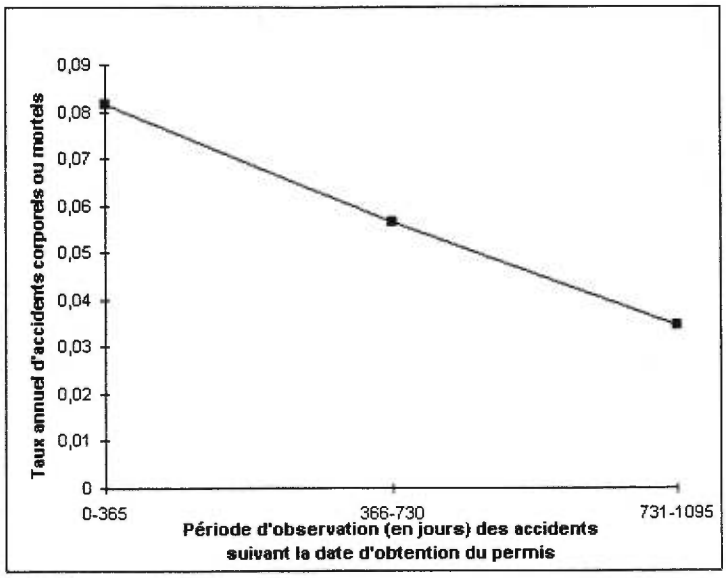

FIG. 4.4.2. *Taux annuel d'accidents avec victimes par nouveau conducteur appartenant au fichier des 3 550 conducteurs ayant plus de 90 jours d'interruption totale*

conducteurs ayant plus de 90 jours d'interruption pour retard de renouvellement, les taux d'accidents pendant et hors sanction sont respectivement 0,014 et 0,048.

Le tableau 4.4.1 présente la répartition et les taux d'accidents avec victimes par année des titulaires de permis masculins pour les trois années suivant l'obtention du permis selon différentes variables explicatives utilisées dans les modèles de régression. Notons que la première colonne de ce tableau est consacrée aux taux pour la population des 53 069 conducteurs alors que la deuxième colonne présente les taux des 3 550 conducteurs ayant plus de 90 jours d'interruption totale.

TAB. 4.4.1. *Taux d'accidents avec victimes par nouveau conducteur par an pour les trois années suivant l'obtention du permis selon différentes variables explicatives utilisées dans les modèles de régression.*

Variables explicatives	53 069 TITULAIRES		3 550 TITULAIRES	
	Nombre	Taux d'accidents avec victimes	Nombre	Taux d'accidents avec victimes
Période d'observation des accidents				
Première année	53069	0,031	3550	0,081
Deuxième année	53069	0,029	3550	0,056
Troisième année	53069	0,028	3550	0,035
Âge à l'obtention du permis				
16 ans	27298	0,030	980	0,070
17 ans	12043	0,031	1034	0,065
18-19 ans	7057	0,031	838	0,053
20-24 ans	2926	0,025	396	0,028
25 ans et plus	3745	0,022	302	0,041
Réussir l'examen théorique et pratique				
16 ans	17511	0,028	512	0,063
17 ans	5937	0,026	430	0,061
18-19 ans	3336	0,026	340	0,057
20-24 ans	1373	0,023	182	0,026
25 ans et plus	1447	0,019	134	0,035
Échoué l'examen théorique et pratique				
16 ans	9787	0,034	468	0,078
17 ans	6106	0,036	604	0,068
18-19 ans	3721	0,035	498	0,051
20-24 ans	1553	0,026	214	0,030
25 ans et plus	2298	0,025	168	0,046
Date d'entrée dans le processus d'accès				
01/03/91 au 29/02/92	26523	0,030	1417	0,056
01/03/92 au 29/02/93	29546	0,029	2133	0,059
Année d'obtention du permis				
1991	8398	0,030	271	0,069
1992	26079	0,030	1604	0,055
1993	18592	0,029	1675	0,058
Durée d'apprentissage				
90 à 180 jours	40419	0,031	2286	0,062
181 à 270 jours	6879	0,029	586	0,055
Plus de 270 jours	5771	0,024	678	0,043
Réussir l'examen théorique et pratique				
1 tentative	29604	0,027	1598	0,054
Plus d'une tentative	23465	0,033	1952	0,060
Ensemble	53069	0,030	3550	0,057

En général, la répartition des groupes d'âge n'est pas la même pour les deux populations étudiées. En effet, nous observons à peu près le même nombre de conducteur qui ont plus de 90 jours d'interruption totale (3 550 titulaires) appartenant aux catégories des 16 ans, 17 ans et 18-19 ans. Par contre, pour le fichier des 53 069 titulaires, 51,4% des conducteurs appartiennent à la catégorie des 16 ans à l'obtention du permis. En examinant la répartition des autres variables explicatives, nous observons que la majorité des conducteurs du fichier des 53 069 titulaires obtiennent leur permis à l'âge de 16 ans. Par conséquent, la plupart des conducteurs de ce fichier (76,2%) détiennent leur permis d'apprenti moins de 6 mois. Nous remarquons enfin que 56% des conducteurs du fichier des 53 069 titulaires réussissent les deux examens en une seule tentative contre 45% pour le fichier des conducteurs ayant plus de 90 jours d'interruption totale.

Examinons maintenant les taux d'accidents avec victimes du tableau 4.4.1. Nous remarquons que les nouveaux titulaires qui ont obtenu leur permis à 20 ans ont des taux d'accidents avec blessés ou décès moins élevés que les autres groupes d'âge. De plus, les taux d'accidents diminuent lorsque le nombre de jours d'apprentissage augmente. Les nouveaux titulaires masculins qui ont eu besoin d'une seule tentative pour réussir l'ensemble des examens ont des taux d'accidents corporels ou mortels moins élevés que les autres. Ce résultat se confirme pour chaque groupe d'âge, sauf pour les 18-19 ans ayant plus de 90 jours d'interruption totale. Il est toutefois intéressant de constater que les conducteurs âgés de 20 ans et plus ayant eu besoin de plus d'une tentative ont des taux d'accidents avec victimes moins élevés que les 16 ans ayant réussi les deux examens dès la première tentative. En terminant, notons que les taux d'accidents des nouveaux titulaires ayant plus de 90 jours d'interruption totale sont plus élevés que ceux du fichier des 53 069 conducteurs, et ce pour chacune des modalités des variables considérées.

4.5. MODÉLISATION

Nous présentons un modèle de régression logistique-normal avec interaction comportant 16 variables explicatives. Rappelons que l'unité d'observation est un nouveau titulaire masculin et que la variable expliquée est la présence ou non d'au moins un accident corporel ou mortel par nouveau conducteur pour la $j^{\text{ième}}$ année suivant l'obtention du permis (j varie de 1 à 3). Notre but est de vérifier si la présence ou non d'accidents avec victimes est influencée par différentes caractéristiques du titulaire (âge à l'obtention du permis, année d'obtention du permis, durée d'apprentissage et nombre de tentatives pour réussir l'examen théorique et pratique).

Pour les variables de type catégorique, une modalité est choisie comme groupe de référence. Pour une variable donnée, les coefficients associés aux modalités mesurent donc l'effet individuel de cette modalité sur la variable dépendante relative au groupe de référence. En sécurité routière, il peut y avoir de l'interaction entre l'âge à l'obtention et la performance deux examens théorique et pratique (Laberge-Nadeau *et al.*, 1999). Nous créons donc une variable à dix catégories qui décrit simultanément ces deux variables. Nous choisissons comme groupe de référence la catégorie des 16 ans à l'obtention du permis qui ont réussi les deux examens dès la première tentative.

Nous utilisons la méthode QQN décrite à la section 3.3 afin de modéliser le fichier des 53 069 conducteurs. En effet, d'après les simulations présentés à la section 3.5 et les résultats théoriques, l'approche QQN est conseillé lorsque m est grand. Comme le modèle avec interaction présenté ici s'apparente au modèle 3.1.3 de la section 3.5 et que le nombre d'individus est très élevé ($m = 53\ 069$), la méthode QQN est un choix judicieux pour l'analyse des 53 069 conducteurs.

En ce qui concerne les 3 550 conducteurs ayant plus de 90 jours d'interruption totale, nous proposons de comparer les résultats obtenus selon différentes approches puisque le nombre d'individus n'est pas très élevé. Notre but est de vérifier si les résultats obtenus à l'aide des méthodes LOGIT, PQL, CPQL0, CPQL1, CPQL2 et QQN (voir section 3.3) sont semblables.

4.5.1. Comparaisons des résultats obtenus selon différentes approches pour les 3 550 conducteurs

Les tableaux 4.5.1 et 4.5.2 présentent respectivement les coefficients estimées et les écarts types estimés du modèle de régression pour les conducteurs ayant plus de 90 jours d'interruption totale. Nous remarquons que les coefficients et les écarts types estimées sont à peu près égaux peu importe l'approche utilisée.

L'estimateur de la composante de la variance n'est pas significativement différent de zéro au seuil 5% et ce, quelque soit la méthodologie utilisée. Par conséquent, les résultats obtenus par la régression logistique sont semblables à ceux des autres méthodes qui tiennent comptent de la corrélation dans le temps. Il semble que l'estimateur de la composante de la variance soit plus élevée pour méthode PQL corrigée. Par contre, cette légère différence n'est pas inquiétante puisque les rapports des estimateurs de θ divisés par les écarts types correspondants sont près de 0 dans tous les cas (estimateurs de θ non significatifs).

L'interprétation des résultats pour le fichier des 3 550 conducteurs peut donc être effectuée à l'aide de n'importe quelle méthodologie présentée au tableau 4.5.1. Nous choisissons toutefois de présenter l'approche QQN dans la section suivante afin de comparer les résultats avec ceux des 53 069 conducteurs.

TAB. 4.5.1. *Comparaison des coefficients estimés par le modèle logistique-normal selon différentes méthodes pour le fichier des 3 550 conducteurs.*

Variables explicatives	LOGIT	PQL	CPQL0	CPQL1	CPQL2	QQN
Constante	-1,734	-1,734	-1,734	-1,768	-1,767	-1,754
Période d'observation des accidents	-0,435	-0,435	-0,435	-0,436	-0,436	-0,436
Date d'entrée dans le processus d'accès						
01/03/91 au 29/02/92	0,008	0,007	0,007	0,007	0,007	0,008
01/03/92 au 29/02/93	.....	.....	.....	.....	.....	.....
Année d'obtention du permis						
1991	.....	.....	.....	.....	.....	.....
1992	-0,210	-0,210	-0,210	-0,212	-0,212	-0,210
1993	-0,108	-0,108	-0,108	-0,109	-0,109	-0,107
Durée d'apprentissage						
90 à 180 jours	.....	.....	.....	.....	.....	.....
181 à 270 jours	-0,187	-0,187	-0,187	-0,188	-0,188	-0,188
Plus de 270 jours	-0,325	-0,325	-0,325	-0,327	-0,327	-0,327
Nombre de tentatives pour réussir l'examen théorique et pratique						
Une tentative						
16 ans	.....	.....	.....	.....	.....	.....
17 ans	0,047	0,047	0,048	0,049	0,049	0,047
18-19 ans	-0,039	-0,038	-0,038	-0,038	-0,038	-0,038
20-24 ans	-0,853	-0,854	-0,854	-0,854	-0,854	-0,855
25 ans et plus	-0,571	-0,571	-0,571	-0,574	-0,574	-0,571
Plus d'une tentative						
16 ans	0,189	0,189	0,189	0,187	0,187	0,190
17 ans	0,126	0,126	0,126	0,126	0,126	0,126
18-19 ans	-0,135	-0,135	-0,135	-0,136	-0,136	-0,136
20-24 ans	-0,738	-0,738	-0,738	-0,741	-0,741	-0,740
25 ans et plus	-0,299	-0,299	-0,299	-0,302	-0,302	-0,302
Composante de la variance	.....	0,056	0,089	0,089	0,089	0,053

TAB. 4.5.2. *Comparaison des écarts types estimés par le modèle logistique-normal selon différentes méthodes pour le fichier des 3 550 conducteurs.*

Variables explicatives	LOGIT	PQL	CPQL0	CPQL1	CPQL2	QQN
Constante	0,221	0,222	0,222	0,238	0,238	0,232
Période d'observation des accidents	0,054	0,054	0,054	0,054	0,054	0,054
Date d'entrée dans le processus d'accès						
01/03/91 au 29/02/92	0,116	0,116	0,117	0,117	0,117	0,117
01/03/92 au 29/02/93	.....	.....	.....	.....	.....	.....
Année d'obtention du permis						
1991	.....	.....	.....	.....	.....	.....
1992	0,168	0,169	0,169	0,169	0,169	0,169
1993	0,202	0,202	0,203	0,203	0,203	0,203
Durée d'apprentissage						
90 à 180 jours	.....	.....	.....	.....	.....	.....
181 à 270 jours	0,125	0,125	0,125	0,126	0,126	0,125
Plus de 270 jours	0,137	0,138	0,138	0,138	0,138	0,138
Nombre de tentatives pour réussir l'examen théorique et pratique						
Une tentative						
16 ans	.....	.....	.....	.....	.....	.....
17 ans	0,160	0,161	0,162	0,162	0,162	0,161
18-19 ans	0,177	0,178	0,179	0,179	0,179	0,179
20-24 ans	0,293	0,294	0,294	0,294	0,294	0,294
25 ans et plus	0,293	0,294	0,295	0,295	0,295	0,295
Plus d'une tentative						
16 ans	0,149	0,150	0,150	0,150	0,150	0,150
17 ans	0,148	0,149	0,149	0,149	0,149	0,149
18-19 ans	0,165	0,166	0,166	0,166	0,166	0,166
20-24 ans	0,265	0,266	0,266	0,267	0,267	0,266
25 ans et plus	0,245	0,246	0,247	0,247	0,247	0,247
Compsante de la variance	.....	0,143	0,225	0,225	0,225	0,184

4.6. Interprétation des résultats

Notre analyse est effectuée selon l'approche spécifique aux sujets puisque nous utilisons le modèle logistique-normal. Nous supposons donc que la probabilité d'avoir au moins un accident corporel ou mortel varie parmi les sujets, reflétant ainsi leur prédisposition naturelle à avoir ce type d'accident et l'influence non mesurable des facteurs environnementaux. L'effet de l'appartenance à un sous-groupe de la population ou l'effet de l'expérience accumulée (période d'observation des accidents) est donc semblable pour chacun des conducteurs. La composante de la variance représente le degré d'hétérogénéité parmi ces conducteurs.

Les tableaux 4.6.1 et 4.6.2 donnent les coefficients estimés à l'aide de l'approche QQN, les rapports de cotes individuels (exponentiel des coefficients estimés) et les intervalles de confiance à 95% correspondants. Les intervalles de confiance sont obtenus à l'aide de la distribution asymptotique des estimateurs des coefficients (distribution normale) et de la variance approximative des estimateurs calculée par des différences finies (voir section 2.4.2). Notons que par définition, le rapport de cotes du groupe de référence est égal à 1. De plus, le rapport de cotes est significativement différent de 1 au niveau 5% lorsque la valeur 1 n'est pas incluse dans l'intervalle de confiance à 95%. Tel que mentionné à la section 1.4, le rapport de cotes est approximativement égal au risque relatif lorsque la proportion de succès (avoir au moins un accident avec victimes) est proche de zéro pour les deux groupes.

4.6.1. Interprétation des résultats pour les 53 069 nouveaux conducteurs

Au tableau 4.6.1, le coefficient estimé correspondant à la période d'observation des accidents est significatif (au niveau 5%) et négatif montrant ainsi que le risque approximatif d'au moins un accident corporel ou mortel diminue lorsque l'expérience augmente. Ceci confirme la tendance observée graphiquement à la section 4.4. De plus, la variance estimée de la composante aléatoire est significative au niveau 5% : il est donc nécessaire de tenir compte de la corrélation dans le temps.

En examinant les résultats, nous constatons que les hommes du fichier des 53 069 conducteurs qui sont plus lents à obtenir leur permis sont significativement (au niveau 5%) moins à risque individuellement que ceux qui ont entre 90 et 180 jours d'apprentissage. De plus, nous constatons qu'un conducteur de 16 à 19 ans à l'obtention du permis qui aurait besoin de plus d'une tentative pour réussir l'ensemble des examens (pratique et théorique) est significativement plus à risque au niveau 5% que le même conducteur qui appartiendrait à la catégorie des 16 ans ayant réussi les deux examens dès le premier essai. Enfin, les 25 ans et plus qui ont réussi les deux examens dès la première tentative ont des risques individuels d'accidents corporels ou mortels approximativement 33% moins élevés que les 16 ans à l'obtention du permis ayant réussi l'ensemble des examens en un seul essai.

