Université de Montréal

Modélisation et prévision pour des séries chronologiques à valeurs entières

par

Yves Lafortune

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

octobre 1998

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Modélisation et prévision pour des séries chronologiques à valeurs entières

présenté par

Yves Lafortune

a été évalué par un jury composé des personnes suivantes :

Yves Lepage

(président-rapporteur)

Roch Roy

(directeur de recherche)

Alain Latour

(membre du jury)

Mémoire accepté le :

Le 3 novembre *1998*

SOMMAIRE

Les séries chronologiques à valeurs discrètes constituent une sphère de recherche de plus en plus active. Diverses méthodes ont été proposées pour étudier de telles séries. L'une d'entre elles, développée entre autres par Zeger (1988) et par Blais, MacGibbon et Roy (1997), est une généralisation du modèle linéaire généralisé au cas d'observations dépendantes. Jusqu'à présent, ce type de modèles permettait l'établissement de prévisions mais il ne permettait pas la construction d'intervalles de prévision pour les valeurs futures de la série. L'objectif premier de ce mémoire est d'apporter une contribution, ne serait-ce que modeste, à la solution de ce problème.

Nous présentons deux solutions, que nous croyons originales, permettant la construction d'intervalles de prévision pour les valeurs futures de la série pour ce type de modèles. La première solution, paramétrique, est basée sur la spécification de la distribution marginale du processus latent $\{\epsilon_t\}$. Nous supposons que la distribution marginale de ϵ_t est une distribution conjuguée naturelle de la distribution de $Y_t|\epsilon_t$. La deuxième solution, non paramétrique, repose sur l'utilisation d'une transformation T qui stabilise la variance de la série et sur l'utilisation des différences, à l'échelle transformée, entre les observations et les valeurs ajustées.

Les performances des deux méthodes sont également évaluées à l'aide de simulations et la méthodologie est finalement appliquée à des données réelles provenant du domaine de l'épidémiologie.

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche, Roch Roy, qui a bien su doser encadrement et autonomie pour faire de moi aujourd'hui un chercheur davantage accompli qu'il y a deux ans. Je veux aussi le remercier pour les nombreuses conversations des dernières années, bénéfiques tant sur le plan professionnel que sur le plan humain.

Je veux également exprimer mes remerciements au Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) pour la bourse d'études qu'il a bien voulu m'octroyer. Ce support financier a grandement facilité la réalisation du projet et a été très apprécié. Cette recherche a également bénéficié du support financier des Fonds FCAR et CRSNG via les fonds de recherche octroyés à mon directeur.

Sur une base plus personnelle, je veux remercier mes parents, Gérard et Lucie, pour leur constant support, de même que pour leur intérêt envers mes études, sans cesse manifesté. Merci du fin fond du coeur. Un merci tout aussi profond à l'endroit de mes deux soeurs, Jeanne et Line, amies et confidentes.

Je tiens également à dire un gros merci à tous(tes) mes amis(es) qui ont marqué, chacun(e) à leur manière, les deux dernières années et qui ont contribué, par leur amitié, à la réalisation de ce mémoire.

Table des matières

Table des figures

Liste des tableaux

INTRODUCTION

Les séries chronologiques à valeurs entières font l'objet d'études depuis seulement une quinzaine d'années. Devant l'inadéquation des méthodes développées pour des séries à valeurs continues, de nouvelles approches ont dû être proposées. Les modèles à espace d'états généralisés tels que décrits dans Brockwell et Davis (1996) constituent l'une de ces approches. Divisés en deux classes de modèles, soient les modèles dictés par les états (traduction libre de "parameter-driven models") et les modèles dictés par les observations (traduction libre de "observation-driven models"), ils utilisent un processus latent $\{\epsilon_t\}$ pour introduire de la dépendance entre les observations $Y_1, \ldots, Y_n$.

Blais, MacGibbon et Roy (1997), inspirés par les travaux de Zeger (1988), ont proposé une généralisation du modèle linéaire généralisé au cas d'observations dépendantes en utilisant l'approche des modèles dictés par les états. Dans le cadre de cette généralisation, ils font intervenir un processus latent $\{\epsilon_t\}$ qui induit des autocorrélations entre les observations $Y_1, \ldots, Y_n$. Conditionnellement aux états ϵ_t, les observations Y_t sont indépendantes et distribuées selon une loi de la famille exponentielle de distribution. Pour estimer les paramètres de régression, une méthode de quasi-vraisemblance, la méthode des équations d'estimation, est utilisée. Jusqu'à présent, ce type de modèles permettait l'établissement de prévisions mais il ne permettait pas la construction d'intervalles de prévision pour les valeurs futures de la série. L'objectif premier de ce mémoire est d'apporter une contribution, ne serait-ce que modeste, à la solution de ce problème.

Ce mémoire est constitué de cinq chapitres. Dans le premier chapitre, nous présentons d'abord les deux classes de modèles à espace d'états généralisés tels que proposés par Brockwell et Davis (1996). Nous présentons ensuite la généralisation du modèle linéaire généralisé, telle que développée par Blais, MacGibbon et Roy (1997), tout en faisant le lien avec les modèles dictés par les états. Finalement, nous montrons comment il est possible d'obtenir des prévisions pour les valeurs futures de la série, tout en indiquant que cette façon de faire ne permet pas l'établissement d'intervalles de prévision pour ces valeurs.

Le deuxième chapitre présente deux solutions, que nous croyons originales, permettant la construction d'intervalles de prévision pour les valeurs futures de la série. La première solution, paramétrique, est basée sur la spécification de la distribution marginale du processus latent $\{\epsilon_t\}$. Nous supposons que la distribution marginale de ϵ_t est une distribution conjuguée naturelle de la distribution de $Y_t|\epsilon_t$. La deuxième solution, non paramétrique, repose sur l'utilisation d'une transformation T qui stabilise la variance de la série et sur l'utilisation des différences, à l'échelle transformée, entre les observations et les valeurs ajustées.

Le chapitre 3 est un complément nécessaire à l'évaluation des performances de la solution paramétrique. Il comprend deux sections. La première, inspirée de Sim (1986), indique une façon de simuler un processus AR(1) avec distribution marginale gamma. La deuxième, inspirée de McKenzie (1985), présente une façon de simuler un processus AR(1) avec distribution marginale bêta.

Quant au chapitre 4, il est consacré à l'étude des performances des deux méthodes de construction d'intervalles de prévision développées au chapitre 2.

Bien que l'étude soit restreinte au cas où $Y_t|\epsilon_t$ est poissonnien, elle couvre trois duos, connus ou estimés, de variance et de corrélation de délai 1, trois distributions marginales pour le processus latent et deux types d'intervalles (unilatéraux et bilatéraux) pour chaque méthode. Une analyse des performances et de l'impact de chacune des composantes est proposée. Finalement, une comparaison sommaire entre les performances des deux méthodes est présentée.

Dans le dernier chapitre, nous appliquons la méthodologie développée au domaine de l'épidémiologie. Nous étudierons la série classique de la poliomyélite, analysée par Zeger (1988), ainsi que deux séries d'incidence de maladies à déclaration obligatoire de la région de Montréal.

Chapitre 1

APPROCHE ACTUELLE

1.1. Modèles à espace d'états généralisés

Dans les dernières années, les modèles à espace d'états linéaires ont eu un impact important sur l'analyse des séries chronologiques et sur plusieurs autres domaines connexes. Utilisés avec les filtres de Kalman, ces modèles ont fourni des solutions adéquates à plusieurs situations complexes. Malheureusement, plusieurs cas pouvaient difficilement être couverts à cause de la rigidité du cadre linéaire. Récemment, des modèles à espace d'états non linéaires ont été développés pour pallier ce manque: les modèles à espace d'états généralisés.

Tels que définis par Brockwell et Davis (1996), les modèles à espace d'états généralisés regroupent deux classes de modèles: les modèles régis par les états (traduction libre de "parameter-driven model") et les modèles régis par les observations (traduction libre de "observation-driven model"). Ils sont tous deux constitués par deux équations: l'équation des observations et l'équation des états. La distinction se situe au niveau de la deuxième équation: le vecteur des états d'un modèle dicté par les états évolue indépendamment du passé du processus observationnel alors que celui d'un modèle dicté par les observations dépend des observations passées.

L'équation des observations détermine la distribution de Y_t, étant donné l'état ϵ_t. De façon plus précise, si on dispose d'une suite d'observations $\{Y_t\}$ et d'une suite d'états $\{\epsilon_t\}$, et si on dénote les vecteurs de dimension t par $\mathbf{Y}^{(t)} = (Y_1, \ldots, Y_t)$ et $\boldsymbol{\epsilon}^{(t)} = (\epsilon_1, \ldots, \epsilon_t)$, on fait l'hypothèse que la distribution de $Y_t|(\epsilon_t, \mathbf{Y}^{(t-1)}, \boldsymbol{\epsilon}^{(t-1)})$ est indépendante de $(\mathbf{Y}^{(t-1)}, \boldsymbol{\epsilon}^{(t-1)})$, c'est-à-dire:

$$f(y_t|\epsilon_t, \mathbf{y}^{(t-1)}, \boldsymbol{\epsilon}^{(t-1)}) = f(y_t|\epsilon_t), \qquad t = 1, 2, \ldots \tag{1.1.1}$$

Pour les modèles dictés par les états, on suppose de plus que $\epsilon_{t+1}|(\epsilon_t, \boldsymbol{\epsilon}^{(t-1)}, \mathbf{Y}^{(t)})$ est indépendant de $(\boldsymbol{\epsilon}^{(t-1)}, \mathbf{Y}^{(t)})$. L'équation des états prend donc la forme suivante:

$$f(\epsilon_{t+1}|\epsilon_t, \mathbf{y}^{(t)}, \boldsymbol{\epsilon}^{(t-1)}) = f(\epsilon_{t+1}|\epsilon_t), \qquad t = 1, 2, \ldots \tag{1.1.2}$$

Pour ce type de modèle, nous pouvons aisément déterminer la distribution conjointe des états et des observations à partir des équations (1.1.1) et (1.1.2). En effet, selon la notation spécifiée ci-haut,

$$\begin{aligned} f(y_1, \ldots, y_n, \epsilon_1, \ldots, \epsilon_n) &= f(\mathbf{y}^{(n)}, \boldsymbol{\epsilon}^{(n)}) \\ &= f(y_n|\epsilon_n, \mathbf{y}^{(n-1)}, \boldsymbol{\epsilon}^{(n-1)}) f(\epsilon_n, \mathbf{y}^{(n-1)}, \boldsymbol{\epsilon}^{(n-1)}), \end{aligned}$$

et les équations (1.1.1) et (1.1.2) entraînent respectivement que

$$\begin{aligned} f(y_1, \ldots, y_n, \epsilon_1, \ldots, \epsilon_n) &= f(y_n|\epsilon_n) f(\epsilon_n, \mathbf{y}^{(n-1)}, \boldsymbol{\epsilon}^{(n-1)}) \\ &= f(y_n|\epsilon_n) f(\epsilon_n|\mathbf{y}^{(n-1)}, \boldsymbol{\epsilon}^{(n-1)}) f(\mathbf{y}^{(n-1)}, \boldsymbol{\epsilon}^{(n-1)}) \\ &= f(y_n|\epsilon_n) f(\epsilon_n|\epsilon_{n-1}) f(\mathbf{y}^{(n-1)}, \boldsymbol{\epsilon}^{(n-1)}), \end{aligned}$$

d'où, en répétant les mêmes étapes $(n-1)$ fois,

$$f(y_1,\dots,y_n,\epsilon_1,\dots,\epsilon_n) = \Big(\prod_{j=2}^{n} f(y_j|\epsilon_j)\Big)\Big(\prod_{j=2}^{n} f(\epsilon_j|\epsilon_{j-1})\Big) f(y_1,\epsilon_1)$$
$$= \Big(\prod_{j=1}^{n} f(y_j|\epsilon_j)\Big)\Big(\prod_{j=2}^{n} f(\epsilon_j|\epsilon_{j-1})\Big) f(\epsilon_1)\,, \qquad (1.1.3)$$

où l'on suppose que ϵ_1 admet une densité initialement déterminée $f(\epsilon_1)$. Il est important de remarquer que (1.1.2) implique que $\{\epsilon_t\}$ est markovien. En effet,

$$f(\epsilon_{t+1},\epsilon_t,\boldsymbol{\epsilon}^{(t-1)}) = \int\cdots\int_{D_{\mathbf{y}^{(t-1)}}} f(\epsilon_{t+1},\epsilon_t,\boldsymbol{\epsilon}^{(t-1)},\mathbf{y}^{(t-1)})\,d\mathbf{y}^{(t-1)}$$
$$= \int\cdots\int_{D_{\mathbf{y}^{(t-1)}}} f(\epsilon_{t+1}|\epsilon_t,\boldsymbol{\epsilon}^{(t-1)},\mathbf{y}^{(t-1)}) f(\epsilon_t,\boldsymbol{\epsilon}^{(t-1)},\mathbf{y}^{(t-1)})\,d\mathbf{y}^{(t-1)}\,,$$

et grâce à l'équation (1.1.2),

$$f(\epsilon_{t+1},\epsilon_t,\boldsymbol{\epsilon}^{(t-1)}) = \int\cdots\int_{D_{\mathbf{y}^{(t-1)}}} f(\epsilon_{t+1}|\epsilon_t) f(\epsilon_t,\boldsymbol{\epsilon}^{(t-1)},\mathbf{y}^{(t-1)})\,d\mathbf{y}^{(t-1)}$$
$$= f(\epsilon_{t+1}|\epsilon_t)\int\cdots\int_{D_{\mathbf{y}^{(t-1)}}} f(\epsilon_t,\boldsymbol{\epsilon}^{(t-1)},\mathbf{y}^{(t-1)})\,d\mathbf{y}^{(t-1)}$$
$$= f(\epsilon_{t+1}|\epsilon_t) f(\epsilon_t,\boldsymbol{\epsilon}^{(t-1)})\,.$$