TAB. 4.6.1. *Modèle logistique-normal pour la probabilité d'au moins un accident avec victimes par année par nouveau conducteur (fichier des 53 069 conducteurs).*

Variables explicatives	Coefficient	Rapport de cotes	IC à 95%	
			Limite inférieure	Limite supérieure
Constante	-3,712			
Période d'observation des accidents	**-0,062**	**0,940**	**0,906**	**0,974**
Date d'entrée dans le processus d'accès				
01/03/91 au 29/02/92	0,093	1,097	1,008	1,194
01/03/92 au 29/02/93	------	1,000	Groupe de référence	
Année d'obtention du permis				
1991	------	1,000	Groupe de référence	
1992	0,004	1,004	0,910	1,108
1993	0,050	1,051	0,922	1,199
Durée d'apprentissage				
90 à 180 jours	------	1,000	Groupe de référence	
181 à 270 jours	**-0,126**	**0,882**	**0,801**	**0,971**
Plus de 270 jours	**-0,316**	**0,729**	**0,648**	**0,820**
Nombre de tentatives pour réussir l'examen théorique et pratique				
Une tentative				
16 ans	------	1,000	Groupe de référence	
17 ans	0,003	1,003	0,898	1,121
18-19 ans	-0,006	0,994	0,865	1,144
20-24 ans	-0,160	0,852	0,684	1,062
25 ans et plus	**-0,400**	**0,670**	**0,530**	**0,847**
Plus d'une tentative				
16 ans	**0,213**	**1,237**	**1,135**	**1,349**
17 ans	**0,326**	**1,385**	**1,250**	**1,535**
18-19 ans	**0,290**	**1,336**	**1,179**	**1,514**
20-24 ans	0,022	1,022	0,840	1,243
25 ans et plus	-0,057	0,944	0,796	1,120
Log de la vraisemblance	-20775,50			
Paramètre de la variance estimée	0,467 ($p<0,01$)			

Note : les résultats en gras sont significatifs au seuil 5%

4.6.2. Interprétation des résultats pour les 3 550 nouveaux conducteurs ayant plus de 90 jours d'interruption totale

Au tableau 4.6.2, le coefficient correspondant à la période d'observation des accidents est négatif et très significatif (au niveau 5%). Nous en concluons que plus l'expérience augmente, plus la probabilité d'accidents corporels ou mortels diminue. Ces coefficients estimés confirme une fois de plus la tendance observée au graphique 4.4.2. L'estimateur de la variance de l'effet aléatoire est non significatif au seuil 5%. Il n'y aurait donc pas d'hétérogénéité naturelle entre les conducteurs de même sous-groupe ayant plus de 90 jours dínterruption totale.

En examinant rapports de cotes du tableau 4.6.2, nous remarquons que les conducteurs qui sont très lents à obtenir leur permis probatoire (plus de 270 jours d'apprentissage) sont significativement moins à risque au niveau 5% que ceux qui détiennent leur permis d'apprenti entre 90 et 180 jours. Nous constatons également que les 20-24 ans à l'obtention du permis ayant réussi les deux examens en une tentative ont des risques individuels d'au moins un accident corporel ou mortel approximativement 57% fois moins élevés que les conducteurs appartenant à la catégorie de référence. De plus, un homme qui aurait entre 20 et 24 ans à l'obtention du permis et qui aurait réussi les deux examens en plus d'une tentative est 52% fois moins à risque que le même homme appartenant à la catégorie des 16 ans qui aurait réussi les deux examens en une seule tentative.

En général, les résultats pour les 3 550 nouveaux conducteurs sont différents comparativement aux 53 069 hommes du fichier précédent. Par contre, les groupes de comparaisons des tableaux 4.6.1 et 4.6.2 ne sont pas les mêmes et les taux d'accidents avec victimes diffèrent d'un facteur 2 entre les deux fichiers (voir tableau 4.4.1). Nous ne pouvons donc pas comparer directement les deux analyses même si les variables agissent de façon analogue.

TAB. 4.6.2. *Modèle logistique-normal pour la probabilité d'au moins un accident avec victimes par année par nouveau conducteur (fichier des 3 550 conducteurs).*

Variables explicatives	Coefficient	Rapport de cotes	IC à 95%	
			Limite inférieure	Limite supérieure
Constante	-1,754			
Période d'observation des accidents	**-0,436**	**0,646**	**0,581**	**0,719**
Date d'entrée dans le processus d'accès				
01/03/91 au 29/02/92	0,008	1,008	0,802	1,266
01/03/92 au 29/02/93	------	1,000	Groupe de référence	
Année d'obtention du permis				
1991	------	1,000	Groupe de référence	
1992	-0,210	0,811	0,582	1,129
1993	-0,107	0,899	0,604	1,337
Durée d'apprentissage				
90 à 180 jours	------	1,000	Groupe de référence	
181 à 270 jours	-0,188	0,829	0,648	1,060
Plus de 270 jours	**-0,327**	**0,721**	**0,551**	**0,945**
Nombre de tentatives pour réussir l'examen théorique et pratique				
Une tentative				
16 ans	------	1,000	Groupe de référence	
17 ans	0,047	1,048	0,764	1,438
18-19 ans	-0,038	0,962	0,678	1,365
20-24 ans	**-0,855**	**0,425**	**0,239**	**0,757**
25 ans et plus	-0,571	0,565	0,317	1,007
Plus d'une tentative				
16 ans	0,190	1,209	0,901	1,622
17 ans	0,126	1,135	0,847	1,520
18-19 ans	-0,136	0,873	0,630	1,208
20-24 ans	**-0,740**	**0,477**	**0,283**	**0,804**
25 ans et plus	-0,302	0,739	0,456	1,199
Log de la vraisemblance	-2227,00			
Paramètre de la variance estimée	0,053 (p>0,1)			

Note : les résultats en gras sont significatifs au seuil 5%

CONCLUSION

Ce mémoire avait pour but l'étude des différentes extensions du modèle de régression logistique pour analyser des données dichotomiques longitudinales. Les différentes approches proposées dans la littérature ont été réparties en deux catégories : le modèle marginal et le modèle spécifique aux sujets. En sécurité routière, les banques de données contiennent généralement plusieurs réponses dans le temps pour un même individu. Les chercheurs du Laboratoire sur la sécurité des transports du CRT étaient donc intéressés à connaître les nouveaux développements de la méthodologie se rapportant à l'analyse de telles données. Ils souhaitaient plus particulièrement une recommandation générale quant à l'utilisation des différentes méthodes de l'approche spécifique aux sujets.

Nous avons tout d'abord effectué un survol des approches proposées dans la littérature. Essentiellement, l'approche marginale consiste à modéliser l'espérance marginale de la réponse en considérant la corrélation comme un paramètre de nuisance. L'approche spécifique aux sujets, quant à elle, consiste plutôt à modéliser les réponses individuelles en supposant que la corrélation des observations pour un même sujet vient d'une variable aléatoire normale commune. Une première distinction entre les deux approches concerne donc les hypothèses de dépendance dans le temps. Nous avons également vu que les deux approches s'interprètent différemment. En effet, les coefficients du modèle marginal décrivent l'effet des changements des variables explicatives sur la réponse moyenne alors que les coefficients du modèle spécifique aux sujets décrivent l'effet pour un seul individu.

Nous avons présenté un exemple concret d'interprétation dans le cadre d'une étude clinique sur un médicament afin d'illustrer ces deux approches. Enfin, nous avons terminé le premier chapitre en indiquant que les coefficients du modèle spécifique aux sujets avec un seul effet aléatoire sont toujours plus grands ou égaux aux coefficients du modèle de régression marginale.

En sécurité routière, l'objectif est souvent de faire de l'inférence sur les individus plutôt que sur la moyenne. Nous nous sommes donc intéressés plus particulièrement au modèle logistique-normal, en décrivant la théorie à la base de différentes méthodes permettant de maximiser la vraisemblance. La méthode PQL, proposée par Breslow et Clayton (1993), consiste à faire une approximation des équations scores à l'aide d'équations d'estimation généralisées pouvant être résolues itérativement. Comme cette méthode mène à des résultats biaisés pour des réponses dichotomiques avec peu de répétitions dans le temps, nous avons décrit en détail la correction proposée par Lin et Breslow (1996). Leur idée est essentiellement d'utiliser les expressions asymptotiques dérivées par Solomon et Cox (1992). Finalement, nous avons étudié une méthode sans biais basée sur la quadrature gaussienne. Suivant le raisonnement de Pinheiro et Bates (1995), nous avons privilégié l'approche QGA qui peut être ajustée à l'aide de la nouvelle procédure PROC NLMIXED de la version 8 de SAS.

Afin de comparer les méthodes PQL, PQL corrigée et QGA, nous avons effectué des simulations selon trois modèles comportant une seule variable aléatoire. Nous avons constaté que les écarts-types estimés pour la variance de l'effet aléatoire sont considérablement plus élevés que les écarts-types simulés lorsqu'il y a trois répétitions dans le temps pour la méthode PQL corrigée. De plus, nos résultats ont montré l'importance du nombre d'observations dans le temps lorsque l'approche spécifique aux sujets est choisie. Nous recommandons l'approche QGA

à l'aide de l'algorithme d'optimisation de quasi-Newton lorsque le nombre d'individus est très grand. Par contre, l'estimateur de la variance de l'effet aléatoire peut être loin de la vraie valeur pour de petits échantillons surtout lorsqu'il y a peu de répétitions dans le temps. Une alternative est alors l'utilisation des estimateurs PQL corrigés.

Dans le dernier chapitre, nous avons appliqué les méthodes spécifiques aux sujets à des données de sécurité routière. Notre but était de modéliser les risques d'accidents avec victimes de nouveaux conducteurs appartenant à deux banques de données distinctes. Il s'agissait d'une étude longitudinale, puisque les réponses (présence ou non d'au moins un accident avec victimes) étaient observées de façon répétée dans le temps. Nous avons présenté un modèle logistique-normal avec interaction pour les 53 069 hommes de la banque de données analysée par Laberge-Nadeau *et al.* (1999) ainsi que pour les 3 550 conducteurs ayant eu des sanctions ou des retards de paiement de plus de 90 jours. Pour la première banque de données, l'approche QGA avec l'algorithme d'optimisation de quasi-Newton a été choisie, car le nombre d'individus était très élevé. Mentionnons que la composante aléatoire du modèle n'étant pas significative pour les conducteurs ayant plus de 90 jours, les différentes approches donnaient des résultats comparables.

Les modèles de régression pour les simulations et pour l'exemple d'application en sécurité routière ne comportaient qu'un seul effet aléatoire. En pratique, il arrive que le modèle ait plusieurs composantes aléatoires. Tel que mentionné au chapitre 1, l'approche QGA est adéquate pour les modèles comportant moins de 3 effets aléatoires, alors que la méthode PQL corrigée est conseillée pour les problèmes de dimensions élevées. Notons toutefois que la procédure PROC NLMIXED de SAS (méthode QGA) ne permet que la modélisation avec un effet aléatoire. Il serait donc intéressant d'améliorer cette procédure.

L'extension des modèles linéaires généralisés à l'analyse de données longitudinales pave la voie à plusieurs autres sujets de recherche. Ainsi, la procédure de Wolfinger et O'Connell (1993), raffinement de l'approche PQL parce qu'elle suppose que le paramètre d'échelle est inconnu, mérite d'être étudiée. Nous avons d'ailleurs constaté que les coefficients estimés pour l'analyse des nouveaux conducteurs étaient différents lorsque nous appliquions cette procédure. En terminant, il arrive régulièrement en sécurité routière que les études portent sur le nombre d'accidents. Il serait donc intéressant d'examiner les différentes méthodes de l'approche spécifique aux sujets, mais cette fois pour des données de comptage.

Annexe A

DISTRIBUTION DES PARAMÈTRES ESTIMÉS POUR LE MODÈLE 3.1.1 LORSQUE LA VRAIE VALEUR DE $\theta = 1$

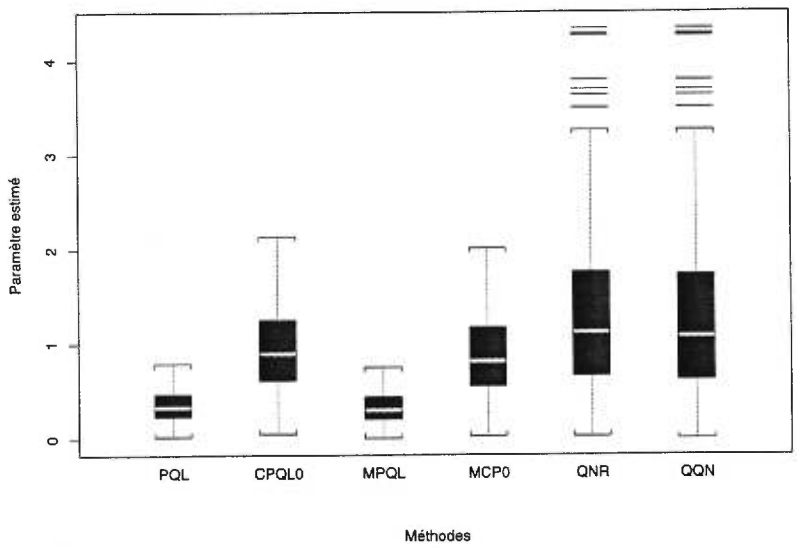

FIG. A.0.1. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 250$ individus (vraie valeur = 1).*

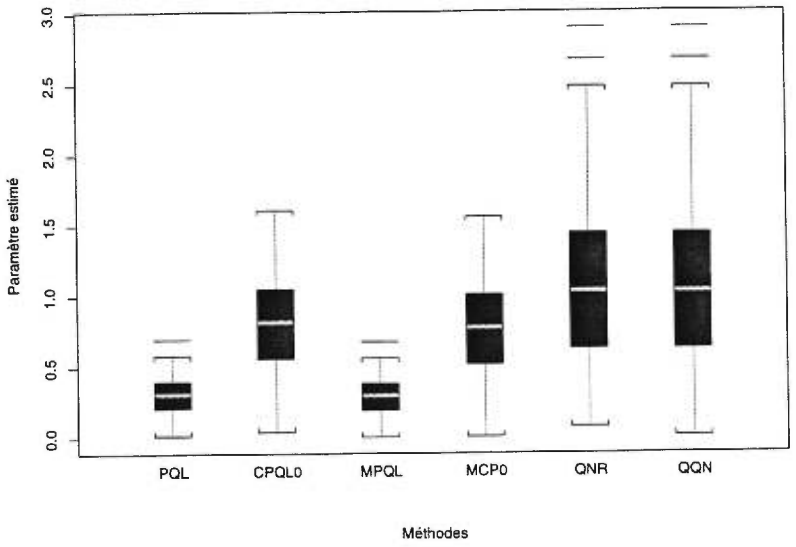

FIG. A.0.2. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 500$ individus (vraie valeur = 1).*

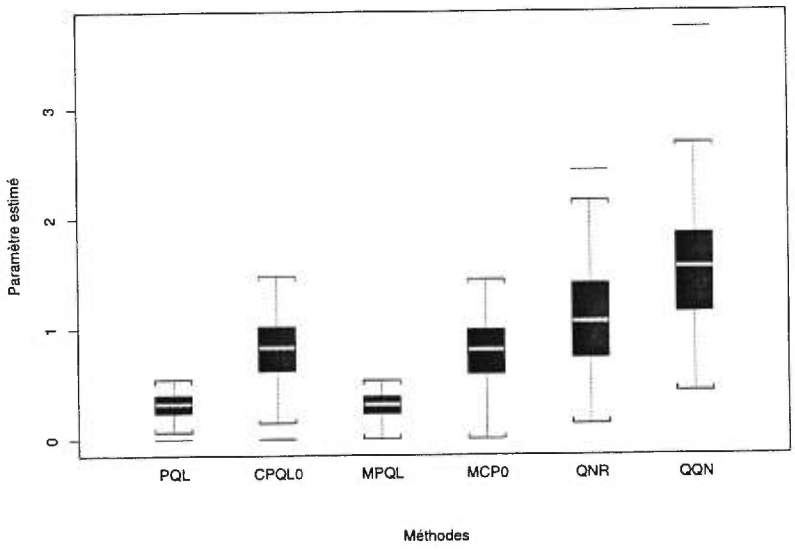

FIG. A.0.3. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 750$ individus (vraie valeur = 1).*

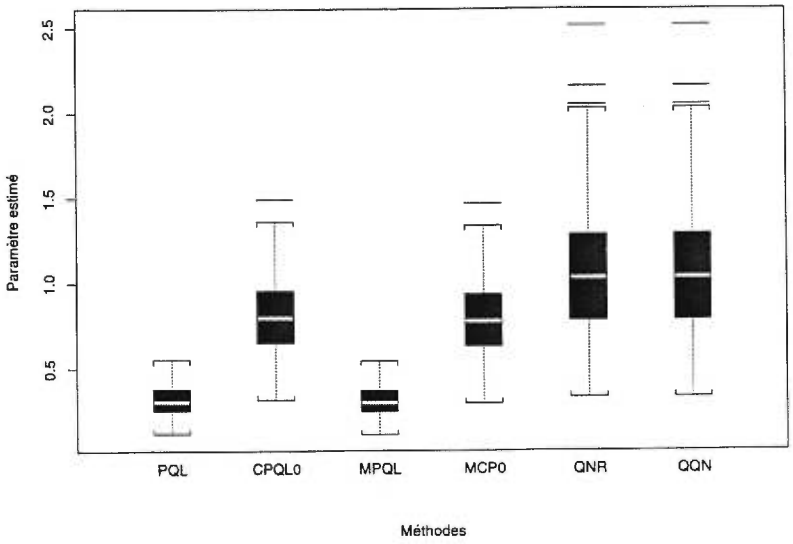

FIG. A.0.4. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 1000$ individus (vraie valeur = 1).*

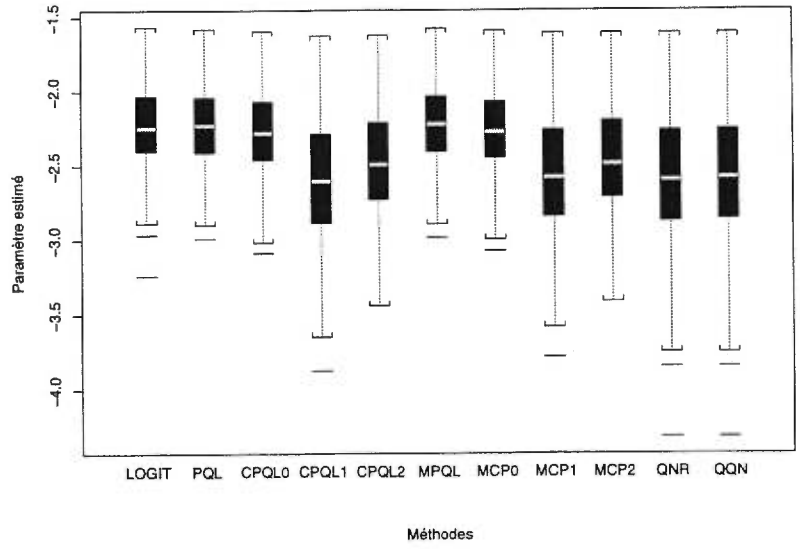

FIG. A.0.5. *Graphiques en boîtes des estimateurs de β_0 pour le modèle 3.1.1 avec $m = 250$ individus (vraie valeur = -2,5).*

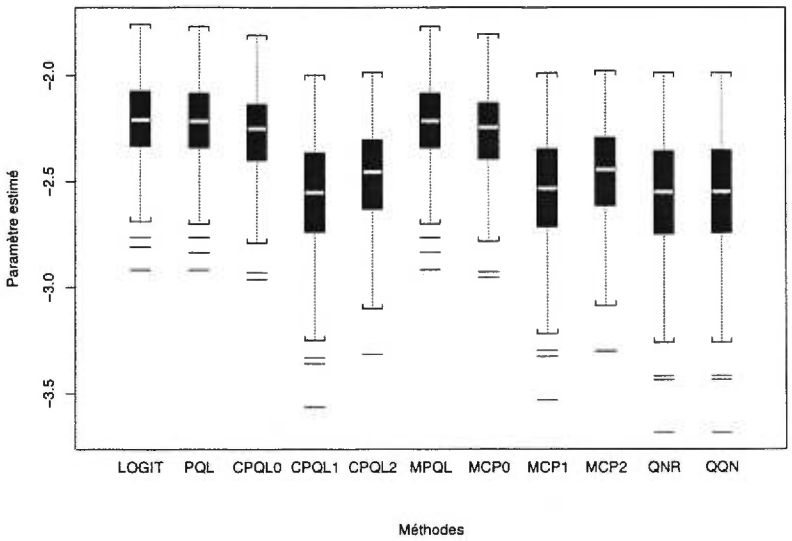

FIG. A.0.6. *Graphiques en boîtes des estimateurs de β_0 pour le modèle 3.1.1 avec $m = 500$ individus (vraie valeur = -2,5).*

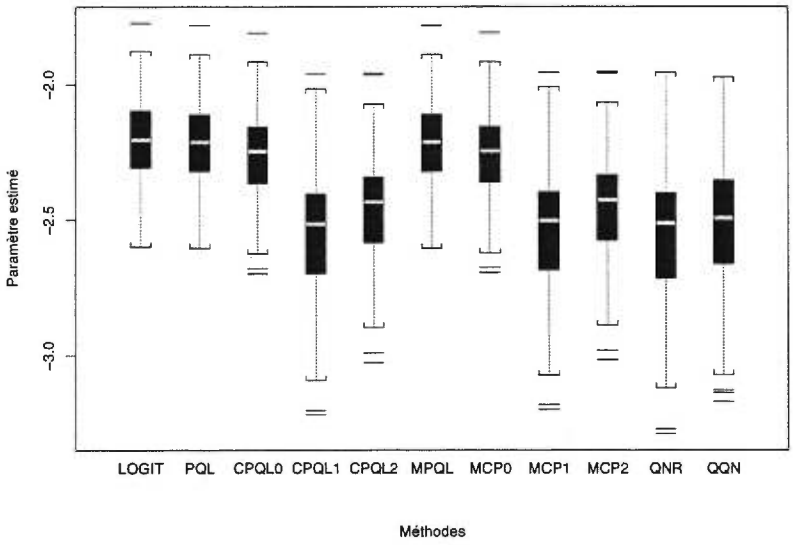

FIG. A.0.7. *Graphiques en boîtes des estimateurs de β_0 pour le modèle 3.1.1 avec $m = 750$ individus (vraie valeur = -2,5).*

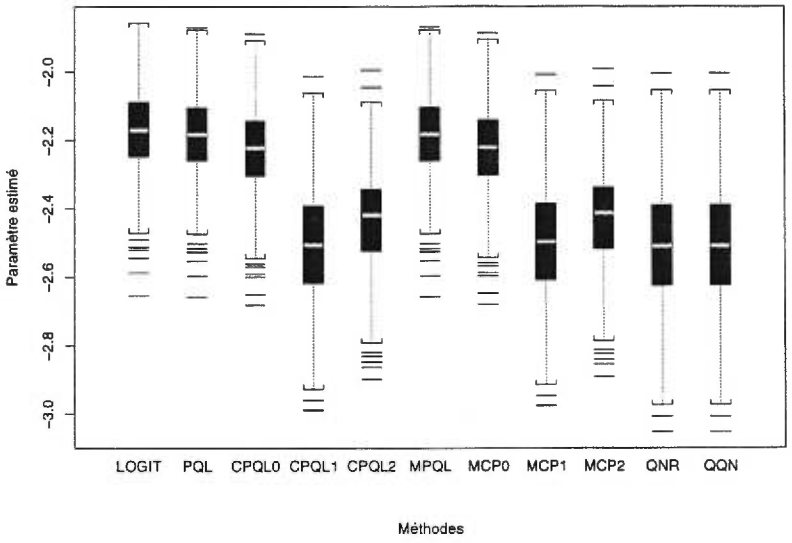

FIG. A.0.8. *Graphiques en boîtes des estimateurs de β_0 pour le modèle 3.1.1 avec $m = 1000$ individus (vraie valeur = -2,5).*

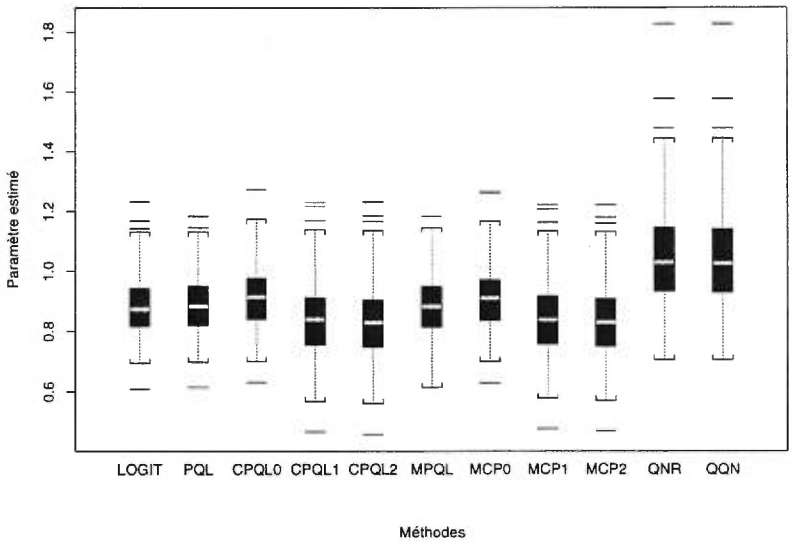

FIG. A.0.9. *Graphiques en boîtes des estimateurs de β_1 pour le modèle 3.1.1 avec $m = 250$ individus (vraie valeur = 1).*

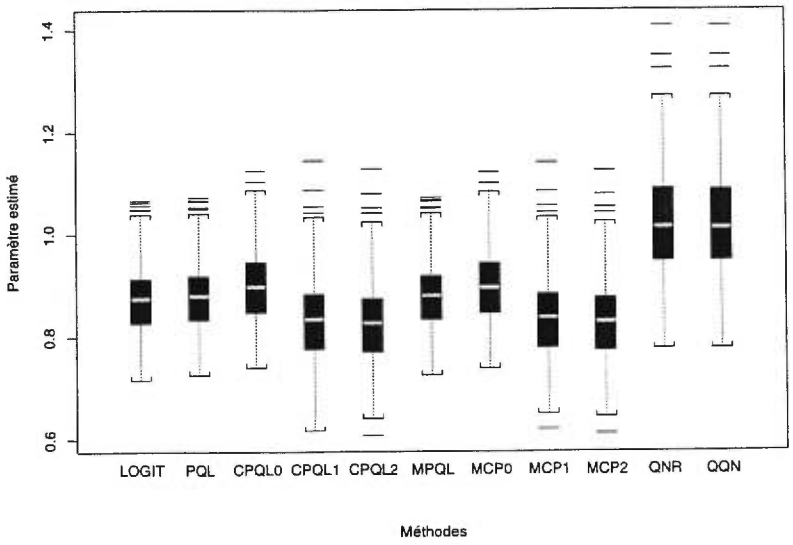

FIG. A.0.10. *Graphiques en boîtes des estimateurs de β_1 pour le modèle 3.1.1 avec $m = 500$ individus (vraie valeur = 1).*

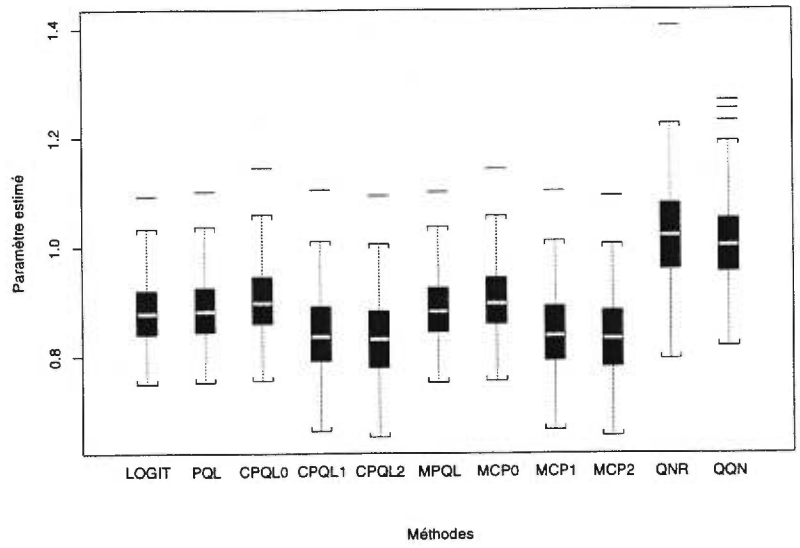

FIG. A.0.11. *Graphiques en boîtes des estimateurs de β_1 pour le modèle 3.1.1 avec $m = 750$ individus (vraie valeur = 1)*

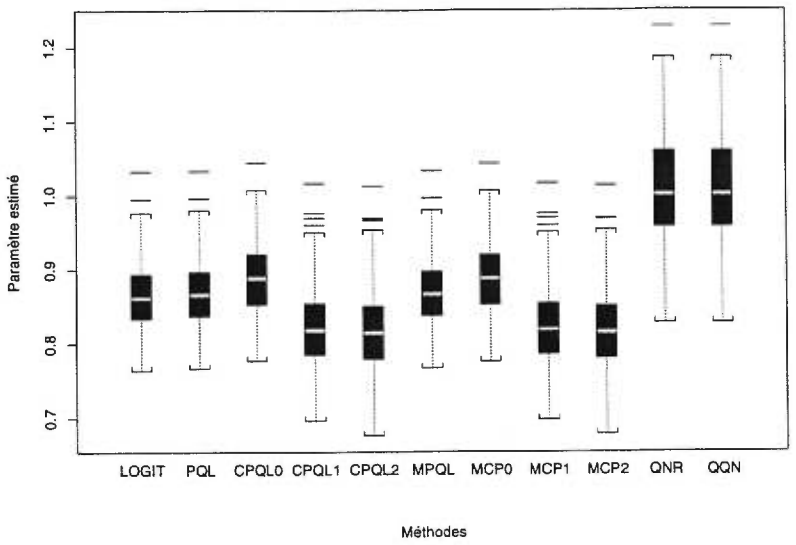

FIG. A.0.12. *Graphiques en boîtes des estimateurs de β_1 pour le modèle 3.1.1 avec $m = 1000$ individus (vraie valeur = 1).*

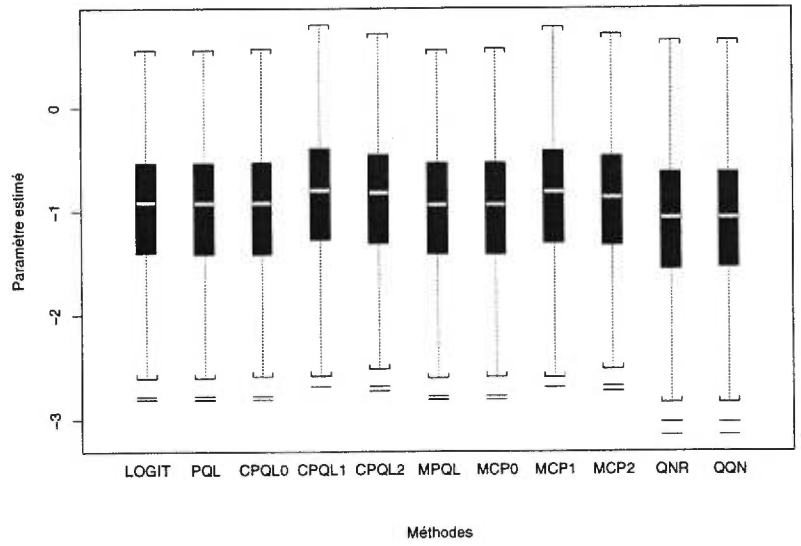

FIG. A.0.13. *Graphiques en boîtes des estimateurs de β_2 pour le modèle 3.1.1 avec $m = 250$ individus (vraie valeur = -1).*