Il s'en suit que

$$f(\epsilon_{t+1}|\epsilon_t,\boldsymbol{\epsilon}^{(t-1)}) = \frac{f(\epsilon_{t+1},\epsilon_t,\boldsymbol{\epsilon}^{(t-1)})}{f(\epsilon_t,\boldsymbol{\epsilon}^{(t-1)})} = f(\epsilon_{t+1}|\epsilon_t)\,,$$

d'où $\{\epsilon_t\}$ est markovien.

Ce résultat nous permet d'écrire la densité conjointe des états sous une forme simplifiée:

$$f(\epsilon_1, \ldots, \epsilon_n) = f(\epsilon_n|\epsilon_{n-1}, \ldots, \epsilon_1) f(\epsilon_1, \ldots, \epsilon_{n-1}),$$

et puisque $\{\epsilon_t\}$ est markovien,

$$f(\epsilon_1, \ldots, \epsilon_n) = f(\epsilon_n|\epsilon_{n-1}) f(\epsilon_1, \ldots, \epsilon_{n-1}),$$

d'où, en répétant $(n-1)$ fois,

$$f(\epsilon_1, \ldots, \epsilon_n) = \prod_{j=2}^{n} f(\epsilon_j|\epsilon_{j-1}) f(\epsilon_1). \tag{1.1.4}$$

Les équations (1.1.3) et (1.1.4) permettent d'obtenir la densité conjointe des observations $Y_1, \ldots, Y_n$, étant donné les états $\epsilon_1, \ldots, \epsilon_n$:

$$f(y_1, \ldots, y_n|\epsilon_1, \ldots, \epsilon_n) = \prod_{j=1}^{n} f(y_j|\epsilon_j) \tag{1.1.5}$$

On en conclut que $Y_1, \ldots, Y_n$ sont conditionnellement indépendants étant donné $\epsilon_1, \ldots, \epsilon_n$. La suite d'états $\{\epsilon_t\}$ est dite: «*processus latent associé au processus* $\{Y_t\}$ » car la structure de dépendance des observations $\{Y_t\}$ est induite par celle des états $\{\epsilon_t\}$.

Pour les modèles dictés par les observations, l'équation des observations demeure la même:

$$f(y_t|\epsilon_t, \mathbf{y}^{(t-1)}, \boldsymbol{\epsilon}^{(t-1)}) = f(y_t|\epsilon_t), \qquad t = 1, 2, \ldots$$

Quant aux états, on suppose qu'ils sont indépendants des états passés et qu'ils dépendent uniquement des observations passées:

$$f(\epsilon_{t+1}|\epsilon_t, \mathbf{y}^{(t)}, \boldsymbol{\epsilon}^{(t-1)}) = f(\epsilon_{t+1}|\mathbf{y}^{(t)}), \qquad t = 0, 1, \ldots \tag{1.1.6}$$

où $f(\epsilon_1|\mathbf{y}^{(0)}) = f(\epsilon_1)$ est une densité initialement déterminée.

Le plus grand avantage du modèle dicté par les observations est la facilité avec laquelle il est possible d'obtenir une densité pour les prévisions:

$$f(y_{t+1}|\mathbf{y}^{(t)}) = \int f(y_{t+1}|\epsilon_{t+1}) f(\epsilon_{t+1}|\mathbf{y}^{(t)}) d\mu(\epsilon_{t+1})$$

(L'intégrale par rapport à $d\mu(\epsilon_{t+1})$ doit être interprétée comme l'intégrale par rapport à $d\epsilon_{t+1}$ dans le cas continu et comme la somme sur toutes les valeurs de ϵ_{t+1} dans le cas discret). Par contre, les propriétés évolutives de la série chronologique sont beaucoup plus difficiles à caractériser pour ce type de modèle que pour les modèles dictés par les états.

1.2. Modèles dictés par les états pour des séries de dénombrement

Dans cette section, nous rappellerons les principaux éléments d'une classe de modèles dictés par les états, initiée par Zeger (1988) et développée, entres autres, par Blais, MacGibbon et Roy (1997), utilisée pour modéliser une série chronologique à valeurs entières. Cette approche tente de décrire le comportement d'une série chronologique à valeurs entières, $\{Y_t\}$, non stationnaire et ne pouvant être stationnarisée par différenciation. Ce pourrait être, par exemple, une série épidémiologique avec des périodes d'épidémie. L'objectif est d'exprimer $\mu_t = \mathbb{E}[\mathrm{Y_t}]$ comme une fonction d'un vecteur de covariables $\mathbf{X}_t$. Définissons immédiatement la notation utilisée. Soient:

$\mathbf{Y} = (Y_1, \ldots, Y_n)^T$: le vecteur des variables aléatoires d'intérêt,

$\mathbf{y} = (y_1, \ldots, y_n)^T$: le vecteur des observations, une réalisation de $\mathbf{Y}$,

$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$: le vecteur des moyennes des variables aléatoires d'intérêt; chaque μ_t satisfaisant

$$\mu_t = \mathbb{E}[Y_t], \tag{1.2.1}$$

$\mathbf{X}$: une matrice de dimension $n \times p$ de variables explicatives dont la première colonne est formée par le vecteur colonne $\mathbf{1}$; $\mathbf{X}_t$ dénote la t^e ligne de $\mathbf{X}$ et x_{tj} dénote la j^e variable explicative de la t^e observation, et ce, pour $j = 1, 2, ..., p$ et $t = 1, 2, ..., n$,

$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$: le vecteur des paramètres de régression,

$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T = \mathbf{X}\boldsymbol{\beta}$: le vecteur des prédicteurs linéaires; chaque η_t satisfaisant

$$\eta_t = \sum_{j=1}^{p} x_{tj}\beta_j = \mathbf{X}_t\boldsymbol{\beta},$$

g : la fonction de lien canonique associée à la distribution (à venir) de $Y_t|\epsilon_t$ avec fonction inverse h (supposée différentiable) telle que $\eta_t = g(\mu_t)$ et

$$\mu_t = h(\eta_t) = h(\mathbf{X}_t\boldsymbol{\beta}). \tag{1.2.2}$$

Soit finalement $\{\epsilon_t\}$, un processus latent.

Nous ferons l'hypothèse que la distribution de $Y_t|(\epsilon_t, \mathbf{Y}^{(t-1)}, \boldsymbol{\epsilon}^{(t-1)})$ est indépendante de $(\mathbf{Y}^{(t-1)}, \boldsymbol{\epsilon}^{(t-1)})$ et qu'il s'agit d'une distribution discrète de la famille exponentielle dont le ou les paramètres sont choisis de sorte que

$$\mathbb{E}[\, Y_t|\epsilon_t] = \mu_t\epsilon_t\,, \tag{1.2.3}$$

$$\mathrm{Var}(\, Y_t|\epsilon_t) = f(\mu_t, \epsilon_t)\,, \tag{1.2.4}$$

où μ_t satisfait à l'équation (1.2.2). Cette hypothèse correspond à l'équation des observations du modèle dicté par les états. Notons que f est complètement déterminée par le choix de la distribution de $Y_t|\epsilon_t$. Par exemple, un choix classique est la distribution de Poisson, c'est-à-dire:

$$Y_t|\epsilon_t \sim \mathrm{P}(\mu_t\epsilon_t)$$

où $\sim$ signifie "admet pour distribution", pour laquelle la formule de variance serait la même que celle de l'espérance et pour laquelle la fonction g serait la fonction logarithmique.

Pour compléter le modèle dicté par les états, il suffirait de spécifier les propriétés distributionnelles du processus latent $\{\epsilon_t\}$. Ces propriétés distributionnelles constitueraient l'équation des états. Pour le moment, aucune hypothèse ne sera

faite sur la distribution de $Y_t|\epsilon_t$, car plusieurs résultats peuvent être obtenus dans un contexte général. Nous supposerons seulement que $\{\epsilon_t\}$ est un processus stationnaire latent satisfaisant

$$\mathbb{E}[\epsilon_t] = 1 \tag{1.2.5}$$

$$\mathrm{Cov}(\epsilon_t, \epsilon_{t+\tau}) = \sigma^2 \rho_\epsilon(\tau) \tag{1.2.6}$$

où $\rho_\epsilon(\tau)$ dénote la corrélation de délai τ pour le processus $\{\epsilon_t\}$. Notons que la condition $\mathbb{E}[\epsilon_t] = 1$ n'est pas du tout restrictive et est imposée uniquement dans le but de faciliter la présentation du modèle. En effet, si $\mathbb{E}[\epsilon_t] = \mathrm{c} \neq 1$, la constante c peut être absorbée par le terme servant d'ordonnée à l'origine. Il suffit de remplacer ϵ_t par ϵ_t/c et β_1 par $\beta_1 + ln(c)$. Notons également que des conditions supplémentaires peuvent devoir être imposées au domaine de ϵ_t pour assurer la définition des paramètres. Par exemple, dans le cas de la distribution de Poisson, on doit imposer une contrainte de non-négativité à ϵ_t pour que le paramètre de la distribution soit positif.

Finalement, nous supposerons aussi que

$$\mathrm{Cov}(\mathbf{Y}) = \mathbf{V}_\mathrm{n} \tag{1.2.7}$$

où $\mathbf{V}_n$ est une matrice $n \times n$ symétrique définie positive de fonctions connues de μ et des paramètres σ^2 et ϕ (paramètre de pondération associé à l'expression de la distribution de $Y_t|\epsilon_t$ sous la forme de la famille exponentielle) qui ne dépendent pas de $\boldsymbol{\beta}$.

On peut résumer moins formellement le tout en disant que, conditionnellement à ϵ_t, un processus latent stationnaire, Y_t suit un modèle linéaire généralisé. Ainsi, les moments marginaux de Y_t peuvent être exprimés comme une fonction des coefficients de régression et des paramètres du processus $\{\epsilon_t\}$. En effet,

Proposition 1.2.1. *Pour le modèle dicté par les états défini ci-avant, on a :*

i) $\mathbb{E}[Y_t] = \mu_t$,

ii) $\mathrm{Var}(Y_t) = \sigma^2 \mu_t^2 + \mathbb{E}[\mathrm{Var}(Y_t|\epsilon_t)]$,

iii) $\mathrm{Cov}(Y_t, Y_{t+\tau}) = \mu_t \mu_{t+\tau} \sigma^2 \rho_\epsilon(\tau), \tau > 0$.