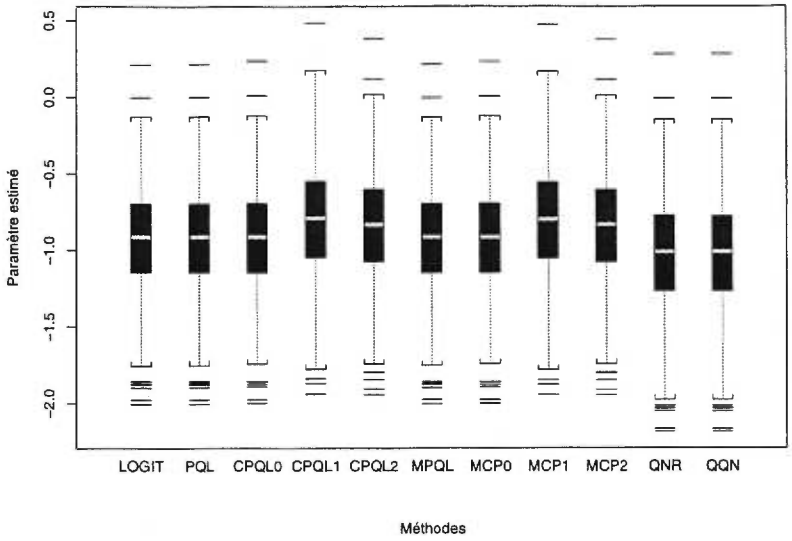

FIG. A.0.14. *Graphiques en boîtes des estimateurs de β_2 pour le modèle 3.1.1 avec $m = 500$ individus (vraie valeur = -1)*

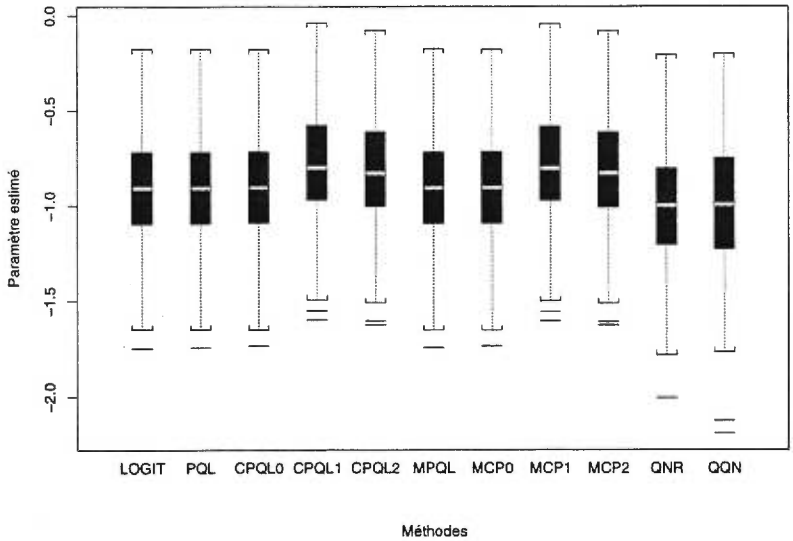

FIG. A.0.15. *Graphiques en boîtes des estimateurs de β_2 pour le modèle 3.1.1 avec $m = 750$ individus (vraie valeur = -1).*

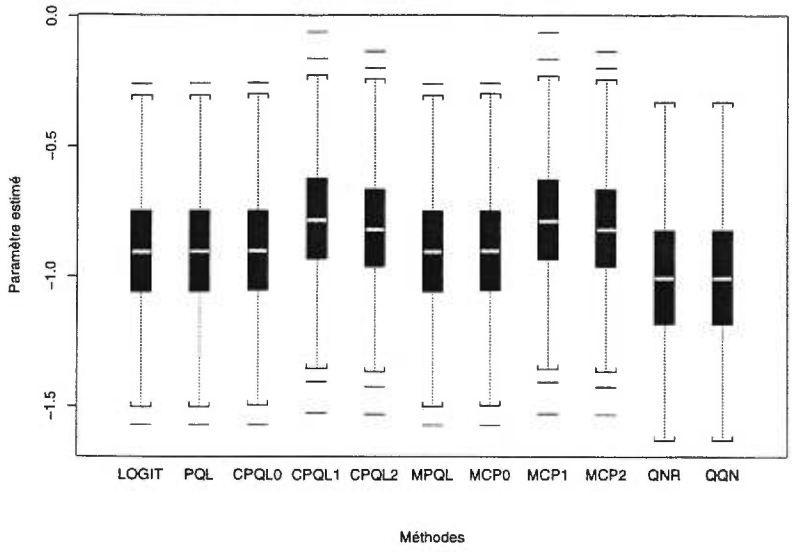

FIG. A.0.16. *Graphiques en boîtes des estimateurs de β_2 pour le modèle 3.1.1 avec $m = 1000$ individus (vraie valeur = -1).*

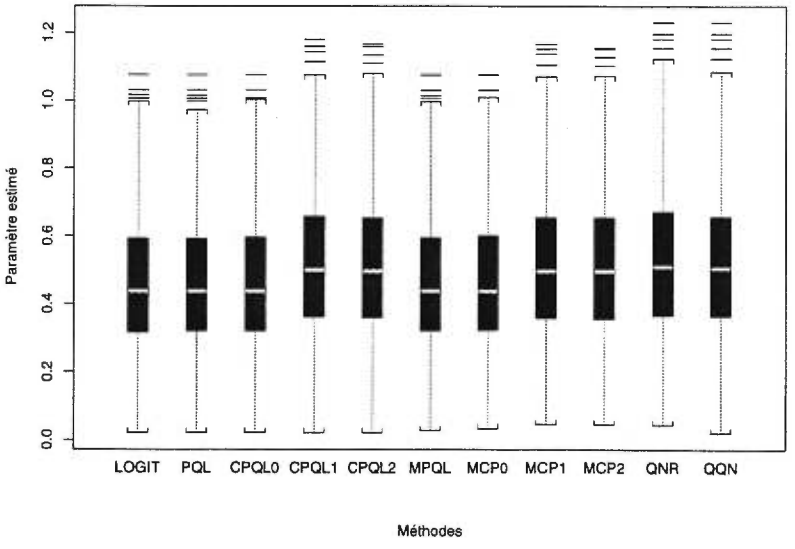

FIG. A.0.17. *Graphiques en boîtes des estimateurs de β_3 pour le modèle 3.1.1 avec $m = 250$ individus (vraie valeur = 0,5).*

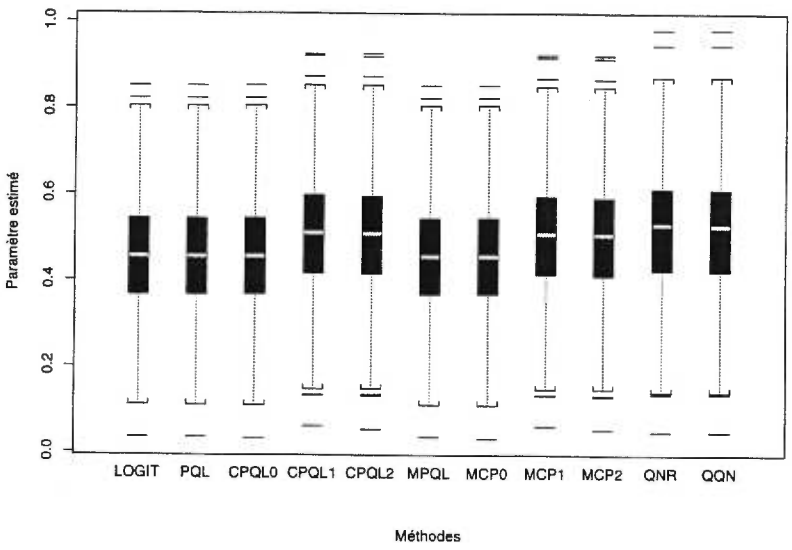

FIG. A.0.18. *Graphiques en boîtes des estimateurs de β_3 pour le modèle 3.1.1 avec $m = 500$ individus (vraie valeur = 0,5).*

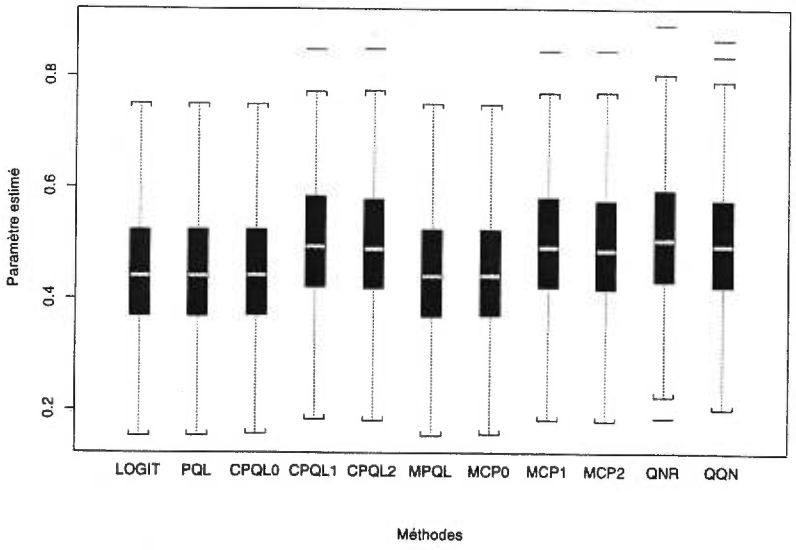

FIG. A.0.19. *Graphiques en boîtes des estimateurs de β_3 pour le modèle 3.1.1 avec $m = 750$ individus (vraie valeur = 0,5).*

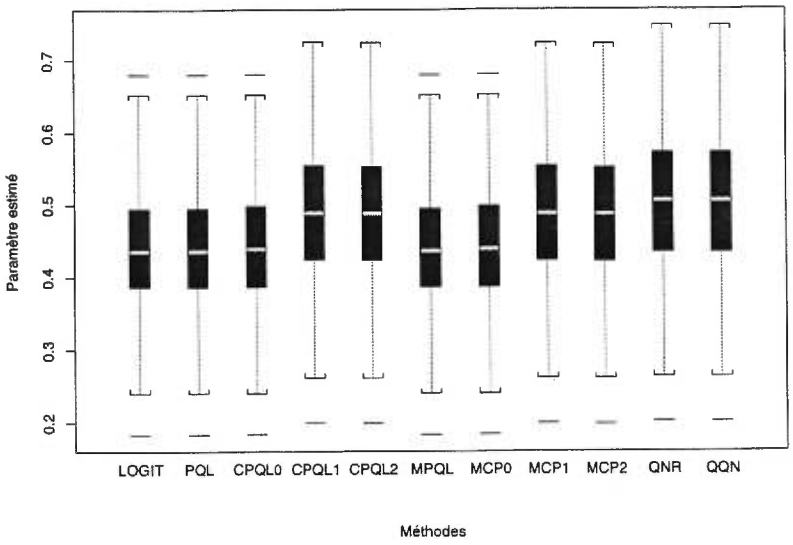

FIG. A.0.20. *Graphiques en boîtes des estimateurs de β_3 pour le modèle 3.1.1 avec $m = 1000$ individus (vraie valeur = 0,5).*

Annexe B

DISTRIBUTION DES PARAMÈTRES ESTIMÉS POUR LE MODÈLE 3.1.1 LORSQUE LA VRAIE VALEUR DE $\theta = 1{,}5$ OU $0{,}5$

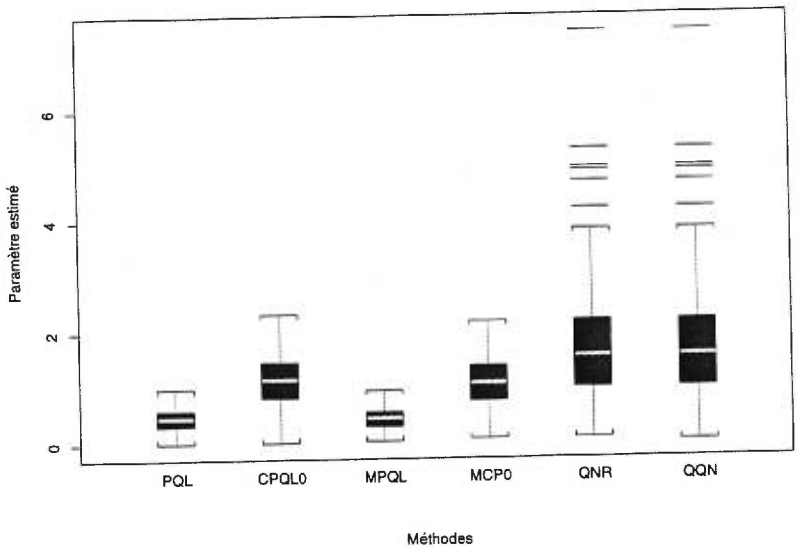

FIG. B.0.1. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 250$ individus (vraie valeur = 1,5).*

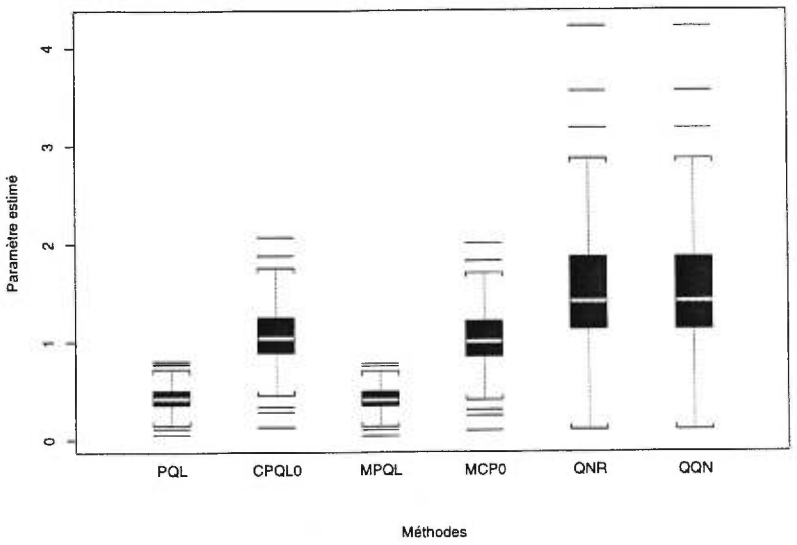

FIG. B.0.2. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 500$ individus (vraie valeur = 1,5).*

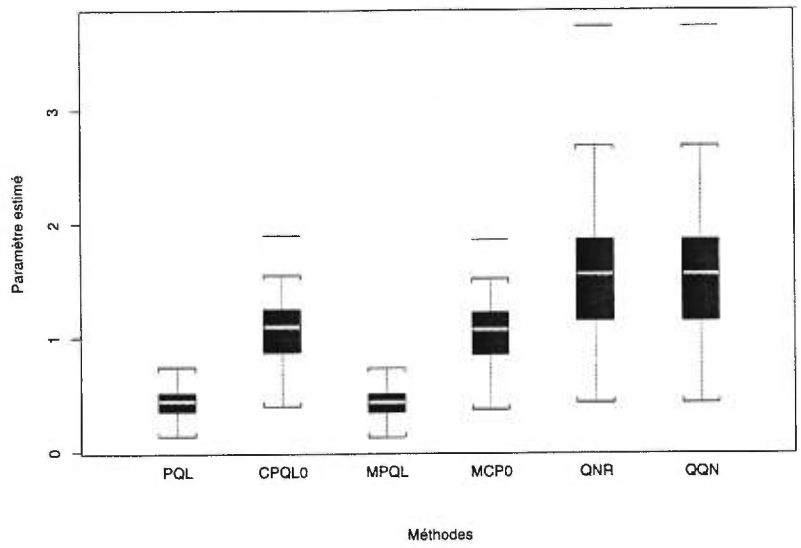

FIG. B.0.3. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 750$ individus (vraie valeur = 1,5).*

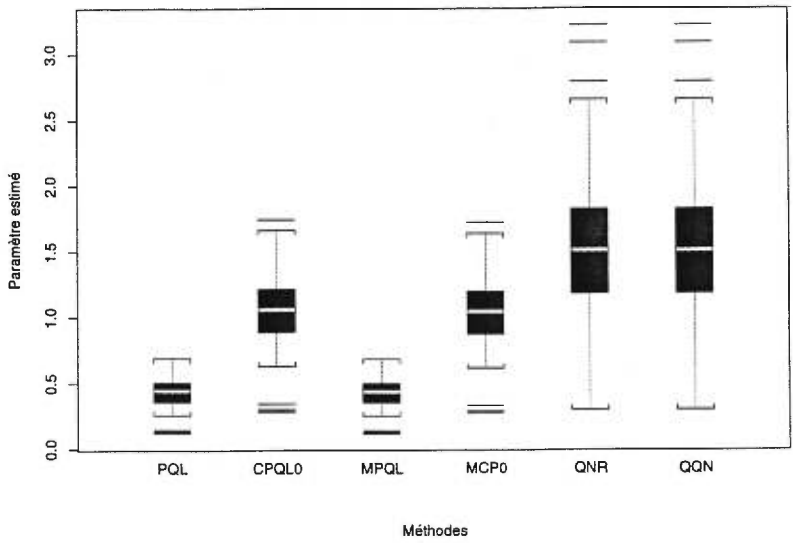

FIG. B.0.4. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 1000$ individus (vraie valeur = 1,5).*

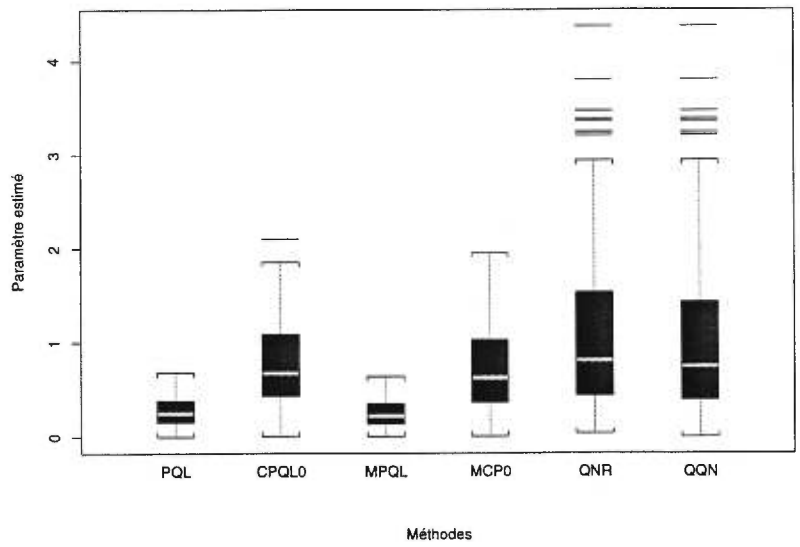

FIG. B.0.5. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 250$ individus (vraie valeur = 0,5).*

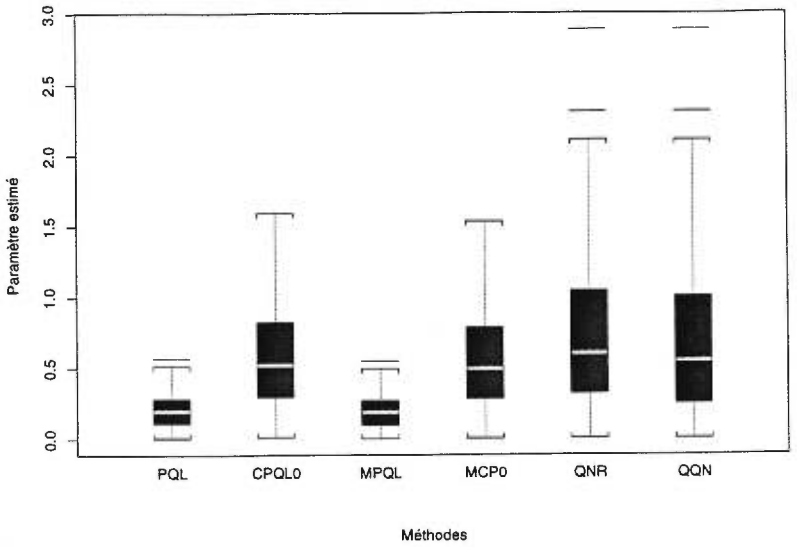

FIG. B.0.6. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 500$ individus (vraie valeur = 0,5).*

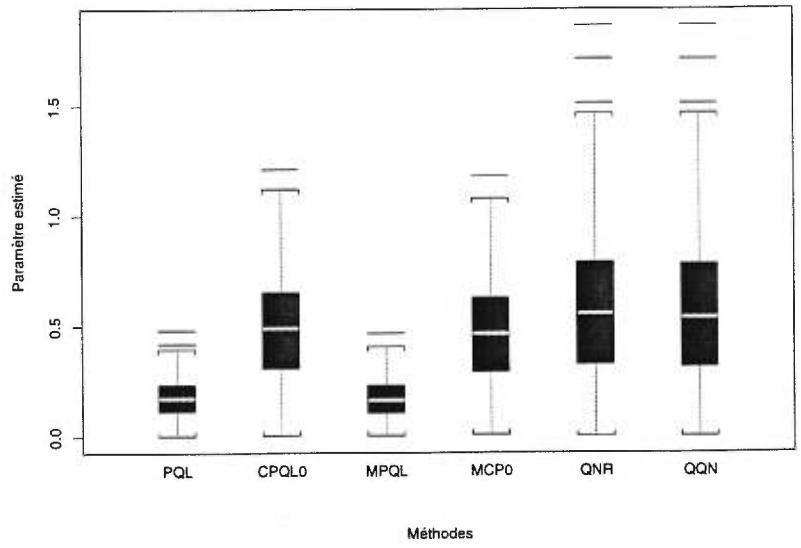

FIG. B.0.7. *Graphiques en boîtes des estimateurs de θ pour $m = 750$ individus (vraie valeur = 0,5).*

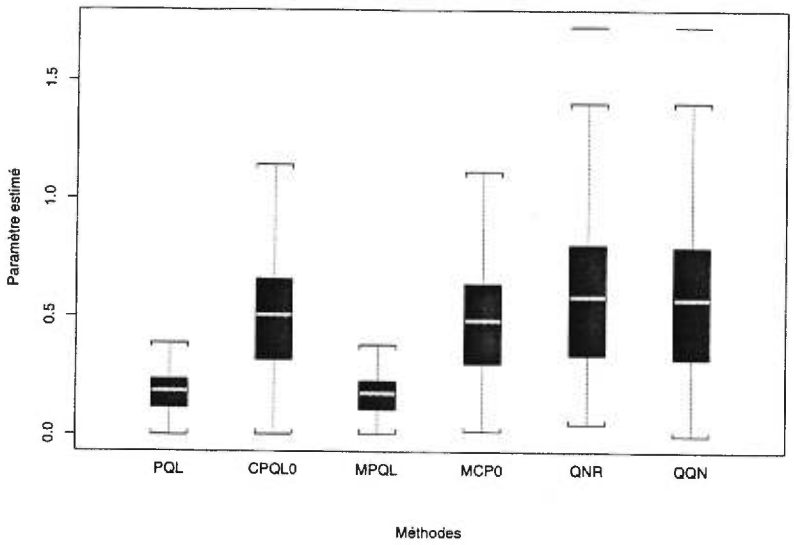

FIG. B.0.8. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.1 avec $m = 1000$ individus (vraie valeur = 0,5).*

Annexe C

MOYENNES ET ÉCARTS TYPES DES PARAMÈTRES POUR LE MODÈLE 3.1.1 LORSQUE $\theta = 1{,}5$ OU 0,5

TAB. C.0.1. *Valeurs moyennes des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour* $m = 250, 500$ *et une vraie valeur de* $\theta = 1,5$.

m	Méthode	θ	β_0	β_1	β_2	β_3
Vraie valeur		1,50	-2,50	1,00	-1,00	0,50
250	LOGIT	—	-2,09	0,84	-0,89	0,42
	PQL	0,46	-2,12	0,84	-0,89	0,43
	CPQL0	1,12	-2,19	0,88	-0,89	0,43
	CPQL1	1,12	-2,58	0,79	-0,72	0,50
	CPQL2	1,12	-2,40	0,77	-0,80	0,49
	MPQL	0,43	-2,11	0,84	-0,89	0,43
	MCP0	1,05	-2,18	0,87	-0,89	0,43
	MCP1	1,05	-2,54	0,79	-0,73	0,49
	MCP2	1,05	-2,39	0,78	-0,80	0,49
	QNR	1,70	-2,59	1,04	-1,04	0,52
	QQN	1,69	-2,59	1,04	-1,04	0,52
500	LOGIT	—	-2,08	0,82	-0,87	0,42
	PQL	0,43	-2,10	0,83	-0,87	0,42
	CPQL0	1,05	-2,16	0,86	-0,87	0,43
	CPQL1	1,05	-2,53	0,77	-0,71	0,49
	CPQL2	1,05	-2,39	0,76	-0,77	0,49
	MPQL	0,41	-2,10	0,83	-0,87	0,42
	MCP0	1,01	-2,16	0,86	-0,87	0,43
	MCP1	1,01	-2,51	0,77	-0,72	0,49
	MCP2	1,01	-2,38	0,76	-0,77	0,48
	QNR	1,50	-2,53	1,00	-1,01	0,51
	QQN	1,50	-2,53	1,00	-1,01	0,51

TAB. C.0.2. *Valeurs moyennes des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour* $m = 750, 1000$ *et une vraie valeur de* $\theta = 1,5$.

m	Méthode	θ	β_0	β_1	β_2	β_3
Vraie valeur		1,50	-2,50	1,00	-1,00	0,50
750	LOGIT	—	-2,06	0,82	-0,85	0,42
	PQL	0,44	-2,08	0,83	-0,85	0,42
	CPQL0	1,06	-2,14	0,86	-0,85	0,42
	CPQL1	1,06	-2,50	0,77	-0,70	0,48
	CPQL2	1,06	-2,36	0,76	-0,76	0,48
	MPQL	0,42	-2,08	0,83	-0,85	0,42
	MCP0	1,03	-2,14	0,85	-0,85	0,42
	MCP1	1,03	-2,49	0,77	-0,70	0,48
	MCP2	1,03	-2,36	0,76	-0,76	0,48
	QNR	1,53	-2,51	1,01	-1,00	0,50
	QQN	1,53	-2,51	1,01	-1,00	0,50
1000	LOGIT	—	-2,06	0,82	-0,87	0,42
	PQL	0,43	-2,08	0,83	-0,87	0,42
	CPQL0	1,05	-2,15	0,86	-0,87	0,42
	CPQL1	1,05	-2,51	0,77	-0,71	0,49
	CPQL2	1,05	-2,37	0,76	-0,77	0,48
	MPQL	0,42	-2,08	0,83	-0,87	0,42
	MCP0	1,03	-2,14	0,86	-0,87	0,42
	MCP1	1,03	-2,50	0,77	-0,72	0,49
	MCP2	1,03	-2,37	0,76	-0,77	0,48
	QNR	1,51	-2,52	1,01	-1,02	0,51
	QQN	1,51	-2,52	1,01	-1,01	0,51

TAB. C.0.3. *Comparaison des écarts types estimés et simulés du modèle 3.1.1 avec 200 répétitions pour $m = 250, 500$ et une vraie valeur de $\theta = 1,5$.*

m	Méthode		θ	β_0	β_1	β_2	β_3
250	LOGIT	Sim. [a]	—	0,24	0,09	0,50	0,18
		Est. [b]	—	0,25	0,10	0,48	0,18
	PQL	Sim.	0,20	0,24	0,09	0,50	0,18
		Est.	0,34	0,25	0,10	0,49	0,18
	CPQL0	Sim.	0,48	0,25	0,09	0,50	0,18
		Est.	0,85	0,27	0,10	0,51	0,18
	CPQL1	Sim.	0,48	0,35	0,12	0,52	0,21
		Est.	0,85	0,40	0,12	0,52	0,19
	CPQL2	Sim.	0,48	0,28	0,13	0,51	0,21
		Est.	0,85	0,40	0,12	0,52	0,19
	MPQL	Sim.	0,19	0,24	0,09	0,51	0,18
		Est.	0,33	0,25	0,10	0,49	0,18
	MCP0	Sim.	0,45	0,25	0,09	0,51	0,18
		Est.	0,83	0,27	0,09	0,50	0,18
	MCP1	Sim.	0,45	0,34	0,12	0,52	0,21
		Est.	0,83	0,39	0,12	0,52	0,19
	MCP2	Sim.	0,45	0,28	0,12	0,52	0,21
		Est.	0,83	0,39	0,12	0,52	0,19
	QNR	Sim.	1,05	0,40	0,16	0,58	0,22
		Est.	1,04	0,41	0,16	0,56	0,21
	QQN	Sim.	1,07	0,40	0,16	0,58	0,22
		Est.	1,05	0,40	0,16	0,56	0,21
500	LOGIT	Sim.	—	0,18	0,06	0,30	0,11
		Est.	—	0,17	0,07	0,33	0,12
	PQL	Sim.	0,12	0,18	0,06	0,30	0,11
		Est.	0,24	0,18	0,07	0,33	0,12
	CPQL0	Sim.	0,30	0,19	0,06	0,31	0,11
		Est.	0,58	0,19	0,07	0,34	0,12
	CPQL1	Sim.	0,30	0,26	0,08	0,31	0,12
		Est.	0,58	0,27	0,09	0,35	0,13
	CPQL2	Sim.	0,30	0,21	0,08	0,31	0,12
		Est.	0,58	0,27	0,09	0,35	0,13
	MPQL	Sim.	0,12	0,18	0,06	0,30	0,11
		Est.	0,23	0,18	0,07	0,33	0,12
	MCP0	Sim.	0,29	0,19	0,06	0,31	0,11
		Est.	0,58	0,18	0,07	0,34	0,12
	MCP1	Sim.	0,29	0,25	0,08	0,31	0,12
		Est.	0,58	0,27	0,08	0,35	0,13
	MCP2	Sim.	0,29	0,21	0,08	0,31	0,12
		Est.	0,58	0,27	0,08	0,35	0,13
	QNR	Sim.	0,60	0,28	0,10	0,35	0,13
		Est.	0,67	0,27	0,11	0,38	0,14
	QQN	Sim.	0,60	0,28	0,10	0,35	0,13
		Est.	0,67	0,27	0,11	0,38	0,14

[a] écart type simulé : écart type estimé par les simulations.

[b] écart type estimé : moyenne des écarts types estimés.

TAB. C.0.4. *Comparaison des écarts types estimés et simulés du modèle 3.1.1 avec 200 répétitions pour $m = 750, 1000$ et une vraie valeur de $\theta = 1,5$.*

m	Méthode		θ	β_0	β_1	β_2	β_3
750	LOGIT	Sim.[a]	—	0,14	0,05	0,30	0,10
		Est.[b]	—	0,14	0,05	0,26	0,10
	PQL	Sim.	0,11	0,14	0,05	0,30	0,10
		Est.	0,19	0,14	0,05	0,27	0,10
	CPQL0	Sim.	0,26	0,15	0,05	0,30	0,10
		Est.	0,47	0,15	0,06	0,28	0,10
	CPQL1	Sim.	0,26	0,20	0,06	0,30	0,11
		Est.	0,47	0,22	0,06	0,29	0,10
	CPQL2	Sim.	0,26	0,17	0,06	0,30	0,11
		Est.	0,47	0,22	0,06	0,29	0,10
	MPQL	Sim.	0,11	0,14	0,05	0,30	0,10
		Est.	0,19	0,14	0,05	0,27	0,10
	MCP0	Sim.	0,25	0,15	0,05	0,30	0,10
		Est.	0,46	0,15	0,06	0,28	0,10
	MCP1	Sim.	0,25	0,20	0,06	0,30	0,11
		Est.	0,46	0,22	0,07	0,28	0,10
	MCP2	Sim.	0,25	0,17	0,06	0,30	0,11
		Est.	0,46	0,22	0,07	0,28	0,10
	QNR	Sim.	0,52	0,22	0,08	0,34	0,12
		Est.	0,55	0,22	0,09	0,31	0,11
	QQN	Sim.	0,52	0,22	0,08	0,34	0,12
		Est.	0,55	0,22	0,09	0,31	0,11
1 000	LOGIT	Sim.	—	0,12	0,04	0,24	0,08
		Est.	—	0,12	0,05	0,23	0,09
	PQL	Sim.	0,10	0,12	0,04	0,24	0,08
		Est.	0,17	0,12	0,05	0,23	0,09
	CPQL0	Sim.	0,23	0,13	0,05	0,24	0,08
		Est.	0,41	0,13	0,05	0,24	0,09
	CPQL1	Sim.	0,23	0,18	0,06	0,25	0,09
		Est.	0,41	0,19	0,06	0,25	0,09
	CPQL2	Sim.	0,23	0,15	0,06	0,25	0,09
		Est.	0,41	0,19	0,06	0,25	0,09
	MPQL	Sim.	0,10	0,12	0,04	0,24	0,08
		Est.	0,17	0,12	0,05	0,23	0,09
	MCP0	Sim.	0,23	0,13	0,05	0,24	0,08
		Est.	0,40	0,13	0,05	0,24	0,09
	MCP1	Sim.	0,23	0,18	0,06	0,25	0,09
		Est.	0,40	0,19	0,06	0,25	0,09
	MCP2	Sim.	0,23	0,15	0,06	0,25	0,09
		Est.	0,40	0,19	0,06	0,25	0,09
	QNR	Sim.	0,47	0,20	0,08	0,27	0,10
		Est.	0,48	0,19	0,07	0,27	0,10
	QQN	Sim.	0,47	0,20	0,08	0,28	0,10
		Est.	0,48	0,19	0,07	0,27	0,10

[a] écart type simulé : écart type estimé par les simulations.