Preuve

i) Le résultat découle de l'identité $\mathbb{E}[Y_t] = \mathbb{E}[\mathbb{E}[Y_t|\epsilon_t]]$ et des équations (1.2.3) et (1.2.5).

ii) Le résultat découle de l'identité $\mathrm{Var}(Y_t) = \mathrm{Var}(\mathbb{E}[Y_t|\epsilon_t]) + \mathbb{E}[\mathrm{Var}(Y_t|\epsilon_t)]$, de l'équation (1.2.3) et du fait que $\mathrm{Var}(\epsilon_t) = \sigma^2$.

iii) Le résultat découle de l'égalité suivante:

$$\begin{aligned}
\mathbb{E}[Y_t Y_{t+\tau}] &= \mathbb{E}[\mathbb{E}[Y_t Y_{t+\tau}|\epsilon_t \epsilon_{t+\tau}]] \\
&= \mathbb{E}[\mu_t \mu_{t+\tau} \epsilon_t \epsilon_{t+\tau}] \\
&= \mu_t \mu_{t+\tau} \mathbb{E}[\epsilon_t \epsilon_{t+\tau}] \\
&= \mu_t \mu_{t+\tau} (\sigma^2 \rho_\epsilon(\tau) + 1)
\end{aligned}$$

□

Le tableau qui suit présente les variances résultant du modèle pour les distributions discrètes les plus courantes de la famille exponentielle.

TABLEAU 1.2.1. *Variance de la variable dépendante Y_t, résultant du modèle dicté par les états, pour les distributions discrètes les plus courantes de la famille exponentielle de distributions.*

Distribution de $Y_t \vert \epsilon_t$	Paramètre variant avec t	$\mathrm{Var}(Y_t \vert \epsilon_t)$	$\mathrm{Var}(Y_t)$
$\mathrm{P}(\lambda_t)$	$\lambda_t = \mu_t \epsilon_t$	$\mu_t \epsilon_t$	$\mu_t + \sigma^2 \mu_t^2$
$\mathrm{B}(n, p_t)$	$p_t = \frac{\mu_t \epsilon_t}{n}$	$\frac{\mu_t \epsilon_t (n - \mu_t \epsilon_t)}{n}$	$\mu_t (\frac{(n-1)}{n} \sigma^2 - 1)$
$\mathrm{BN}(r, p_t)$	$p_t = \frac{r}{\mu_t \epsilon_t + r}$	$\frac{\mu_t \epsilon_t (\mu_t \epsilon_t + r)}{r}$	$\mu_t + \mu_t^2 (\frac{(r+1)}{r} \sigma^2 + 1)$

Ici, la distribution binomiale négative doit être interprétée comme un compte du nombre d'échecs avant d'obtenir r succès avec une probabilité de succès de p_t à chaque essai. Pour obtenir la distribution géométrique, il suffit de poser $r = 1$ comme premier paramètre de la distribution binomiale négative. Fait intéressant à souligner, ce type de modèle permet d'introduire, dans la majorité des cas, de la sur-dispersion ($\mathrm{Var}(Y_t) > \mathbb{E}[Y_t]$), ce qui est habituellement le cas avec des données réelles.

Pour l'estimation des paramètres de régression, la méthode d'équations d'estimation est utilisée. La fonction de score $U(\theta; y_t)$ est définie comme la dérivée de la fonction de log-vraisemblance l_t:

$$U(\theta; y_t) = \frac{\partial l_t}{\partial \theta}.$$

Sa variance $F_t(\theta)$ satisfait à l'équation

$$F_t(\theta) = \mathrm{Var}\left(\frac{\partial l_t}{\partial \theta}\right) = -\mathbb{E}\left[\frac{\partial^2 l_t}{\partial \theta^2}\right].$$

Si Y_t admet une distribution issue de la famille exponentielle des distributions, sa fonction de densité (masse) peut être écrite sous la forme

$$f_{Y_t}(y_t; \theta, \phi) = \exp\left\{\frac{y_t\theta - b(\theta)}{a(\phi)} + c(y_t, \phi)\right\}$$

où $a(\cdot), b(\cdot)$ et $c(\cdot)$ sont des fonctions mesurables données et ϕ est un paramètre de pondération. Par exemple, si Y_t admet une distribution de Poisson avec paramètre μ,

$$\theta = \log(\mu), b(\theta) = \exp(\theta), \phi = 1, a(\phi) = \phi, c(y_t, \phi) = -\log(y_t!).$$

Dans le cas général, la fonction de log-vraisemblance a la forme suivante:

$$l_t = \left\{\frac{y_t\theta - b(\theta)}{a(\phi)} + c(y_t, \phi)\right\}.$$

On a donc:

$$\frac{\partial l_t}{\partial \theta} = \frac{y_t - b'(\theta)}{a(\phi)}\,,$$
$$\frac{\partial^2 l_t}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)}\,,$$

où $(')$ dénote une dérivée par rapport à θ.

Il s'en suit que, dans le cas de la famille exponentielle, la fonction de score et sa variance satisfont aux équations:

$$U(\theta; y_t) = \frac{y_t - b'(\theta)}{a(\phi)}\,,$$
$$F_t(\theta) = \frac{b''(\theta)}{a(\phi)}\,.$$

Notons aussi que les deux équations suivantes:

$$\mathbb{E}\left[\frac{\partial l_t}{\partial \theta}\right] = 0\,,$$
$$\mathbb{E}\left[\frac{\partial^2 l_t}{\partial \theta^2}\right] + \mathbb{E}\left[\left(\frac{\partial l_t}{\partial \theta}\right)^2\right] = 0\,,$$

résultat de la différenciation de l'identité $\int f_{Y_t}(y_t; \theta, \phi)\, dy_t = 1$ par rapport à θ, impliquent que

$$\mathbb{E}[Y_t] = \mu_t = b'(\theta)\,,$$
$$\mathrm{Var}(Y_t) = b''(\theta) a(\phi)\,.$$

Bradley (1973) et Wedderburn (1974) ont soulevé le fait que, dans le cas où Y_t a une distribution de la famille exponentielle, la fonction de score ne dépend des paramètres θ et ϕ que via la moyenne et la variance de Y_t. Wedderburn a suggéré d'utiliser la fonction de score propre à la famille exponentielle même lorsque la distribution sous-jacente est inconnue. Dans un tel cas, la fonction d'estimation prend le nom de fonction de quasi-score et l'estimateur relié, celui d'estimateur de quasi-vraisemblance. Dans le cas général, on définit le quasi-score, un vecteur de dimension p, par

$$\mathbf{U}_n(\boldsymbol{\beta}) = \mathbf{D}_n^T \mathbf{V}_n^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \tag{1.2.8}$$

où $\mathbf{D}_n$ est une matrice $n \times p$ dont le $(t,j)^e$ élément est donné par $\partial \mu_t / \partial \beta_j$ (Voir McCullagh et Nelder, 1989, chapitre 9). L'équation à résoudre est donc:

$$\mathbf{U}_n(\hat{\boldsymbol{\beta}}) = 0 \tag{1.2.9}$$

et $\hat{\boldsymbol{\beta}}$, une racine des équations d'estimation, est l'estimateur de quasi-vraisemblance. En pratique, si $\mathbf{U}_n$ dépend de ϕ, un vecteur de paramètres indépendants de $\boldsymbol{\beta}$, des estimateurs convergents de ϕ sont alors substitués dans (1.2.8). Le vecteur $\mathbf{U}_n(\boldsymbol{\beta})$ satisfait aux équations

$$\begin{aligned} \mathbb{E}[\mathbf{U}_n(\boldsymbol{\beta})] &= 0 \\ \mathrm{Cov}(\mathbf{U}_n(\boldsymbol{\beta})) &= \mathbf{F}_{\beta,n} \\ -\mathbb{E}\left[\frac{\partial \mathbf{U}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right] &= \mathbf{F}_{\beta,n} \end{aligned}$$

où $\mathbf{F}_{\beta,n} = \mathbf{D}_n^T \mathbf{V}_n^{-1} \mathbf{D}_n$.

Cette matrice joue le même rôle que la matrice d'information de Fisher dans l'inférence à vraisemblance maximale.

L'équation (1.2.9) est, en général, non linéaire en $\hat{\boldsymbol{\beta}}$ et le calcul de $\hat{\boldsymbol{\beta}}$ doit se faire de façon itérative. Dans ce mémoire, nous utilisons l'algorithme décrit au chapitre 9 de McCullagh et Nelder (1989). Il s'agit d'un algorithme de type Newton-Raphson. En débutant les itérations en une valeur arbitraire $\hat{\boldsymbol{\beta}}_0$, suffisamment près de $\hat{\boldsymbol{\beta}}$, la suite des paramètres estimés générée par l'algorithme est donnée par

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} + ((\hat{\mathbf{D}}^{(k)})^T(\hat{\mathbf{V}}^{(k)})^{-1}\hat{\mathbf{D}}^{(k)})^{-1}(\hat{\mathbf{D}}^{(k)})^T(\hat{\mathbf{V}}^{(k)})^{-1}(\mathbf{Y} - \hat{\boldsymbol{\mu}}^{(k)})$$

où $\cdot^{(k)}$ dénote l'estimation résultant de la k^e itération. Le détail de l'estimation ne sera pas repris ici. Le lecteur est référé à Blais (1996) pour davantage de précision. Mentionnons simplement que la méthode des moments est utilisée afin d'estimer les paramètres de nuisance σ^2 et ρ_ϵ, à partir des relations décrites dans la proposition 1.2.1. Il en résulte les estimateurs suivants:

$$\hat{\mu}_t^{(k)} = exp(\mathbf{X}_t\hat{\boldsymbol{\beta}}^{(k)})$$

$$\hat{\mathbf{D}}^{(k)} = (\hat{\mathbf{D}}_{tj}^{(k)}), \text{où } \hat{\mathbf{D}}_{tj}^{(k)} = x_{tj}\hat{\mu}_t^{(k)}$$

$$\hat{\sigma^2}^{(k)} = \frac{\sum_{t=1}^{n}(y_t - \hat{\mu}_t^{(k)})^2 - \hat{\mu}_t^{(k)}}{\sum_{t=1}^{n}(\hat{\mu}_t^{(k)})^2}$$

$$\hat{\rho}_\epsilon^{(k)} = \frac{\sum_{t=1}^{n-1}(y_t - \hat{\mu}_t^{(k)})(y_{t+1} - \hat{\mu}_{t+1}^{(k)})}{\hat{\sigma^2}^{(k)}\sum_{t=1}^{n-1}\hat{\mu}_t^{(k)}\hat{\mu}_{t+1}^{(k)}}$$

$$\hat{\mathbf{V}}^{(k)} = (\hat{v}_{st}^{(k)})_{n\times n},$$

où

$$\hat{v}_{st}^{(k)} = \begin{cases} \hat{\mu}_t^{(k)} + \hat{\sigma^2}^{(k)}(\hat{\mu}_t^{(k)})^2 & , \quad s = t\,, \\ \hat{\mu}_s^{(k)}\hat{\mu}_t^{(k)}\hat{\sigma^2}^{(k)}(\hat{\rho}_\epsilon^{(k)})^{|s-t|} & , \quad s \neq t\,. \end{cases}$$

L'estimation de quasi-vraisemblance $\hat{\boldsymbol{\beta}}$ est obtenue par itération jusqu'à ce qu'il y ait convergence et les estimations des paramètres de nuisance sont obtenues en utilisant les formules de la page précédente et l'estimation à quasi-vraisemblance.

1.3. Distribution asymptotique et inférence

Sous certaines conditions, il est possible de démontrer la normalité asymptotique de l'estimateur à quasi-vraisemblance. Pour le détail des conditions, le lecteur est référé à Blais, MacGibbon et Roy (1997). Il est donc possible de faire de l'inférence sur les paramètres de régression, mais également sur l'espérance μ_t. En effet, si on suppose que

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{D} \mathcal{N}_p(0, \mathbf{B}^{-1})$$

alors

$$\sqrt{n}(\mathbf{X}_t\hat{\boldsymbol{\beta}}_n - \mathbf{X}_t\boldsymbol{\beta}) \xrightarrow{D} \mathcal{N}(0, \mathbf{X}_t\mathbf{B}^{-1}\mathbf{X}_t^T)$$

où D signifie "converge en distribution". Donc, un intervalle de confiance de niveau asymptotique $(1-\alpha)\times 100\%$ pour $\mathbf{X}_t\boldsymbol{\beta}$ est donné par:

$$\left(\mathbf{X}_t\hat{\boldsymbol{\beta}}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{X}_t\hat{\mathbf{B}}^{-1}\mathbf{X}_t^T}{n}}\,,\ \mathbf{X}_t\hat{\boldsymbol{\beta}}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{X}_t\hat{\mathbf{B}}^{-1}\mathbf{X}_t^T}{n}}\right)$$

où $\hat{\mathbf{B}}$ est un estimateur convergent de $\mathbf{B}$ et où n est supposé grand. Puisque h est continue (car elle est différentiable) et bijective (car elle possède une inverse g), h est monotone. Par conséquent, un intervalle de confiance de niveau asymptotique $(1-\alpha)\times 100\%$ pour μ_t est donné par

$$\left(h\left(\mathbf{X}_t\hat{\boldsymbol{\beta}}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{X}_t\hat{\mathbf{B}}^{-1}\mathbf{X}_t^T}{n}}\right)\,,\ h\left(\mathbf{X}_t\hat{\boldsymbol{\beta}}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{X}_t\hat{\mathbf{B}}^{-1}\mathbf{X}_t^T}{n}}\right)\right)$$

si h est monotone croissante et par

$$\left(h\left(\mathbf{X}_t\hat{\boldsymbol{\beta}}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{X}_t\hat{\mathbf{B}}^{-1}\mathbf{X}_t^T}{n}} \right) , h\left(\mathbf{X}_t\hat{\boldsymbol{\beta}}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{X}_t\hat{\mathbf{B}}^{-1}\mathbf{X}_t^T}{n}} \right) \right)$$

si h est monotone décroissante.

Mais l'intérêt d'établir de tels intervalles de confiance est toutefois moins capital qu'à première vue. En effet, si $\mathbf{X}$ est constitué de variables explicatives déterministes, la prévision d'horizon l est donnée par:

$$Y_n(l) = \hat{Y}_{n+l} = \hat{\mu}_{n+l} = h(\mathbf{X}_{n+l}\hat{\boldsymbol{\beta}}) .$$

L'intervalle établi ci-avant constitue donc un intervalle de prévision pour la moyenne $\mathbb{E}[Y_{n+l}]$ et non pas pour la valeur future Y_{n+l}. Pour mieux comprendre la distinction, notons simplement qu'un intervalle de prévision de niveau $1-\alpha$ pour la moyenne est défini par des bornes I_1 et S_1 telles que

$$P[I_1 < \mu_{n+l} = \mathbb{E}[Y_{n+l}] < S_1] = 1 - \alpha ,$$

alors qu'un intervalle de prévision de niveau $1-\alpha$ pour la valeur future Y_{n+l} est défini par des bornes I_2 et S_2 telles que

$$P[I_2 < Y_{n+l} < S_2] = 1 - \alpha .$$

Or, en séries chronologiques, ce sont les intervalles de prévision des valeurs futures qui ont un intérêt particulier. Il est donc nécessaire d'entrevoir une autre solution.

Chapitre 2

APPROCHE MODIFIÉE

La normalité asymptotique de l'estimateur à quasi-vraisemblance, démontrée par Blais, MacGibbon et Roy (1997), est un résultat majeur. La section 1.3 a toutefois permis de constater que ce résultat ne permet pas d'établir des intervalles de prévision pour les valeurs futures. Le présent chapitre propose deux solutions possibles pour construire de tels intervalles. La première solution est une méthode paramétrique basée sur la spécification des propriétés distributionnelles du processus latent. La deuxième méthode propose plutôt une solution non paramétrique basée sur l'utilisation d'une transformation T qui stabilise la variance de la série et sur l'utilisation des différences $\delta_t = T(y_t) - T(\hat{\mu}_t)$.

2.1. SOLUTION PARAMÉTRIQUE

Jusqu'à présent, les propriétés distributionnelles du processus latent n'ont pas été spécifiées. Nous avons seulement supposé que le processus latent $\{\epsilon_t\}$ était stationnaire et satisfaisait aux équations (1.2.5) et (1.2.6). Dans le cadre de cette solution, nous supposerons plutôt que $\{\epsilon_t\}$ est un processus stationnaire latent dont la distribution marginale de ϵ_t est une distribution conjuguée naturelle de la distribution de $Y_t|\epsilon_t$, respectant les équations (1.2.5) et (1.2.6). Notons que tous les résultats des sections 1.2 et 1.3 demeurent vrais car les hypothèses précédentes sont toujours présentes.

La distribution de $Y_t|\epsilon_t$, la distribution marginale de ϵ_t, conjuguée naturelle de la précédente, et la distribution résultante pour Y_t sont présentées dans le tableau qui suit.

TABLEAU 2.1.1. *Distribution résultante pour la variable Y_t, selon le modèle dicté par les états, pour les distributions discrètes les plus courantes de la famille exponentielle de distributions.*

Distribution de $Y_t\|\epsilon_t$	Distribution de ϵ_t	Fonction de masse pour Y_t
$\mathrm{P}(\mu_t\epsilon_t)$	$\mathrm{Gamma}(a\,,b)$	$\frac{\Gamma(a+y_t)}{\Gamma(a)y_t!}\left(\frac{b}{b+\mu_t}\right)^a\left(\frac{\mu_t}{b+\mu_t}\right)^{y_t}$ $y_t=0,1,\ldots$
$\mathrm{B}(n\,,\frac{\mu_t\epsilon_t}{n})$	$\mathrm{Bêta}(a\,,b)$	$\binom{n}{y_t}\left(\frac{\mu_t}{n}\right)^{y_t}\frac{\Gamma(a+b)\Gamma(a+y_t)}{\Gamma(a)\Gamma(a+b+y_t)}$ $\times F(y_t-n,a+y_t,a+b+y_t;\frac{\mu_t}{n})$ $y_t=0,1,\ldots$
$\mathrm{BN}(r\,,\frac{r}{\mu_t\epsilon_t+r})$	$\mathrm{Bêta}(a\,,b)$	$\binom{r+y_t-1}{r-1}\left(\frac{\mu_t}{r}\right)^{y_t}\frac{\Gamma(a+b)\Gamma(a+y_t)}{\Gamma(a)\Gamma(a+b+y_t)}$ $\times F(r+y_t,a+y_t,a+b+y_t;\frac{-\mu_t}{r})$ $y_t=0,1,\ldots$

Dans ce dernier tableau, $\Gamma(\cdot)$ désigne la fonction gamma usuelle et la fonction F est la fonction hypergéométrique définie par:

$$F(a,b,c;x)=\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 u^{b-1}(1-u)^{c-b-1}(1-ux)^{-a}\,du\,.$$

Les principales propriétés de cette fonction peuvent être trouvées dans Spiegel (1989). Par la distribution $\mathrm{Gamma}(a,b)$, nous voulons désigner la distribution dont la fonction de densité est définie par

$$f(\epsilon_t)=\frac{b^a}{\Gamma(a)}\epsilon_t^{a-1}e^{-b\epsilon_t}\,,\qquad \epsilon_t\geq 0\,,$$

alors que par la distribution Bêta(a, b), nous voulons désigner la distribution dont la fonction de densité est définie par:

$$f(\epsilon_t) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \epsilon_t^{a-1}(1-\epsilon_t)^{b-1}\,, \qquad 0 \leq \epsilon_t \leq 1\,.$$

Nous démontrerons le premier résultat, les deux autres résultats pouvant être obtenus de la même façon.

$$\begin{aligned} f_{Y_t}(y_t) &= \int_0^\infty f_{Y_t,\varepsilon_t}(y_t, \epsilon_t)\, d\epsilon_t \\ &= \int_0^\infty f_{Y_t|\varepsilon_t}(y_t|\epsilon_t) f_{\varepsilon_t}(\epsilon_t)\, d\epsilon_t\,, \end{aligned}$$

et puisque $Y_t|\epsilon_t \sim \mathrm{P}(\mu_t\epsilon_t)$ et que $\epsilon_t \sim \mathrm{Gamma}(a, b)$,

$$\begin{aligned} f_{Y_t}(y_t) &= \int_0^\infty \frac{(\mu_t\epsilon_t)^{y_t}}{y_t!}\, e^{-\mu_t\epsilon_t}\, \frac{b^a}{\Gamma(a)}\, \epsilon_t^{a-1}\, e^{-b\epsilon_t}\, d\epsilon_t \\ &= \frac{b^a \mu_t^{y_t}}{\Gamma(a) y_t!} \int_0^\infty \epsilon_t^{a+y_t-1}\, e^{-(b+\mu_t)\epsilon_t}\, d\epsilon_t \\ &= \frac{\Gamma(a+y_t)}{\Gamma(a) y_t!} \left(\frac{b}{b+\mu_t}\right)^a \left(\frac{\mu_t}{b+\mu_t}\right)^{y_t}. \end{aligned}$$

Le tableau précédent présente les résultats sous leur forme la plus générale. Voici quelques remarques plus spécifiques s'y rapportant:

i) Dans le cas où $\epsilon_t \sim \mathrm{Gamma}(a, b)$ avec $a \in \mathbb{N}$, la distribution de Y_t est binomiale négative de paramètres a (pour le nombre de succès) et $b/(b+\mu_t)$ (pour la probabilité de succès).

ii) L'hypothèse $\mathbb{E}[\epsilon_t] = 1$ impose des conditions sur les paramètres. Dans le cas où $\epsilon_t \sim \mathrm{Gamma}(a, b)$, cela entraîne que $a = b$. Pour $\epsilon_t \sim$ Bêta(a, b), cela entraînerait que $b = 0$; alors, il est préférable d'imposer $\mathbb{E}[\epsilon_t] = \mathrm{c} \neq 1$ et

d'utiliser ϵ_t/c tel que mentionné précédemment. La distribution résultante pour Y_t, de même que celle du processus latent $\{\epsilon_t/c\}$, doivent alors être déterminées de nouveau.

Cette hypothèse est très utile à la construction d'intervalles de prévision. En effet, fort de cette hypothèse, il est possible de déterminer la distribution résultante pour Y_{n+l}. Or, cette distribution ne dépend que des paramètres de la loi marginale de ϵ_t et de μ_t, des quantités pouvant être estimées uniquement à partir des observations $y_1, \ldots, y_n$. En autant que les variables explicatives soient déterministes, on peut estimer la distribution de Y_{n+l} en remplaçant a, b et μ_{n+l} par leurs estimations respectives. Une fois l'estimation réalisée, il est donc possible de construire un intervalle de prévision pour Y_{n+l} par simulation ou de façon théorique selon la distribution estimée.

Supposons, pour fin d'illustration, que l'on utilise le modèle présenté à la section 1.2 pour décrire $\{Y_t\}_{t=1}^n$ et où l'on suppose que $Y_t|\epsilon_t \sim \mathrm{P}(\mu_t\epsilon_t)$. La méthode paramétrique suggère les étapes suivantes afin de construire un intervalle de prévision pour Y_{n+l}:

i) Utiliser l'algorithme présenté pour obtenir $\hat{\boldsymbol{\beta}}$, l'estimateur de $\boldsymbol{\beta}$.

ii) Supposer que $\epsilon_t \sim \mathrm{Gamma}(a, b)$.

iii) Estimer les paramètres a et b. Puisque le modèle suppose que $\mathbb{E}[\epsilon_t] = 1$, on imposera la contrainte $\hat{a} = \hat{b}$. Il en résulte aussi que $\mathrm{Var}(\epsilon_t) = \sigma^2 = 1/b$. La méthode des moments suggère de prendre $\hat{b} = 1/\hat{\sigma^2}$, où $\hat{\sigma^2}$ est l'estimation de la variance de ϵ_t, fournie par l'algorithme d'estimation.

iv) Construire l'intervalle de prévision désiré en utilisant la distribution estimée de Y_{n+l}, soit:

$$\hat{f}_{Y_{n+l}}(y) = \frac{\Gamma(\hat{a}+y)}{\Gamma(\hat{a})y!}\left(\frac{\hat{b}}{\hat{b}+\hat{\mu}_{n+l}}\right)^{\hat{a}}\left(\frac{\hat{\mu}_{n+l}}{\hat{b}+\hat{\mu}_{n+l}}\right)^{y}.$$

Il peut sembler un peu curieux de postuler une distribution pour un processus latent, donc pour un phénomène non observé. L'hypothèse semble, au premier coup d'oeil, très restrictive, voire grossière. Toutefois, cette hypothèse est nettement moins restrictive qu'à première vue. En effet, la distribution spécifiée est la distribution marginale de ϵ_t. Les états $\epsilon_1, \ldots, \epsilon_n$ demeurent donc conjointement dépendants, et à plus forte raison, les observations $Y_1, \ldots, Y_n$. De plus, les nombreuses contraintes imposées par le modèle ont pour effet d'uniformiser l'allure de la distribution de ϵ_t. Il s'agit donc, principalement, d'une contrainte de définition, car elle impose un domaine aux valeurs admissibles pour les états. L'impact de cette hypothèse sur les taux de couverture des intervalles de prévision, lorsqu'elle est vérifiée ou non, sera étudié un peu plus tard. Pour l'instant, mentionnons seulement qu'il s'agit d'une première méthode de construction d'intervalles de prévision.