[b] écart type estimé : moyenne des écarts types estimés.

TAB. C.0.5. *Valeurs moyennes des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour* $m = 250, 500$ *et une vraie valeur de* $\theta = 0,5$.

m	Méthode	θ	β_0	β_1	β_2	β_3
Vraie valeur		0,50	-2,50	1,00	-1,00	0,50
250	LOGIT	—	-2,36	0,94	-1,07	0,51
	PQL	0,27	-2,37	0,94	-1,07	0,52
	CPQL0	0,77	-2,42	0,96	-1,06	0,52
	CPQL1	0,77	-2,70	0,90	-0,94	0,57
	CPQL2	0,77	-2,60	0,89	-0,98	0,56
	MQLP	0,25	-2,36	0,94	-1,09	0,52
	MCP0	0,71	-2,40	0,95	-1,09	0,52
	MCP1	0,71	-2,66	0,90	-0,97	0,57
	MCP2	0,71	-2,58	0,89	-1,01	0,57
	QNR	1,06	-2,71	1,08	-1,17	0,58
	QQN	0,99	-2,69	1,07	-1,16	0,58
500	LOGIT	—	-2,33	0,93	-1,00	0,49
	PQL	0,20	-2,34	0,93	0,99	0,49
	CPQL0	0,56	-2,37	0,95	-0,99	0,49
	CPQL1	0,56	-2,58	0,90	-0,90	0,52
	CPQL2	0,56	-2,53	0,90	-0,92	0,52
	MQLP	0,19	-2,35	0,94	-0,99	0,48
	MCP0	0,53	-2,37	0,95	-0,98	0,48
	MCP1	0,53	-2,57	0,91	-0,90	0,52
	MCP2	0,53	-2,52	0,90	-0,92	0,52
	QNR	0,71	-2,59	1,04	-1,05	0,53
	QQN	0,66	-2,57	1,03	-1,06	0,53

TAB. C.0.6. *Valeurs moyennes des paramètres estimés du modèle 3.1.1 avec 200 répétitions pour* $m = 750, 1000$ *et une vraie valeur de* $\theta = 0,5$.

m	Méthode	θ	β_0	β_1	β_2	β_3
Vraie valeur		1,50	-2,50	1,00	-1,00	0,50
750	LOGIT	—	-2,33	0,94	-0,98	0,47
	PQL	0,18	-2,33	0,94	-0,97	0,46
	CPQL0	0,49	-2,36	0,95	-0,97	0,46
	CPQL1	0,49	-2,54	0,91	-0,89	0,49
	CPQL2	0,49	-2,50	0,90	-0,91	0,49
	MQLP	0,17	-2,33	0,94	-0,98	0,46
	MCP0	0,48	-2,35	0,95	-0,97	0,47
	MCP1	0,48	-2,53	0,91	-0,90	0,49
	MCP2	0,48	-2,49	0,91	-0,91	0,49
	QNR	0,60	-2,54	1,02	-1,03	0,50
	QQN	0,59	-2,53	1,02	-1,02	0,50
1000	LOGIT	—	-2,35	0,93	-0,94	0,47
	PQL	0,18	-2,35	0,93	-0,94	0,47
	CPQL0	0,49	-2,37	0,94	-0,94	0,47
	CPQL1	0,49	-2,55	0,90	-0,86	0,50
	CPQL2	0,49	-2,52	0,90	-0,87	0,50
	MQLP	0,17	-2,35	0,93	-0,94	0,47
	MCP0	0,47	-2,37	0,94	-0,94	0,47
	MCP1	0,47	-2,55	0,90	-0,86	0,50
	MCP2	0,47	-2,51	0,90	-0,87	0,50
	QNR	0,60	-2,56	1,02	-0,99	0,51
	QQN	0,59	-2,55	1,01	-0,99	0,51

TAB. C.0.7. *Comparaison des écarts types estimés et simulés du modèle 3.1.1 avec 200 répétitions pour* $m = 250, 500$ *et une vraie valeur de* $\theta = 0,5$.

m	Méthode		θ	β_0	β_1	β_2	β_3
250	LOGIT	Sim. [a]	—	0,28	0,11	0,59	0,20
		Est. [b]	—	0,28	0,11	0,60	0,22
	PQL	Sim.	0,16	0,28	0,11	0,58	0,20
		Est.	0,34	0,29	0,11	0,60	0,22
	CPQL0	Sim.	0,47	0,30	0,11	0,58	0,20
		Est.	0,99	0,30	0,11	0,61	0,22
	CPQL1	Sim.	0,47	0,41	0,13	0,59	0,21
		Est.	0,99	0,48	0,14	0,63	0,23
	CPQL2	Sim.	0,47	0,35	0,13	0,59	0,21
		Est.	0,99	0,48	0,14	0,63	0,23
	MPQL	Sim.	0,15	0,28	0,11	0,58	0,20
		Est.	0,34	0,29	0,11	0,60	0,22
	MCP0	Sim.	0,44	0,29	0,11	0,58	0,20
		Est.	0,98	0,30	0,11	0,61	0,22
	MCP1	Sim.	0,44	0,39	0,12	0,59	0,21
		Est.	0,98	0,47	0,14	0,63	0,23
	MCP2	Sim.	0,44	0,35	0,12	0,59	0,21
		Est.	0,98	0,47	0,14	0,63	0,23
	QNR	Sim.	0,85	0,43	0,17	0,61	0,22
		Est.	0,95	0,44	0,17	0,64	0,24
	QGA	Sim.	0,86	0,43	0,17	0,62	0,22
		Est.	0,92	0,43	0,17	0,64	0,23
500	LOGIT	Sim.	—	0,20	0,07	0,41	0,14
		Est.	—	0,20	0,08	0,40	0,15
	PQL	Sim.	0,12	0,20	0,07	0,41	0,15
		Est.	0,24	0,20	0,08	0,40	0,15
	CPQL0	Sim.	0,35	0,20	0,07	0,41	0,15
		Est.	0,66	0,21	0,08	0,40	0,15
	CPQL1	Sim.	0,35	0,28	0,08	0,42	0,16
		Est.	0,66	0,32	0,09	0,42	0,15
	CPQL2	Sim.	0,35	0,24	0,08	0,41	0,16
		Est.	0,66	0,32	0,09	0,42	0,15
	MPQL	Sim.	0,12	0,20	0,07	0,41	0,15
		Est.	0,24	0,20	0,08	0,40	0,15
	MCP0	Sim.	0,33	0,20	0,07	0,41	0,15
		Est.	0,66	0,21	0,08	0,40	0,15
	MCP1	Sim.	0,33	0,27	0,08	0,41	0,15
		Est.	0,66	0,32	0,09	0,42	0,15
	MCP2	Sim.	0,33	0,24	0,08	0,41	0,15
		Est.	0,66	0,32	0,09	0,42	0,15
	QNR	Sim.	0,52	0,28	0,11	0,43	0,16
		Est.	0,56	0,28	0,11	0,42	0,15
	QGA	Sim.	0,54	0,29	0,11	0,42	0,16
		Est.	0,55	0,28	0,11	0,42	0,15

[a] écart type simulé : écart type estimé par les simulations.

[b] écart type estimé : moyenne des écarts types estimés.

TAB. C.0.8. *Comparaison des écarts types estimés et simulés du modèle 3.1.1 avec 200 répétitions pour* $m = 750, 1000$ *et une vraie valeur de* $\theta = 0,5$.

m	Méthode		θ	β_0	β_1	β_2	β_3
750	LOGIT	Sim.[a]	—	0,16	0,06	0,33	0,12
		Est.[b]	—	0,16	0,06	0,32	0,12
	PQL	Sim.	0,10	0,16	0,06	0,32	0,11
		Est.	0,19	0,16	0,06	0,32	0,12
	CPQL0	Sim.	0,26	0,17	0,06	0,32	0,11
		Est.	0,53	0,17	0,06	0,32	0,12
	CPQL1	Sim.	0,26	0,21	0,06	0,32	0,12
		Est.	0,53	0,26	0,08	0,33	0,12
	CPQL2	Sim.	0,26	0,19	0,07	0,32	0,12
		Est.	0,53	0,26	0,08	0,33	0,12
	MPQL	Sim.	0,09	0,16	0,06	0,32	0,11
		Est.	0,19	0,16	0,06	0,32	0,12
	MCP0	Sim.	0,25	0,17	0,06	0,32	0,11
		Est.	0,53	0,17	0,06	0,32	0,12
	MCP1	Sim.	0,25	0,21	0,06	0,33	0,12
		Est.	0,53	0,26	0,08	0,33	0,12
	MCP2	Sim.	0,25	0,19	0,07	0,32	0,12
		Est.	0,53	0,26	0,08	0,33	0,12
	QNR	Sim.	0,37	0,22	0,08	0,33	0,12
		Est.	0,43	0,22	0,09	0,33	0,12
	QGA	Sim.	0,37	0,21	0,08	0,33	0,12
		Est.	0,43	0,22	0,09	0,33	0,12
1 000	LOGIT	Sim.	—	0,15	0,05	0,27	0,10
		Est.	—	0,14	0,05	0,27	0,10
	PQL	Sim.	0,09	0,15	0,05	0,27	0,10
		Est.	0,17	0,14	0,05	0,27	0,10
	CPQL0	Sim.	0,24	0,15	0,05	0,27	0,10
		Est.	0,46	0,14	0,05	0,28	0,10
	CPQL1	Sim.	0,24	0,19	0,06	0,28	0,10
		Est.	0,46	0,22	0,07	0,29	0,10
	CPQL2	Sim.	0,24	0,18	0,06	0,28	0,10
		Est.	0,46	0,22	0,07	0,29	0,10
	MPQL	Sim.	0,08	0,15	0,05	0,27	0,10
		Est.	0,16	0,14	0,05	0,27	0,10
	MCP0	Sim.	0,23	0,15	0,05	0,27	0,10
		Est.	0,46	0,14	0,05	0,28	0,10
	MCP1	Sim.	0,23	0,19	0,06	0,28	0,10
		Est.	0,46	0,22	0,07	0,29	0,10
	MCP2	Sim.	0,23	0,18	0,06	0,28	0,10
		Est.	0,46	0,22	0,07	0,29	0,10
	QNR	Sim.	0,33	0,20	0,08	0,29	0,11
		Est.	0,37	0,20	0,08	0,29	0,11
	QGA	Sim.	0,34	0,20	0,08	0,28	0,10
		Est.	0,37	0,19	0,08	0,29	0,11

[a] écart type simulé : écart type estimé par les simulations.

[b] écart type estimé : moyenne des écarts types estimés.

Annexe D

DISTRIBUTION DES PARAMÈTRES ESTIMÉS POUR LE MODÈLE 3.1.3

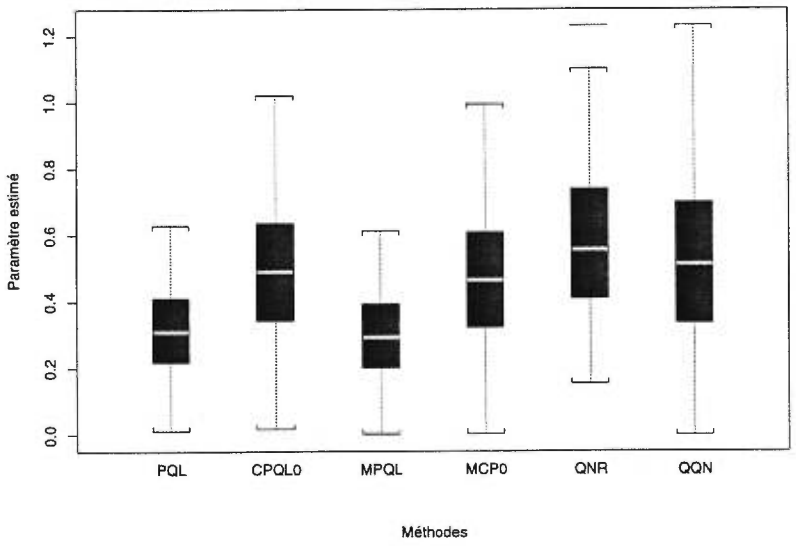

FIG. D.0.1. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.3 avec $m = 1000$ individus (vraie valeur = 0,5).*

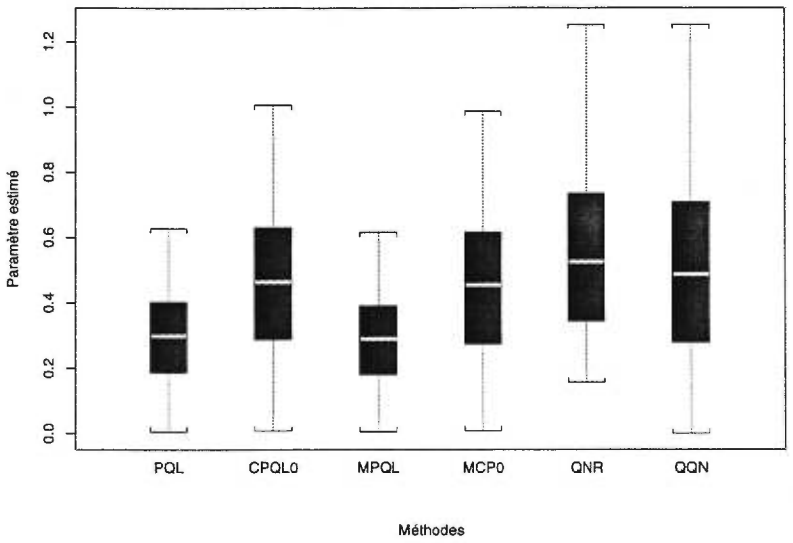

FIG. D.0.2. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.3 avec $m = 1500$ individus (vraie valeur =0,5).*

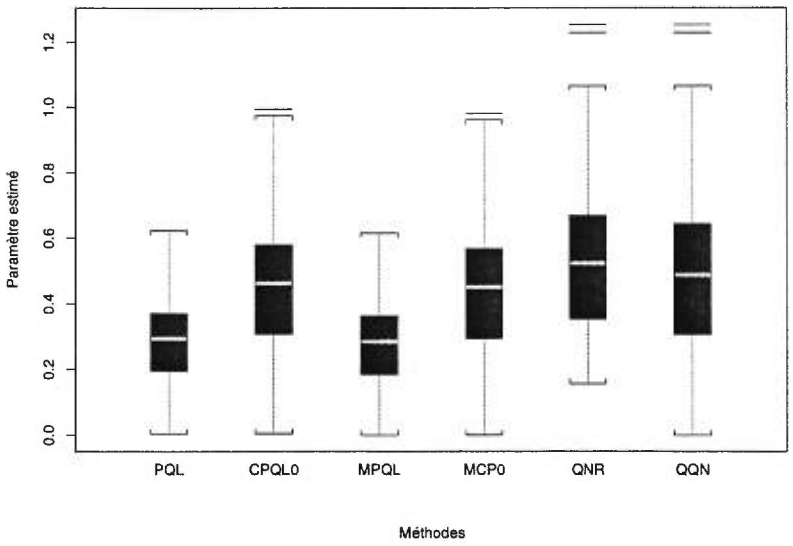

FIG. D.0.3. *Graphiques en boîtes des estimateurs de θ pour le modèle 3.1.3 avec $m = 2000$ individus (vraie valeur = 0,5).*

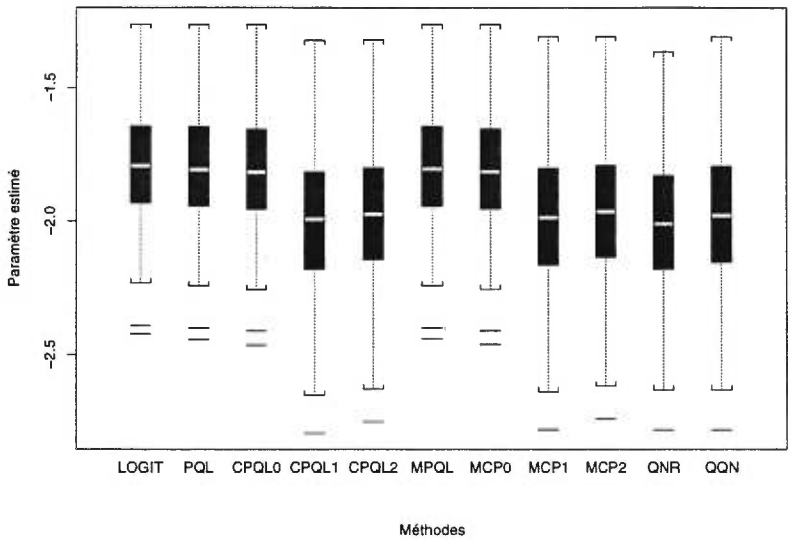

FIG. D.0.4. *Graphiques en boîtes des estimateurs de β_0 pour le modèle 3.1.3 avec $m = 1000$ individus (vraie valeur = -1,96).*

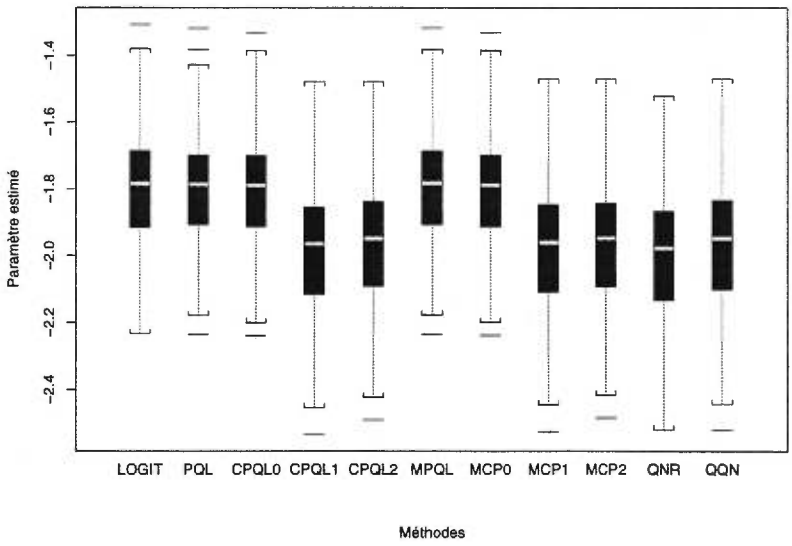

FIG. D.0.5. *Graphiques en boîtes des estimateurs de β_0 pour le modèle 3.1.3 avec $m = 1500$ individus (vraie valeur = -1,96).*

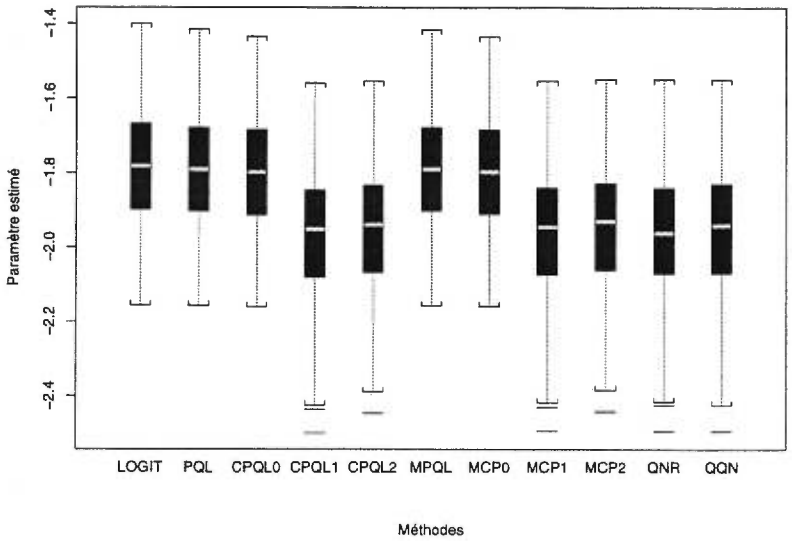

FIG. D.0.6. *Graphiques en boîtes des estimateurs de β_0 pour le modèle 3.1.3 avec $m = 2000$ individus (vraie valeur = -1,96).*

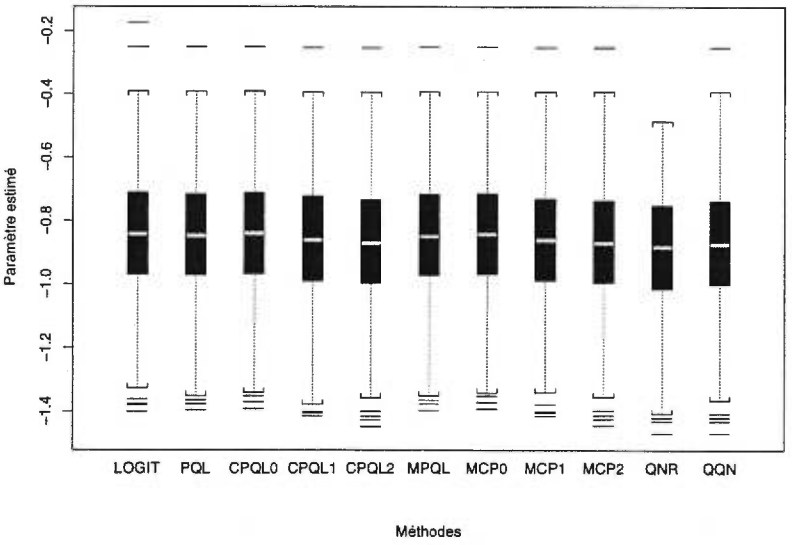

FIG. D.0.7. *Graphiques en boîtes des estimateurs de β_1 pour le modèle 3.1.3 avec $m = 1000$ individus (vraie valeur = -0,86).*

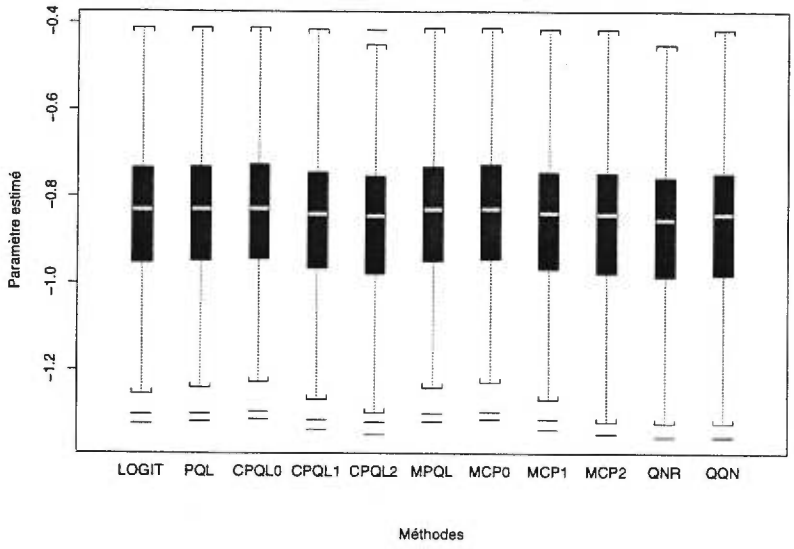

FIG. D.0.8. *Graphiques en boîtes des estimateurs de β_1 pour le modèle 3.1.3 avec $m = 1500$ individus (vraie valeur = -0,86).*

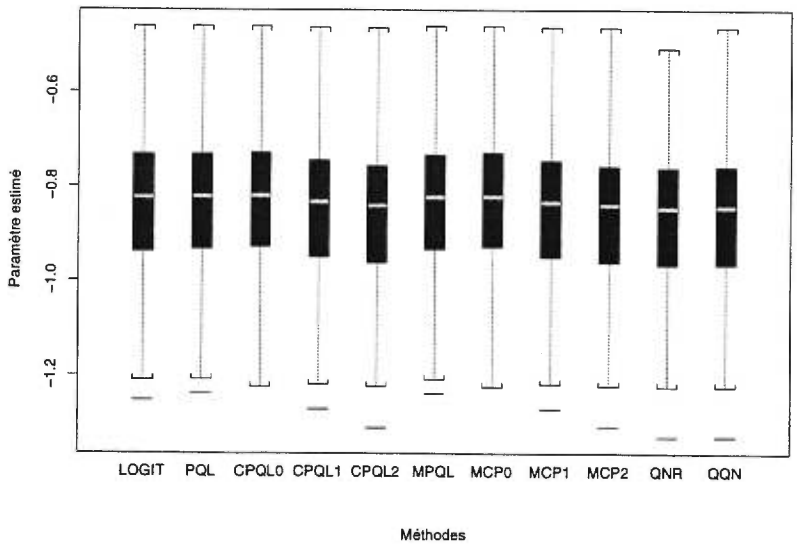

FIG. D.0.9. *Graphiques en boîtes des estimateurs de β_1 pour le modèle 3.1.3 avec $m = 2000$ individus (vraie valeur = -0,86).*

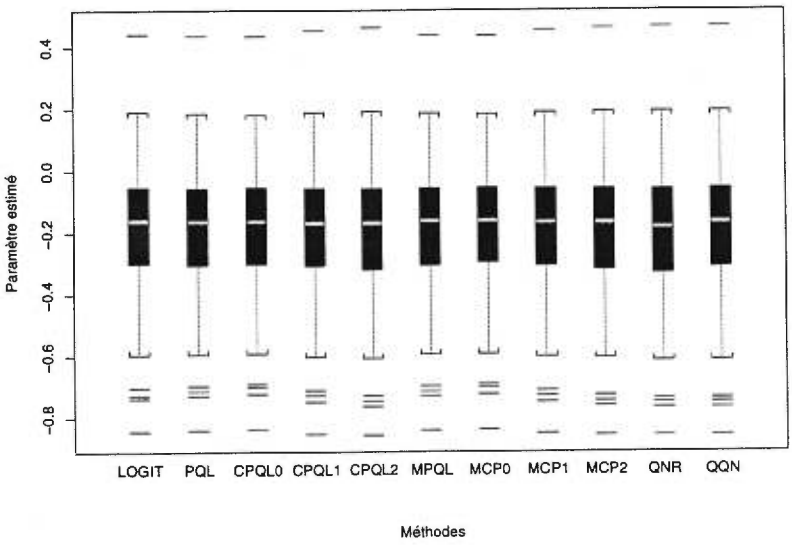

FIG. D.0.10. *Graphiques en boîtes des estimateurs de β_2 pour le modèle 3.1.3 avec $m = 1000$ individus (vraie valeur = -0,17).*

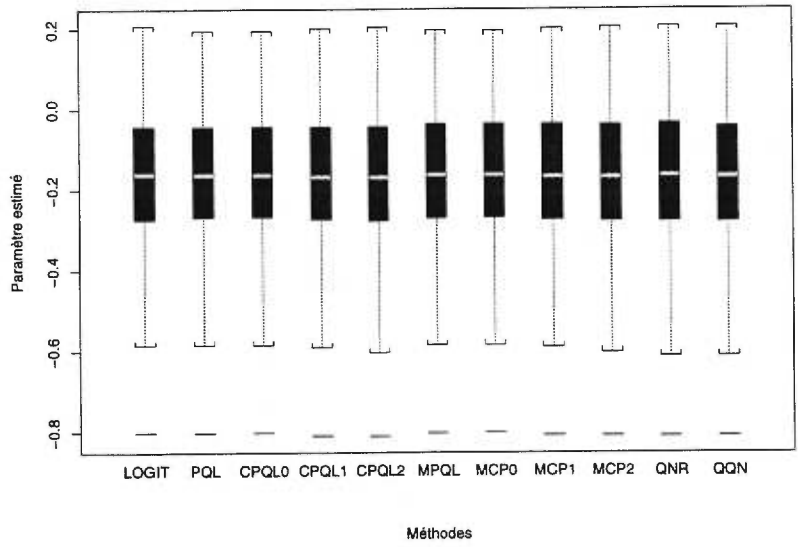

FIG. D.0.11. *Graphiques en boîtes des estimateurs de β_2 pour le modèle 3.1.3 avec $m = 1500$ individus (vraie valeur = -0,17).*

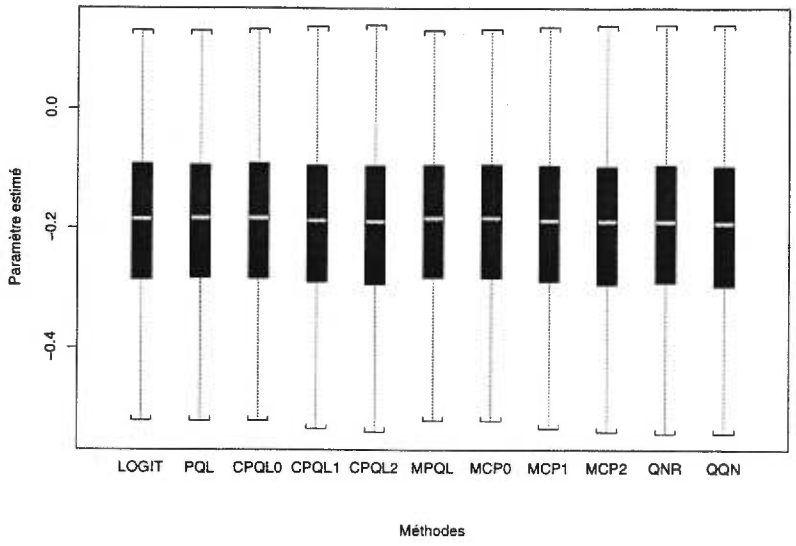

FIG. D.0.12. *Graphiques en boîtes des estimateurs de β_2 pour le modèle 3.1.3 avec $m = 2000$ individus (vraie valeur = -0,17).*

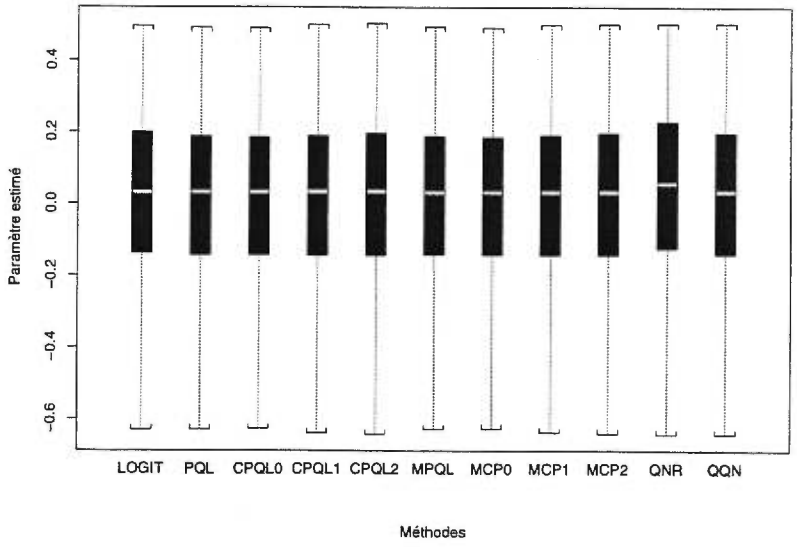

FIG. D.0.13. *Graphiques en boîtes des estimateurs de β_3 pour le modèle 3.1.3 avec $m = 1000$ individus (vraie valeur = 0,04).*

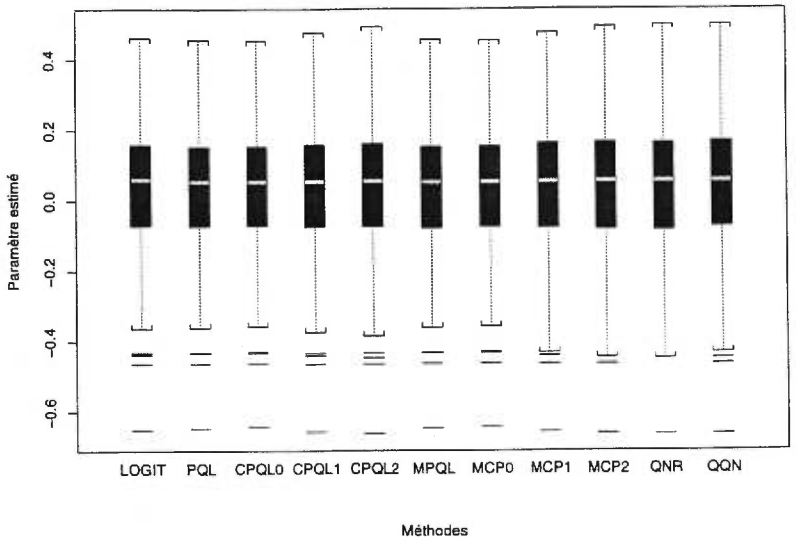

FIG. D.0.14. *Graphiques en boîtes des estimateurs de β_3 pour le modèle 3.1.3 avec $m = 1500$ individus (vraie valeur = 0,04).*