2.2. Solution non paramétrique

Contrairement à la méthode paramétrique, la solution non paramétrique ne spécifie aucune distribution pour le processus latent. Elle est plutôt basée sur la relation entre $\mathbb{E}[Y_t]$ et $\mathrm{Var}(Y_t)$.

En séries chronologiques, il est commun d'utiliser une transformation pour stabiliser la variance lorsque celle-ci varie selon le niveau, c'est-à-dire selon la moyenne (Voir Wei (1989), section 4.3.2). De façon plus explicite, si $\mu_t = \mathbb{E}[Y_t]$ et si $\mathrm{Var}(Y_t) = cf(\mu_t)$, où c est une constante et f est une fonction intégrable, la transformation qui stabilise la variance de Y_t est donnée par:

$$T(y_t) = \int \frac{1}{\sqrt{f(y_t)}}\, dy_t\,.$$

Le tableau qui suit présente les transformations qui stabilisent la variance résultant du modèle dicté par les états pour les distributions les plus courantes de la famille exponentielle.

Tableau 2.2.1. *Transformation stabilisant la variance pour la variable Y_t, selon le modèle dicté par les états, pour les distributions discrètes les plus courantes de la famille exponentielle de distributions.*

Distribution de $Y_t \vert \epsilon_t$	Variance de Y_t	$T(Y_t)$
$\mathrm{P}(\mu_t \epsilon_t)$	$\mu_t + \sigma^2 \mu_t^2$	$\mathrm{arccosh}(2\sigma^2 Y_t + 1)$
$\mathrm{B}(n\,, \frac{\mu_t \epsilon_t}{n})$	$\mu_t(\frac{n-1}{n}\sigma^2 - 1)$	$2\sqrt{Y_t}$
$\mathrm{BN}(r\,, \frac{r}{\mu_t \epsilon_t + r})$	$\mu_t + \mu_t^2(\frac{r+1}{r}\sigma^2 + 1)$	$\mathrm{arccosh}(2(\frac{r+1}{r}\sigma^2 + 1)Y_t + 1)$

Nous démontrerons encore une fois seulement le premier résultat, les deux autres résultats pouvant être obtenus de la même façon. Puisque

$$\mathrm{Var}(Y_t) = \mu_t + \sigma^2 \mu_t^2 = \sigma^2 f(\mu_t)\,,$$

où $f(\mu_t) = \mu_t^2 + \frac{\mu_t}{\sigma^2}$, la transformation à effectuer pour stabiliser la variance est donnée par:

$$\begin{aligned} T(y_t) &= \int \frac{1}{\sqrt{y_t^2 + \frac{y_t}{\sigma^2}}}\, dy_t \\ &= \int \frac{1}{\sqrt{(y_t + \frac{1}{2\sigma^2})^2 - (\frac{1}{2\sigma^2})^2}}\, dy_t \end{aligned}$$

en complétant le carré. Et puisque $\int \frac{1}{\sqrt{z^2 - a^2}}\, dz = \mathrm{arccosh}(\frac{z}{a})$,

$$\begin{aligned} T(y_t) &= \mathrm{arccosh}\left(\frac{(y_t + \frac{1}{2\sigma^2})}{\frac{1}{2\sigma^2}}\right) \\ &= \mathrm{arccosh}(2\sigma^2 y_t + 1) \end{aligned}$$

Notons que $T(Y_t)$ est bien défini, et ce pour toutes les distributions considérées, car $Y_t \geq 0$. On définira évidemment l'arccosinus hyperbolique et la racine carrée par leur branche positive respective, tel qu'habituellement. Finalement, mentionnons qu'il existe une équivalence entre l'arccosinus hyperbolique et le logarithme naturel donnée par:

$$\mathrm{arccosh}(z) = ln(z + \sqrt{z^2 - 1})\,.$$

Maintenant, si on développe $T(Y_t)$ en série de Taylor d'ordre 1 autour de μ_t, c'est-à-dire,

$$T(Y_t) \approx T(\mu_t) + T'(\mu_t)(Y_t - \mu_t)\,,$$

on peut alors remarquer, dans un premier temps, que

$$\mathbb{E}[T(Y_t) - T(\mu_t)] \approx 0\,,$$

et, dans un deuxième temps, que

$$\begin{aligned}\mathrm{Var}(T(Y_t) - T(\mu_t)) &\approx (T'(\mu_t))^2 \mathrm{Var}(Y_t) \\ &= \textit{constante}\end{aligned}$$

car $T(Y_t)$ est la transformation qui stabilise la variance. De plus, on peut également remarquer que la corrélation entre $T(Y_t) - T(\mu_t)$ et $T(Y_{t+k}) - T(\mu_{t+k})$ est approximativement la même que celle entre Y_t et Y_{t+k}. En effet,

$$\begin{aligned}&\mathrm{Cor}(T(Y_t) - T(\mu_t), T(Y_{t+k}) - T(\mu_{t+k})) \\ &= \frac{\mathrm{Cov}(T(Y_t) - T(\mu_t), T(Y_{t+k}) - T(\mu_{t+k}))}{\sqrt{\mathrm{Var}((T(Y_t) - T(\mu_t))\mathrm{Var}(T(Y_{t+k}) - T(\mu_{t+k}))}} \\ &\approx \frac{T'(\mu_t)T'(\mu_{t+k})\mathrm{Cov}(Y_t - \mu_t, Y_{t+k} - \mu_{t+k})}{\sqrt{(T'(\mu_t))^2(T'(\mu_{t+k}))^2\mathrm{Var}(Y_t - \mu_t)\mathrm{Var}(Y_{t+k} - \mu_{t+k})}} \\ &= \mathrm{Cor}(Y_t, Y_{t+k})\,. \qquad (2.2.1)\end{aligned}$$

Donc, $\{T(Y_t) - T(\mu_t)\}$ n'est pas stationnaire en covariance, car $\{Y_t\}$ ne l'est pas. Toutefois, $\{T(Y_t) - T(\mu_t)\}$ est approximativement stationnaire en moyenne et en variance. Puisque $\hat{\boldsymbol{\beta}}$ est un estimateur convergent pour $\boldsymbol{\beta}$, si on définit δ_t par

$$\delta_t = T(Y_t) - T(\hat{\mu}_t) = T(Y_t) - T(h(\mathbf{X}_t\hat{\boldsymbol{\beta}}))\,,$$

δ_t est convergent pour $T(Y_t) - T(\mu_t)$ et $\{\delta_t\}$ devrait être, lui aussi, approximativement stationnaire en moyenne et en variance. Dans la mesure où T possède un inverse T^{-1}, la méthode non paramétrique suggère les étapes suivantes pour construire un intervalle de prévision pour Y_{n+l}:

i) Déterminer $\hat{\boldsymbol{\beta}}$, l'estimateur de $\boldsymbol{\beta}$.
ii) Calculer $\delta_t = T(y_t) - T(\hat{\mu}_t)$ pour $t = 1, \ldots, n$.
iii) Déterminer les quantiles nécessaires à l'établissement de l'intervalle de prévision désiré, par exemple $q_{\delta;0.05}$ et $q_{\delta;0.95}$ dans le cas d'un intervalle bilatéral de niveau de confiance 90%.
iv) Construire l'intervalle de prévision à partir des quantiles, par exemple

$$Y_{n+l} \in \left(T^{-1}(q_{\delta;0.05} + T(\hat{\mu}_{n+l}))\,,\; T^{-1}(q_{\delta;0.95} + T(\hat{\mu}_{n+l}))\right).$$

Si on suppose que les $T(Y_t) - T(\mu_t)$ sont identiquement distribués, l'intervalle ainsi construit devrait être de niveau approximatif 90%. En effet,

$$
\begin{aligned}
& P[Y_{n+l} \in \left(T^{-1}(q_{\delta;0.05} + T(\hat{\mu}_{n+l}))\,,\, T^{-1}(q_{\delta;0.95} + T(\hat{\mu}_{n+l}))\right)] \\
& = P[T(Y_{n+l}) - T(\hat{\mu}_{n+l}) \in (q_{\delta;0.05}\,,\, q_{\delta;0.95})] \\
& \approx 90\%\,,
\end{aligned}
$$

car les $\delta_t = T(y_t) - T(\hat{\mu}_t)$ sont alors approximativement identiquement distribués.

Remarques:

i) Il s'agit d'une méthode heuristique. À première vue, rien ne permet d'affirmer que les $T(Y_t) - T(\mu_t)$ sont identiquement distribués. De plus, il est clair qu'ils sont dépendants car les Y_t sont corrélés. Les quantiles $q_{\delta;0.05}$ et $q_{\delta;0.95}$ pourraient donc ne pas correspondre aux quantiles de δ_{n+l}.

ii) Etant donné la forme parabolique des fonctions inverses des fonctions arccosinus hyperbolique et racine carrée, il est préférable de redéfinir ces fonctions inverses par:

$$
T^{-1}(x) = \begin{cases} T^{-1}(x) & ,\ si\ x > 0\,, \\ T^{-1}(0) & ,\ si\ x \leq 0. \end{cases}
$$

Chapitre 3

SIMULATION DE PROCESSUS AR(1) NON GAUSSIEN

Afin de vérifier l'adéquation de la méthode paramétrique par des simulations, des processus stationnaires avec une distribution marginale fixée doivent être construits. Dans les sections qui suivent, nous présentons deux tels processus. Le premier processus, élaboré par Sim (1986), admet une distribution marginale gamma, alors que le deuxième, développé par McKenzie (1985), admet une distribution marginale bêta.

3.1. Simulation d'un processus gamma

Le processus AR(1) gamma est construit à partir de l'équation aux différences stochastiques suivante:

$$\epsilon_n = V_n \epsilon_{n-1} + a_n \,, \qquad n \geq 1 \,, \tag{3.1.1}$$

où $\{V_n\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) selon la fonction de répartition $F_V(v) = v^\alpha$ $(\alpha \geq 0)$ définie sur l'intervalle $[0, 1)$ et où $\{a_n\}$ est une suite de variables aléatoires i.i.d. exponentielle

de paramètre λ. On suppose aussi que les suites $\{V_n\}$ et $\{a_n\}$ sont mutuellement indépendantes. On suppose finalement que $\epsilon_0 \sim \text{Gamma}(\alpha+1, \lambda)$.

Théorème 3.1.1.

Le processus $\{\epsilon_n\}$ *défini par (3.1.1) est un processus stationnaire au sens large qui admet une distribution marginale* $\text{Gamma}(\alpha+1, \lambda)$ *avec une structure de corrélation donnée par* $\rho_j = \text{Cor}(\epsilon_t, \epsilon_{t+j}) = \left(\frac{\alpha}{\alpha+1}\right)^j$, $j \geq 1$.

Preuve

Mentionnons d'abord que la preuve ci-dessous est différente de celle proposée par Sim (1986) et constitue donc une contribution originale. Elle est plus directe et plus simple. Soit la proposition suivante:

$P(n)$: ϵ_n *admet une distribution marginale* $\text{Gamma}(\alpha+1, \lambda)$.