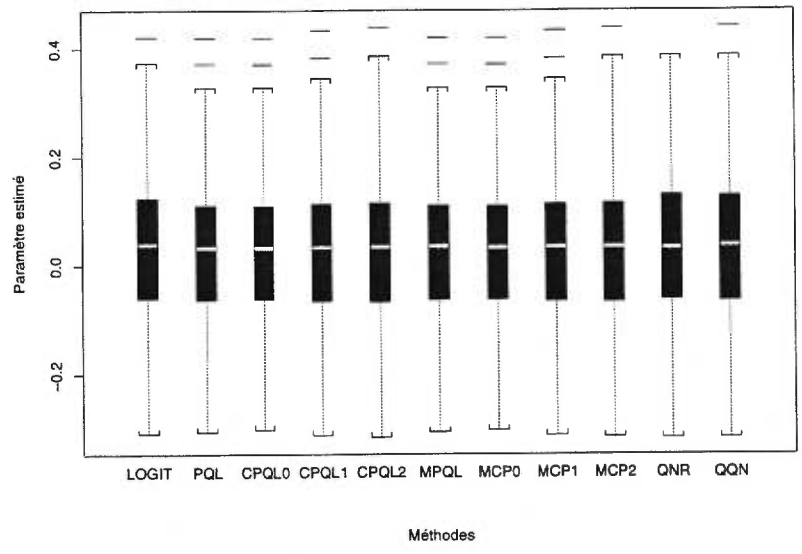

FIG. D.0.15. *Graphiques en boîtes des estimateurs de β_3 pour le modèle 3.1.3 avec $m = 2000$ individus (vraie valeur = 0,04).*

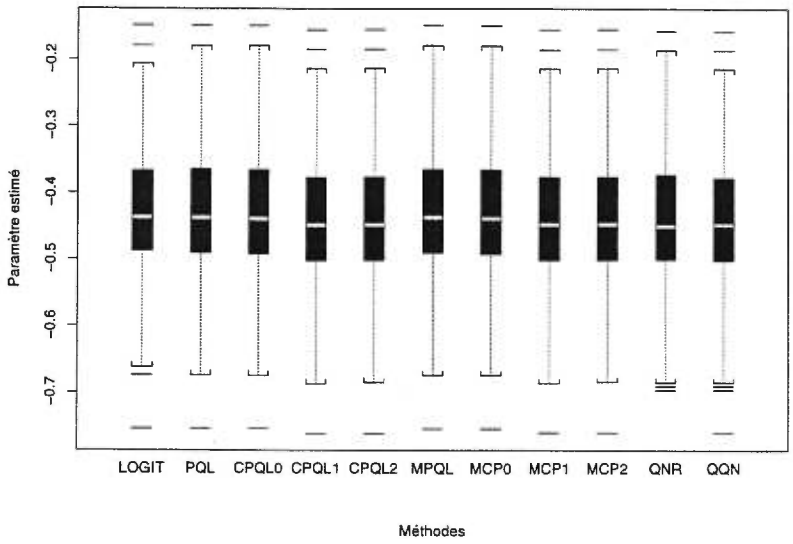

FIG. D.0.16. *Graphiques en boîtes des estimateurs de β_4 pour le modèle 3.1.3 avec $m = 1000$ individus (vraie valeur = -0,44).*

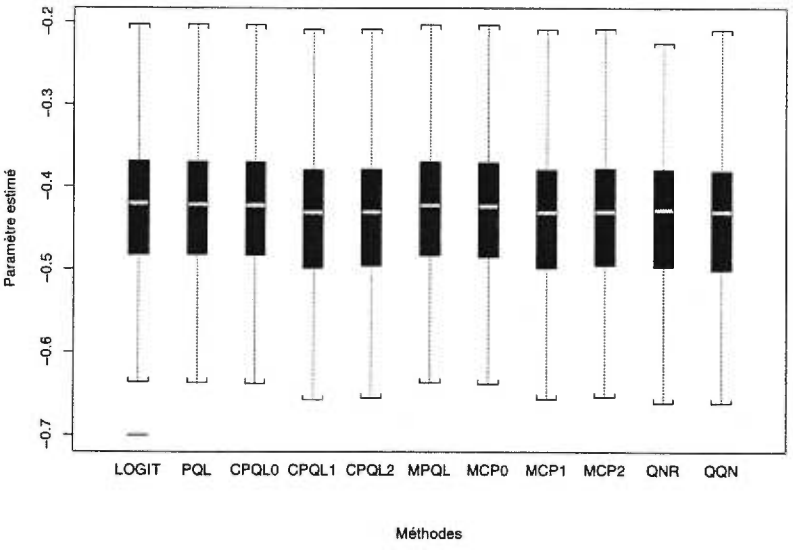

FIG. D.0.17. *Graphiques en boîtes des estimateurs de β_4 pour le modèle 3.1.3 avec $m = 1500$ individus (vraie valeur = -0,44).*

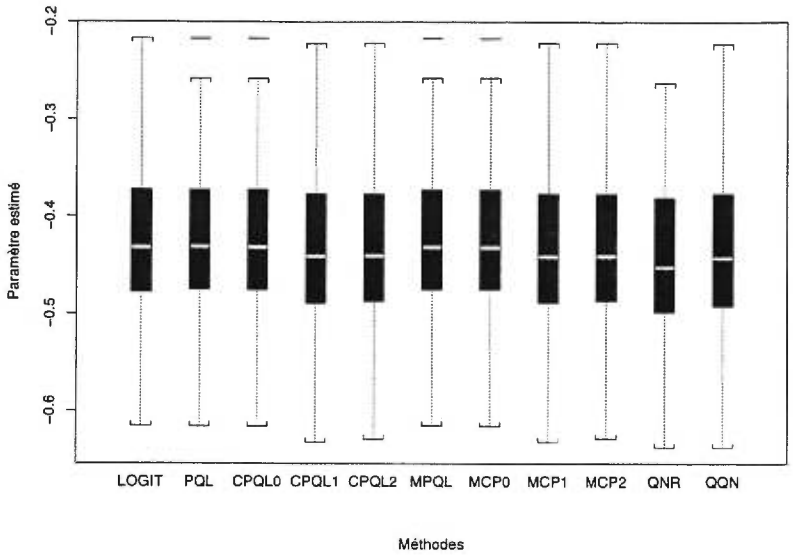

FIG. D.0.18. *Graphiques en boîtes des estimateurs de β_4 le modèle 3.1.3 avec pour $m = 2000$ individus (vraie valeur = -0,44).*

BIBLIOGRAPHIE

ABRAMOWITZ, M. ET STEGUN, I.A. (1964), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.

AGRESTI, A. (1989), A survey of models for repeated ordered categorical response data, *Statistics in Medicine*, **8**, 1209-1224.

AGRESTI, A. (1990), *Categorical Data Analysis*, Wiley, New York.

AGRESTI, A. (1996), *An Introduction to Categorical Data Analysis*, Wiley, New York.

ANDERSON, D.A. ET AITKIN, M. (1985), Variance component models with binary response, *Journal of the Royal Statistical Society*, **B**, **47**, 203-210.

BARNDORFF-NIELSEN, O.E. ET COX, D.R. (1989), *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

BRESLOW, N.E. ET CLAYTON, D.G. (1993), Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9-25.

BRESLOW, N.E. ET LIN, X. (1995), Bias correction in generalized linear mixed models with a single component of dispersion, *Biometrika*, **82**, 81-91.

CAREY, V.C., ZEGER, S.L. ET DIGGLE, P.J. (1993), Modelling multivariate binary data with alternating logistic regressions, *Biometrika*, **80**, 517-526.

CONEWAY, M.R. (1990), A random effects model for binary data, *Biometrics*, **46**, 317-328.

COX, D.R. ET REID, N. (1987), Parameter orthogonality and approximate conditional inference, *Journal of the Royal Statistical Society*, **B**, **49**, 1-39.

CROUCH, A.C. ET SPIEGELMAN, E. (1990), The evaluation of integrals of the form $\int f(t)\exp(-t^2)dt$: application to logistic-normal models, *Journal of the American Statistical Association*, **85**, 464-469.

DAVIDIAN, M. ET GALLANT, A.R. (1992), Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine, *Journal of Pharmacokinetics and Biopharmaceutics*, **20**, 529-556.

DAVIS, P.J. ET RABINOWITZ, P. (1984), *Methods of Numerical Integration*, 2de edition, Academic Press, New York.

DEMPSTER, A.P., LAIRD, N.M. ET RUBIN, D.B. (1977), Maximum likelihood from incomplete data via EM algorithm (with discussion), *Journal of the Royal Statistical Society*, **B**, **39**, 1-38.

DIGGLE, P.J., LIANG, K.-Y. ET ZEGER, S.L. (1994), *Analysis of Longitudinal Data*, Clarendon Press, Oxford.

FITZMAURICE, G.M., HEATH, A.F. ET CLIFFORD, P. (1996), Logistic regression models for binary panel data with attrition, *Journal of the Royal Statistical Society*, **A**, **159**, 249-263.

FITZMAURICE, G.M., LAIRD, N.M. ET ROTNITSKY, A.G. (1993), Regression models for discrete longitudinal responses (with discussion), *Statistical Science*, **8**, 284-309.

FLETCHER, R. (1987), *Practical Methods of Optimization*, 2de edition, Wiley, Chichester.

GOLUB, G.H. ET WELSCH, J.H. (1969), Calculation of Gaussien quadrature rules, *Mathematical Computing*, **23**, 221-230.

GOODWIN, E.T. (1949), The evaluation of integrals of the form $\int_{-\infty}^{+\infty} f(x)e^{-x^2}dx$, *Proceedings of the Cambridge Philosophical Society*, **45**, 241-245.

GRIZZLE, J.E., STAMER C.F. ET KOCH, G.G. (1969), Analysis of categorical data by linear models, *Biometrics*, **25**, 489-504.

HARVILLE, D.A. (1977), Maximum likelihood approaches to variance component estimation and to related problems (with discussion), *Journal of the American Statistical Association*, **72**, 320-340.

KENWARD, M.G. ET JONES, B. (1992), Alternative approaches to the analysis of binary and categorical repeated measurements, *Journal of Biopharmaceutical Statistics*, **2(2)**, 137-170.

KOCH, G.G., LANDIS, J.R., FREEMAN, J.L., FREEMAN, D.H. ET LEHNEN R.G. (1977), A general methodology for the analysis of experiments with repeated measurement of categorical data, *Biometrics*, **33**, 133-158.

LABERGE-NADEAU, C., MAAG, U., BOURBEAU, R., DESJARDINS, D., MESSIER, S. ET HIRSCH, P. (1999), *Le lien entre la performance aux examens (théorique et pratique) pour l'obtention d'un permis et le taux d'implication dans les accidents*, Laboratoire sur la sécurité des transports du Centre de recherche sur les transports de l'Université de Montréal, Publication CRT-99-56, Montréal.

LIANG, K.-Y. ET ZEGER, S.L. (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13-22.

LIANG, K.-Y., ZEGER, S.L. ET QAQISH, B. (1992), Multivariate regression analysis for categorical data (with discussion), *Journal of the Royal Statistical Society,* **B**, **54**, 3-40.

LIN, X. ET BRESLOW, N.E. (1996a), Bias correction in generalized linear mixed models with multiple components of dispersion, *Journal of the American Statistical Association*, **91**, 1007-1016.

LIN, X. ET BRESLOW, N.E. (1996b), Analysis of correlated binomial data in logistic-normal models, *Journal of Statistical Computation and Simulation*, **55**, 133-146.

LINDSTROM, M.J. ET BATES, D.M. (1988), Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *Journal of the American Statistical Association*, **83**, 1014-1022.

LINDSTROM, M.J. ET BATES, D.M. (1990), Nonlinear mixed effects models for repeated measures data, *Biometrics*, **46**, 673-687.

LIPSITZ, S.R., FITZMAURICE, G.M., ORAV, E.J. ET LAIRD, N.M. (1994), Performance of generalized estimating equations in practical situations, *Biometrics*, **50**, 270-278.

LIPSITZ, S.R., KIM, K., ET ZHAO, L. (1994), Analysis of repeated categorical data using generalized estimating equations, *Statistics in Medicine*, **13**, 1149-1163.

LIPSITZ, S.R., LAIRD, N.M. ET HARRINGTON, D.P. (1991), Generalized estimating equations for correlated binary data: using odds ratios as a measure of association, *Biometrika*, **78**, 153-160.

LIU, Q. ET PIERCE, D.A. (1993), Heterogeneite in Mantel-Haenszel-type models, *Biometrika*, **80**, 543-556.

LITTELL, R.C., MILLIKEN, G.A., STROUP, W.W. ET WOLFINGER, R.G. (1996), *SAS System for Mixed Models*, SAS Institute Inc., Cary, NC.

MCCULLAGH, P. (1983), Quasi-likelihood functions, *Annals of Statistics*, **11**, 59-67.

MCCULLAGH, P. ET NELDER, J.A. (1989), *Generalized Linear Models*, 2de edition, Chapman and Hall, London.

NELDER, J.A. ET WEDDERBURN, R.W.M. (1972), Generalized linear models, *Journal of the Royal Statistical Society,* **A**, **135**, 370-384.

NEUHAUS, J.M., KALBFLEISCH, J.D. ET HAUCK, W.W. (1991), A comparaison of cluster-specific and population averaged approaches for analyzing correlated binary data, *International Statistical Review*, **59**, 25-36.

PINHEIRO, J.C. ET BATES, D.M. (1995), Approximations to the log-likelihood function in the nonlinear mixed-effects model, *Journal of Computational and Graphical Statistics*, **4**, 12-35.

PRENTICE, R.L. (1988), Correlated binary regression with covariates specific to each binary observation, *Biometrics*, **44**, 1033-1048.

ORME, C.D. ET FRY, T.R.L. (1995), Maximum likelihood estimation in binary data models using panel data under alternative distributional assumptions, *Economics Letters*, **49**, 359-366.

ROSNER, B. (1984), Multivariate methods in ophthalmology with application to other paired-data situations, *Biometrics*, **40**, 1025-1035.

SAS/STAT SOFTWARE (1996), *Changes and Enhancements through release 6.11*, SAS Institute Inc., Cary, NC.

SAS/IML SOFTWARE (1990), *Usage and Reference, Version 6*, SAS Institute Inc., Cary, NC.

SCHALL, R. (1991), Estimation in generalized linear models with random effects, *Biometrika*, **78**, 719-727.

SKELLAM, J.G. (1948), A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials, *Journal of the Royal Statistical Society*, **B**, **10**, 257-261.

SOLOMON, P.J. ET COX, D.R. (1992), Nonlinear components of variance models, *Biometrika*, **79**, 1-11.

STIRALLI, R., LAIRD, N., ET WARE, J.H. (1984), Random effects models for serial observations with binary responses, *Biometrics*, **40**, 961-971.

STOKES, M.E., DAVIS, C.S. ET KOCH, G.G. (1995), *Categorical Data Analysis Using the SAS System*, SAS Institute Inc., Cary, NC.

THALL, P.F. ET VAIL, S.C. (1990), Some covariance models for longitudinal count data with overdispersion, *Biometrics*, **40**, 657-671.

WEDDERBURN, R.W.M. (1974), Quasi-likelihood functions, generalized linear models ans the Gaussian method, *Biometrika*, **61**, 439-447.

WOLFINGER, R.D. (1993a), *The GLIMMIX SAS Macro*, SAS Institute Inc., Cary, NC.

WOLFINGER, R.D. (1993b), Laplace's approximation for nonlinear mixed models, *Biometrika*, **80**, 791-795.

WOLFINGER, R.D. ET O'CONNELL, M. (1993), Generalized linear mixed models: a pseudo-likelihood approach, *Journal of Statistical Computation and Simulation*, **48**, 233-243.

ZEGER, S.L. ET KARIM, M.R. (1991), Generalized linear models with random effects: a Gibbs sampling approach, *Journal of the American Statistical Association*, **86**, 79-86.

ZEGER, S.L. ET LIANG, K.-Y. (1986), Longitudinal data analysis for discrete and continuous outcomes, *Biometrics*, **42**, 121-130.

ZEGER, S.L. ET LIANG, K.-Y. (1992), An overview of methods for the analysis of longitudinal data, *Statistics in Medicine*, **11**, 1825-1839.

ZEGER, S.L., LIANG, K.-Y. ET ALBERT, P.S. (1988), Models for longitudinal data: a generalized estimating equation approach, *Biometrics*, **44**, 1049-1060.

ZHAO, L.P. ET PRENTICE, R.L. (1990), Correlated binary regression using a generalized quadratic model, *Biometrika*, **77**, 642-648.