Nous allons vérifier par induction que la proposition $P(n)$ est vraie pour tout $n \in \mathbb{N}$. Pour $n = 0$, $P(n)$ est vraie par la dernière hypothèse du modèle. On suppose maintenant que $P(n-1)$ est vraie. On doit montrer que cela implique que $P(n)$ est également vraie alors. Or,

$$\begin{aligned} F_{\epsilon_n}(z) &= P[\epsilon_n \leq z] \\ &= P[V_n \epsilon_{n-1} + a_n \leq z]. \end{aligned}$$

En remarquant la non-négativité des trois variables impliquées, on constate que a_n est limitée à l'intervalle $[0, z]$. Une fois a_n fixée, on peut développer cette

probabilité selon deux domaines réguliers. On obtient alors, par l'indépendance mutuelle de V_n , a_n et ϵ_{n-1},

$$F_{\epsilon_n}(z) = \int_0^z \int_0^{z-a} \int_0^1 \alpha v^{\alpha-1} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \epsilon^\alpha e^{-\lambda\epsilon} \lambda e^{-\lambda a} \, dv \, d\epsilon \, da$$
$$+ \int_0^z \int_{z-a}^\infty \int_0^{\frac{z-a}{\epsilon}} \alpha v^{\alpha-1} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \epsilon^\alpha e^{-\lambda\epsilon} \lambda e^{-\lambda a} \, dv \, d\epsilon \, da \, .$$

Puis, en intégrant par rapport à v, on obtient:

$$F_{\epsilon_n}(z) = \int_0^z \int_0^{z-a} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \epsilon^\alpha e^{-\lambda\epsilon} \lambda e^{-\lambda a} \, d\epsilon \, da$$
$$+ \int_0^z \int_{z-a}^\infty (z-a)^\alpha \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} e^{-\lambda\epsilon} \lambda e^{-\lambda a} \, d\epsilon \, da \, .$$

En inversant l'ordre d'intégration de la première intégrale double, on a:

$$F_{\epsilon_n}(z) = \int_0^z \int_0^{z-\epsilon} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \epsilon^\alpha e^{-\lambda\epsilon} \lambda e^{-\lambda a} \, da \, d\epsilon$$
$$+ \int_0^z \int_{z-a}^\infty (z-a)^\alpha \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} e^{-\lambda\epsilon} \lambda e^{-\lambda a} \, d\epsilon \, da \, .$$

En intégrant les deux intégrales doubles, respectivement par rapport à a et à ϵ, on trouve:

$$F_{\epsilon_n}(z) = \int_0^z \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \epsilon^\alpha e^{-\lambda\epsilon}(1 - e^{-\lambda(z-\epsilon)})\, d\epsilon + \int_0^z (z-a)^\alpha \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \frac{e^{-\lambda(z-a)}}{\lambda} \lambda e^{-\lambda a}\, da\,,$$

ce qui se réduit à

$$F_{\epsilon_n}(z) = \int_0^z \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}\, \epsilon^\alpha\, e^{-\lambda\epsilon}\, d\epsilon - \int_0^z \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}\, \epsilon^\alpha\, e^{-\lambda z}\, d\epsilon + \int_0^z (z-a)^\alpha \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} e^{-\lambda z}\, da\,.$$

En effectuant le changement de variable $u = z - a$ dans la troisième intégrale, on obtient:

$$F_{\epsilon_n}(z) = \int_0^z \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}\, \epsilon^\alpha\, e^{-\lambda\epsilon}\, d\epsilon - \int_0^z \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \epsilon^\alpha e^{-\lambda z}\, d\epsilon + \int_0^z \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} u^\alpha e^{-\lambda z}\, du\,,$$

d'où,

$$F_{\epsilon_n}(z) = \int_0^z \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}\, \epsilon^\alpha\, e^{-\lambda\epsilon}\, d\epsilon\,,$$

c'est-à-dire, $\epsilon_n \sim \text{Gamma}(\alpha+1, \lambda)$. Donc, le processus $\{\epsilon_n\}$ admet bien une distribution marginale Gamma$(\alpha+1, \lambda)$.

De plus, pour $j \geq 1$, on a

$$\begin{aligned} \mathrm{Cov}(\epsilon_n, \epsilon_{n-j}) &= \mathrm{Cov}(V_n \epsilon_{n-1} + a_n, \epsilon_{n-j}) \\ &= \mathrm{Cov}(V_n \epsilon_{n-1}, \epsilon_{n-j}), \end{aligned}$$

car ϵ_n est une fonction seulement de $a_n, a_{n-1}, a_{n-2}, \ldots$ Puisque V_n est indépendant à la fois de ϵ_{n-1} et de ϵ_{n-j}, il s'en suit que:

$$\begin{aligned} \mathrm{Cov}(\epsilon_n, \epsilon_{n-j}) &= \mathbb{E}[V_n]\mathrm{Cov}(\epsilon_{n-1}, \epsilon_{n-j}) \\ &= \left(\frac{\alpha}{\alpha+1}\right) \mathrm{Cov}(\epsilon_{n-1}, \epsilon_{n-j}), \end{aligned}$$

d'où $\rho_j = \left(\frac{\alpha}{\alpha+1}\right)^j$.

Finalement, puisque les deux premiers moments de ϵ_n sont indépendants de n, car $\epsilon_n \sim \mathrm{Gamma}(\alpha+1, \lambda)$, et puisque $\mathrm{Cor}(\epsilon_r, \epsilon_s) = f(|r-s|) = \left(\frac{\alpha}{\alpha+1}\right)^{|r-s|}$, le processus $\{\epsilon_n\}$ est stationnaire au sens large.

□

Il s'agit là d'un processus à corrélation positive. L'obtention d'un processus à corrélation négative est impossible en conservant les mêmes hypothèses d'indépendance, soient $\{V_n\}$ i.i.d. $F_V(v)$, $\{a_n\}$ i.i.d. $\mathrm{Exp}(\lambda)$, $\{V_n\}$ et $\{a_n\}$ mutuellement indépendantes et $\epsilon_0 \sim \mathrm{Gamma}(\alpha+1, \lambda)$. En effet, puisque la distribution Gamma admet seulement des valeurs positives, une corrélation négative entraînerait une dépendance entre V_n et a_n ou entre ϵ_{n-1} et a_n.

3.2. Simulation d'un processus bêta

Avant de présenter le modèle, nous démontrerons deux résultats qui seront ultérieurement utilisés.

Lemme 3.2.1.

i) Si $U \sim \text{Bêta}(a, b)$, *alors* $V = 1 - U \sim \text{Bêta}(b, a)$.

ii) Si $U \sim \text{Bêta}(a, b)$ *et* $V \sim \text{Bêta}(a + b, c)$ *et si* U *et* V *sont des variables aléatoires indépendantes, alors* $W = UV \sim \text{Bêta}(a, b + c)$.

Preuve

i) Puisque $g(u) = 1 - u$ est une fonction strictement décroissante et dérivable, la densité de V est donnée par:

$$\begin{aligned} f_V(v) &= f_U(g^{-1}(u)) \left| \frac{dg^{-1}(u)}{du} \right| \\ &= f_U(1 - v), \end{aligned}$$

d'où $V \sim \text{Bêta}(b, a)$.

ii) Pour tout $s \geq 0$, on a:

$$\begin{aligned} \mathbb{E}[W^s] &= \mathbb{E}[U^s]\mathbb{E}[V^s] \\ &= \frac{\Gamma(a+b)\Gamma(a+s)}{\Gamma(a)\Gamma(a+b+s)} \frac{\Gamma(a+b+c)\Gamma(a+b+s)}{\Gamma(a+b)\Gamma(a+b+c+s)} \\ &= \frac{\Gamma(a+b+c)\Gamma(a+s)}{\Gamma(a)\Gamma(a+b+c+s)}. \end{aligned}$$

Et puisque les moments déterminent de façon unique la distribution d'une variable aléatoire, $W \sim \text{Bêta}(a, b + c)$.

□

Le processus $AR(1)$ Bêta avec corrélation positive est construit à partir de l'équation suivante:

$$\epsilon_n = 1 - U_n(1 - V_n\epsilon_{n-1})\,, \qquad n \geq 1\,, \tag{3.2.1}$$

où $\{U_n\}$ est une suite de variables aléatoires i.i.d. Bêta$(b, a-p)$ et $\{V_n\}$ est une suite de variables aléatoires i.i.d. Bêta$(p, a-p)$, où $0 < p < a$. On suppose aussi que les suites $\{U_n\}$ et $\{V_n\}$ sont mutuellement indépendantes et que U_n et V_n sont indépendantes de $\epsilon_{n-1}, \epsilon_{n-2}, \ldots$ On suppose finalement que $\epsilon_0 \sim$ Bêta(a, b).

Théorème 3.2.1.

Le processus $\{\epsilon_n\}$ défini par (3.2.1) est un processus stationnaire au sens large qui admet une distribution marginale Bêta(a, b) *avec une structure de corrélation positive donnée par* $\rho_j = \left(\frac{pb}{a(a+b-p)}\right)^j$, $j \geq 1$.

Preuve

Soit la proposition suivante:

$$P(n) : \epsilon_n \textit{ admet une distribution marginale } \text{Bêta}(a, b)\,.$$

Nous allons vérifier par induction que la proposition $P(n)$ est vraie pour tout $n \in \mathbb{N}$. Pour $n = 0$, $P(n)$ est vraie par la dernière hypothèse du modèle. On

suppose maintenant que $P(n-1)$ est vraie. On doit montrer que cela implique que $P(n)$ est également vraie alors. Or, si $\epsilon_{n-1} \sim \text{Bêta}(a,b)$, le lemme 3.2.1 nous permet d'affirmer, via respectivement $ii)$, $i)$ et $ii)$ une seconde fois, que:

$$\begin{aligned} V_n\epsilon_{n-1} &\sim \text{Bêta}(p, a+b-p)\,, \\ 1 - V_n\epsilon_{n-1} &\sim \text{Bêta}(a+b-p, p)\,, \\ U_n(1 - V_n\epsilon_{n-1}) &\sim \text{Bêta}(b, a)\,, \end{aligned}$$

d'où

$$\epsilon_n = 1 - U_n(1 - V_n\epsilon_{n-1}) \sim \text{Bêta}(a,b)$$

par le lemme 3.2.1 $i)$.

Pour déterminer la structure de corrélation, il est utile de réécrire l'équation sous la forme

$$\epsilon_n = U_n V_n \epsilon_{n-1} + (1 - U_n)\,.$$

Alors, il est facile de vérifier que, pour $j \geq 1$,

$$\begin{aligned} \text{Cov}(\epsilon_n, \epsilon_{n-j}) &= \text{Cov}\big(U_n\, V_n \epsilon_{n-1} + (1 - U_n), \epsilon_{n-j}\big) \\ &= \text{Cov}\big(U_n\, V_n \epsilon_{n-1}, \epsilon_{n-j}\big)\,, \end{aligned}$$

puisque ϵ_n est fonction uniquement de $U_n, U_{n-1}, U_{n-2}, \ldots$ et de $V_n, V_{n-1}, V_{n-2}, \ldots$ Étant donné l'indépendance mutuelle entre U_n et V_n et la remarque précédente, il s'en suit que

$$\begin{aligned}\mathrm{Cov}(\epsilon_n, \epsilon_{n-j}) &= \mathbb{E}[U_n]\mathbb{E}[V_n]\mathrm{Cov}(\epsilon_{n-1}, \epsilon_{n-j}) \\ &= \frac{pb}{a(a+b-p)}\mathrm{Cov}(\epsilon_{n-1}, \epsilon_{n-j}),\end{aligned}$$

d'où $\rho_j = \left(\frac{pb}{a(a+b-p)}\right)^j$.

Finalement, puisque les deux premiers moments de ϵ_n sont indépendants de n, car $\epsilon_n \sim$ Bêta(a, b), et puisque $\mathrm{Cor}(\epsilon_r, \epsilon_s) = f(|r-s|) = (\frac{pb}{a(a+b-p)})^{|r-s|}$, le processus $\{\epsilon_n\}$ est stationnaire au sens large.

□

Fait à noter, les restrictions imposées par la non-négativité des paramètres entraînent que $0 < p < a$, d'où $0 < \rho < 1$, c'est-à-dire que l'ensemble des corrélations positives sont possibles.

Pour obtenir une corrélation négative, le modèle est plutôt le suivant:

$$\epsilon_n = U_n(1 - V_n\epsilon_{n-1}), \tag{3.2.2}$$

où $\{U_n\}$ est une suite de variables aléatoires i.i.d. Bêta$(a, b-p)$ et $\{V_n\}$ est une suite de variables aléatoires i.i.d. Bêta$(p, a-p)$. Les hypothèses d'indépendance demeurent les mêmes que dans le modèle à corrélation positive.

Théorème 3.2.2.

Le processus $\{\epsilon_n\}$ *défini par (3.2.2) est un processus stationnaire au sens large qui admet une distribution marginale* Bêta(a, b) *et la structure de corrélation est donnée par* $\rho_j = \left(\frac{-p}{a+b-p}\right)^j$, $j \geq 1$.

Preuve

La preuve est identique à la précédente et ne sera donc pas reprise.

□

Chapitre 4

PERFORMANCES DANS LE CADRE DE SIMULATIONS

Dans ce chapitre, nous voulons étudier les deux méthodes de construction d'intervalles de prévision développées pour les modèles dictés par les états. Dans un premier temps, nous introduirons le cadre d'étude. Ensuite, nous présenterons les résultats des simulations, sous forme de graphiques et de tableaux. Nous discuterons des résultats et tenterons d'analyser le tout. Finalement, nous comparerons sommairement les deux méthodes.

4.1. Cadre d'étude

En ce qui a trait au cadre d'étude, notons d'abord que les simulations traitent uniquement le cas où $Y_t|\epsilon_t$ est poissonnien, car il s'agit du cas le plus classique, comme le cas normal dans le contexte continu. Il s'agit aussi du seul cas relativement facile à traiter avec les capacités informatiques disponibles actuellement, car les distributions résultantes pour Y_t dans les deux autres cas sont plutôt rébarbatives, lors de l'utilisation de la méthode paramétrique. Le vecteur de variables explicatives utilisé est le suivant:

$$\mathbf{X}_t = \left(1, t/n, cos(\frac{2\pi t}{12}), sin(\frac{2\pi t}{12})\right) .$$

Il s'agit d'un choix plutôt classique lorsque l'on traite des données mensuelles. Il inclut la composante constante, la tendance linéaire et les harmoniques annuelles. Le vecteur des paramètres de régression est fixé à la valeur suivante:

$$\boldsymbol{\beta} = (2,25;\ -1,25; 0,5; 0,5)\ .$$

Ce choix se justifie ainsi. Le but premier est de former un vecteur faisant en sorte que $\mu_t = exp(\mathbf{X}_t\boldsymbol{\beta})$ prenne des valeurs à l'intérieur d'un intervalle suffisamment large pour les valeurs de t supérieurs à 100. Ajoutons aussi que les valeurs des composantes du vecteur doivent être suffisamment grandes pour avoir un impact sur μ_t. Le signe négatif de β_2 a pour but de se rapprocher des séries épidémiologiques où les progrès de la médecine font en sorte que la tendance linéaire est généralement à la baisse. Les coefficients β_3 et β_4 ont été choisis du même signe, des signes opposés ne faisant que déplacer les courbes et ne changeant pas l'impact sur l'étude. Le graphique de la page suivante représente les valeurs résultantes pour $\mu_t = exp(\mathbf{X}_t\boldsymbol{\beta})$, pour t variant sur les entiers de 1 à 112.

Dans le but d'évaluer les performances et la pertinence des deux méthodes développées au chapitre 2, nous générons d'abord une réalisation d'un processus autorégressif d'ordre 1 stationnaire afin de simuler le processus latent $\{\epsilon_t\}$. Ensuite, nous générons les observations Y_t telles que les $Y_t|\epsilon_t$ soient indépendantes et distribuées selon des lois de Poisson de paramètre $\mu_t\epsilon_t$. La longueur des séries générées est de 112 observations, les cent premières servant à l'estimation et les 12 dernières à l'évaluation des performances. Nous reprenons le tout 2000 fois, afin d'obtenir de meilleures estimations des taux de couverture réels. Nous avons choisi de construire des intervalles de prévision au niveau 90% car il s'agit d'un choix classique, mais surtout parce que le fait que Y_t soit discret donnerait des taux trop près de 1 avec un niveau fixé à 95%.

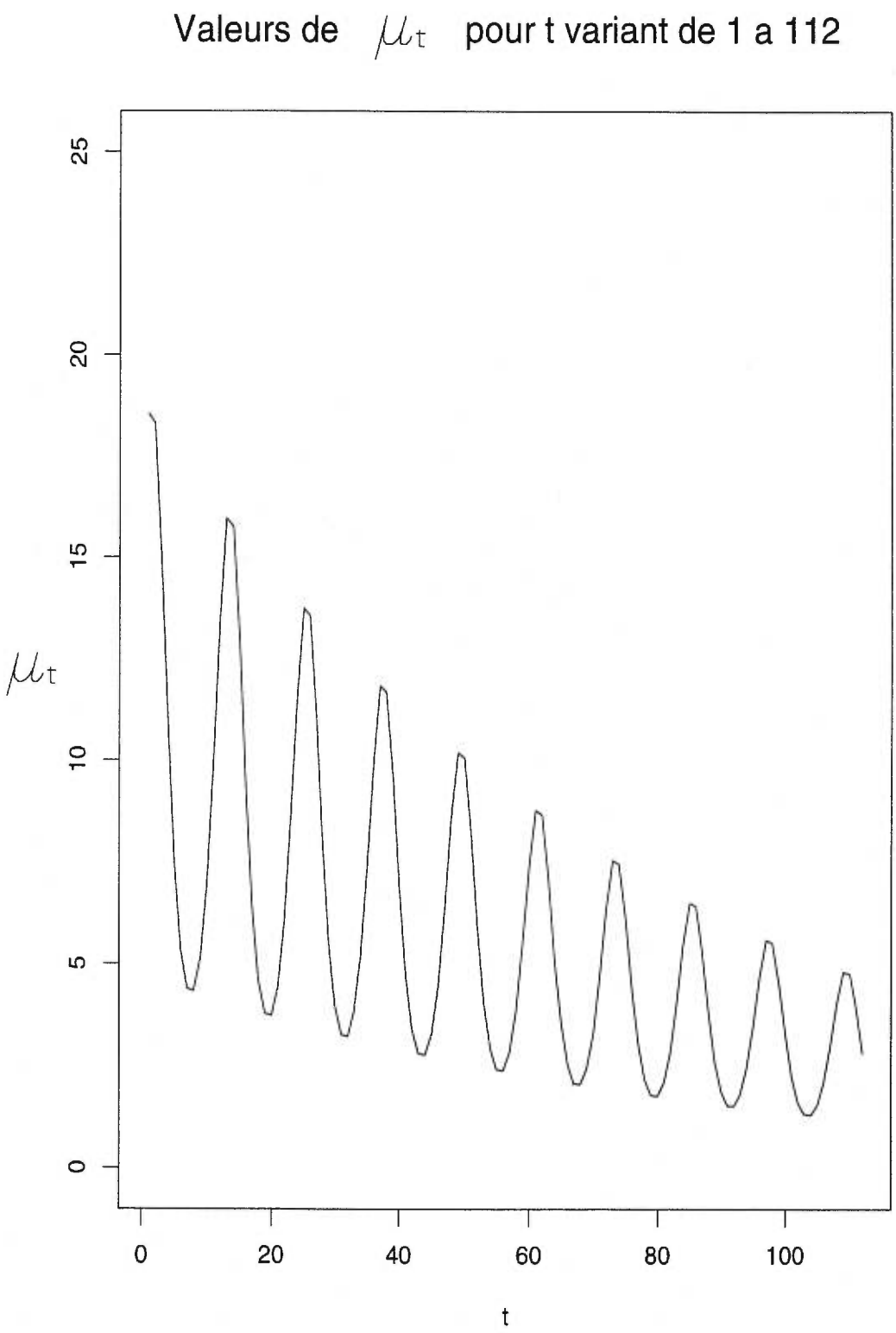

FIGURE 4.1.1. *Valeurs de $\mu_t = exp(\mathbf{X}_t\boldsymbol{\beta})$ pour t variant sur les entiers de 1 à 112, avec $\mathbf{X}_t$ et $\boldsymbol{\beta}$ tels que définis à la section 4.1.*

Afin de ne pas négliger l'effet des paramètres de nuisance, nous reprenons les analyses pour trois duos de variance (σ^2) et de corrélation de délai 1 (ϕ). Les combinaisons choisies sont $(1/2, 1/2)$, $(1/4, 3/4)$ et $(3/4, 1/4)$. Les raisons motivant ces choix seront invoquées incessamment. Notons que c'est pour les mêmes raisons qu'il est impossible de faire varier ces paramètres un à un, bien qu'une telle étude serait souhaitable. Notons finalement que puisqu'il est possible de le faire dans le cadre de simulations, nous évaluons les performances aussi bien lorsque les paramètres de nuisance sont connus que lorsqu'ils sont estimés.

Finalement, afin de pouvoir faire une analyse juste de la performance des deux méthodes, et plus particulièrement de la méthode paramétrique, nous reprenons les analyses pour des processus latents avec des distributions marginales différentes: gamma, lognormale et «bêta modifiée ».

Afin de simuler un processus stationnaire avec distribution marginale gamma, la méthode présentée à la section 3.1 est utilisée. Toutefois, une contrainte supplémentaire est ajoutée pour respecter le modèle de la section 1.2: on souhaite que $\mathbb{E}[\epsilon_t] = 1$. Puisque $\epsilon_t \sim \text{Gamma}(\alpha + 1, \lambda)$, il s'en suit que $\lambda = \alpha + 1$ et, par conséquent, que $\text{Var}(\epsilon_t) = \sigma^2 = 1/\lambda$. Et puisque, selon le théorème 3.1.1, la corrélation de délai 1 est donnée par $\phi = \frac{\alpha}{\alpha+1}$, il en découle la relation suivante:

$$\sigma^2 + \phi = 1$$

Voilà ce qui force la main quant aux choix des combinaisons de variance et de corrélation de délai 1 et qui explique l'impossiblité de les faire varier un à un.

Un processus plus commun est le processus normal ou lognormal, selon le point de vue. En effet si l'on suppose que $\{\nu_t\}$ est un processus stationnaire avec $\nu_t \sim \mathcal{N}(\nu, \tau^2)$ et que l'on modélise l'espérance conditionnelle de Y_t par

$$\mathbb{E}[Y_t|\epsilon_t] = exp(\mathbf{X}_t\boldsymbol{\beta} + \nu_t) = exp(\mathbf{X}_t\boldsymbol{\beta})exp(\nu_t) = \mu_t\epsilon_t\,,$$

il s'en suit que $\epsilon_t = exp(\nu_t)$ est lognormale. De multiples relations unissent alors les deux processus (Davis, Dunsmuir et Wang (1997)). Puisque

$$\mathbb{E}[\epsilon_t] = \mathbb{E}[e^{\nu_t}] = \mathrm{M}_{\nu_t}(1) = e^{\nu+\tau^2/2}\,,$$

où $M_{\nu_t}(\cdot)$ dénote la fonction génératrice des moments de ν_t, la contrainte $\mathbb{E}[\epsilon_t] = 1$ implique alors que $\nu = -\tau^2/2$. De la même façon, puisque

$$\mathbb{E}[{\epsilon_t}^2] = \mathbb{E}[e^{2\nu_t}] = \mathrm{M}_{\nu_t}(2) = e^{2\nu+2\tau^2} = e^{\tau^2}\,,$$

il s'en suit que $\mathrm{Var}(\epsilon_t) = \sigma^2 = e^{\tau^2} - 1$, c'est-à-dire que

$$\tau^2 = \log(\sigma^2 + 1)\,.$$

Finalement, une relation unit aussi les fonctions d'autocovariances et, par le fait même, les fonctions d'autocorrélations. En effet,

$$\begin{aligned}\gamma_\epsilon(h) &= \mathrm{Cov}(\epsilon_{t+h}, \epsilon_t)\\ &= \mathbb{E}[\epsilon_{t+h}\epsilon_t] - 1\\ &= \mathbb{E}[e^{\nu_{t+h}}e^{\nu_t}] - 1\,.\end{aligned}$$

Par les propriétés de la fonction exponentielle et en remarquant qu'il s'agit alors de la fonction génératrice des moments de $\nu_{t+h} + \nu_t$ évaluée en $s = 1$, on a

$$\begin{aligned}\gamma_\epsilon(h) &= \mathbb{E}[e^{\nu_{t+h}+\nu_t}] - 1 \\ &= M_{\nu_{t+h}+\nu_t}(1) - 1\,.\end{aligned}$$

Or, puisque le processus $\{\nu_t\}$ est stationnaire et gaussien et que $\nu_t \sim \mathcal{N}(-\tau^2/2, \tau^2)$ pour tout t, on a

$$\nu_{t+h} + \nu_t \sim \mathcal{N}\big(-\tau^2, 2(\tau^2 + \gamma_\nu(h))\big)\,.$$

Alors,

$$\begin{aligned}\gamma_\epsilon(h) &= exp\{-\tau^2 + (\tau^2 + \gamma_\nu(h))\} - 1 \\ &= e^{\gamma_\nu(h)} - 1\,,\end{aligned}$$

ce qui implique que $\rho_\epsilon(h) = \frac{e^{\gamma_\nu(h)}-1}{\sigma^2}$, ou inversement que

$$\rho_\nu(h) = \frac{\log\left(\rho_\epsilon(h)\sigma^2 + 1\right)}{\log(\sigma^2 + 1)}\,.$$

Donc, pour obtenir un processus latent $\{\epsilon_t\}$ distribué marginalement selon une loi lognormale de moyenne 1, de variance σ^2 et dont la corrélation de délai 1 est ϕ, il suffit de simuler un processus gaussien $\{\nu_t\}$ de moyenne $-\log(\sigma^2+1)/2$, de variance $\log(\sigma^2+1)$ et dont la corrélation de délai 1 est $\log\left(\phi\sigma^2 + 1\right)/\log(\sigma^2+1)$ selon la méthode habituelle et de poser $\epsilon_t = exp(\nu_t)$.

La méthode présentée à la section 3.2 est utilisée quant à elle pour simuler un processus stationnaire avec distribution marginale bêta. Dans ce cas, nous avons déjà soulevé le fait que l'équation (1.2.5) imposait une condition non désirée sur les paramètres, soit $b = 0$. Pour résoudre ce problème, nous avons tout d'abord simulé un processus stationnaire $\{Z_t\}$ avec distribution marginale Bêta(a, b) satisfaisant aux équations suivantes:

$$\mathbb{E}[Z_t] = 1/c\,,$$
$$\mathrm{Var}(Z_t) = \sigma^2/c^2\,,$$
$$\rho_z(1) = \phi\,,$$

pour un c initialement déterminé. Nous avons ensuite posé $\epsilon_t = cZ_t$ et utilisé le processus $\{\epsilon_t\}$. Le choix d'une valeur pour c étant illimitée, ces conditions ne définissent pas un processus $\{\epsilon_t\}$ unique, mais tel n'est pas le but de la démarche qui vise à générer des processus $\{\epsilon_t\}$ avec distribution marginale autre que gamma de sorte à pouvoir mieux évaluer les performances de la méthode paramétrique. Notons toutefois que le processus $\{\epsilon_t\}$ ainsi construit respecte les conditions imposées par le modèle, soient $\mathbb{E}[\epsilon_t] = 1$, $\mathrm{Var}(\epsilon_t) = \sigma^2$ et $\rho_\epsilon(1) = \phi$. Ajoutons finalement que le processus $\{\epsilon_t\}$ n'admet pas une distribution bêta, bien que construit à partir d'un processus stationnaire avec distribution marginale bêta, d'où l'appellation «bêta modifiée ».

4.2. Résultats des simulations

Dans les pages qui suivent, nous présentons, sous forme graphique, les résultats des simulations obtenus conformément au cadre d'étude fixé à la section précédente. Les résultats détaillés peuvent être consultés, sous forme de tableaux, à l'annexe. Les résultats sont présentés dans l'ordre dans lequel ils sont analysés. Les résultats concernant la méthode paramétrique précèdent donc les résultats non paramétriques, alors qu'à l'intérieur de chaque méthode, les taux unilatéraux précèdent leurs homologues bilatéraux.

Il est à noter que, dans le cas de la méthode paramétrique, les vrais taux sous-jacents ne sont pas exactement de 90%, car la distribution résultante pour Y_t est discrète. Malgré cela, puisque le but ultime est de développer des intervalles de prévision de niveau 90% et qu'en pratique, la vraie distribution de Y_t est inconnue, le taux de référence utilisé sera tout de même 90%.

Dans le cas de la méthode non paramétrique, le taux de référence utilisé est également 90%. Pour bien en comprendre la raison, rappelons d'abord que, par exemple, l'intervalle bilatéral a la forme suivante:

$$Y_{n+l} \in \left(T^{-1}(q_{\delta;0.05} + T(\hat{\mu}_{n+l})), T^{-1}(q_{\delta;0.95} + T(\hat{\mu}_{n+l}))\right) ,$$

ou, sous une forme équivalente,

$$T(Y_{n+l}) - T(\hat{\mu}_{n+l}) \in (q_{\delta;0.05}, q_{\delta;0.95}) .$$

Sous l'hypothèse que les $T(Y_t) - T(\mu_t)$ sont identiquement distribués, les différences $\delta_t = T(Y_t) - T(\hat{\mu}_t)$, approximativement stationnaires en moyenne et en

variance, devraient aussi être approximativement identiquement distribués. Cet intervalle devrait donc être de niveau approximatif 90%.

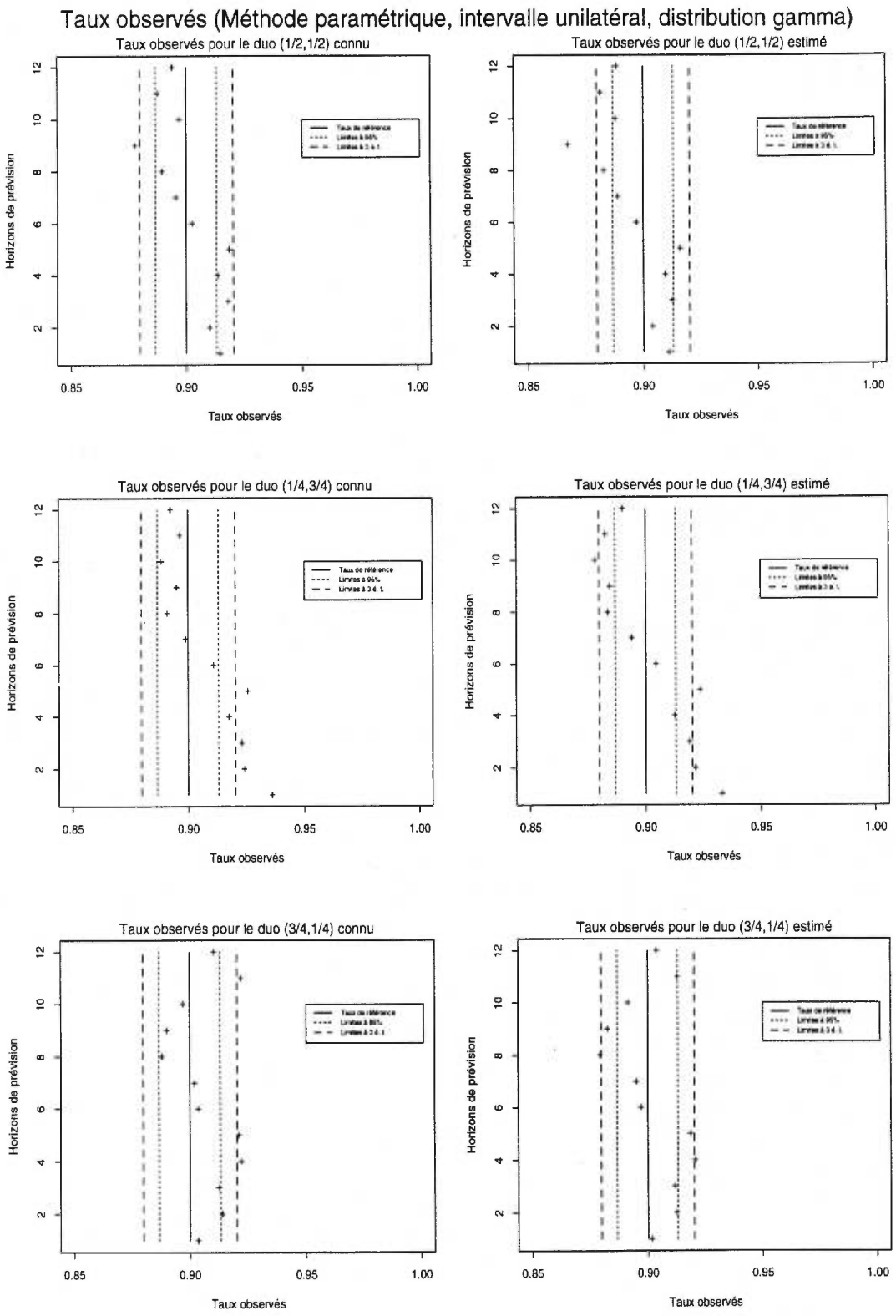

FIGURE 4.2.1. *Taux observés (Méthode paramétrique, intervalle unilatéral, $\epsilon_t \sim$ gamma)*

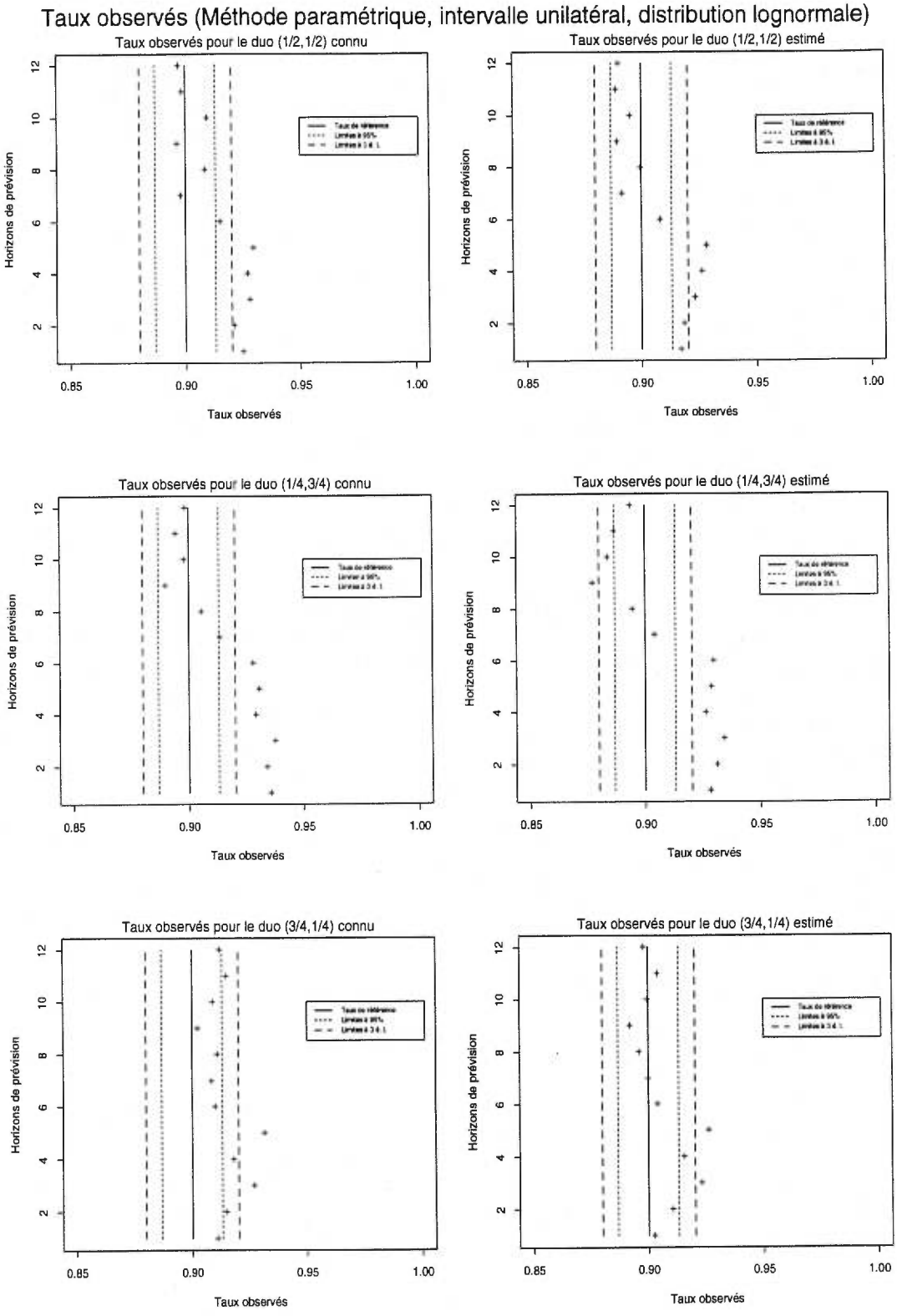

FIGURE 4.2.2. *Taux observés (Méthode paramétrique, intervalle unilatéral, $\epsilon_t \sim$ lognormale)*

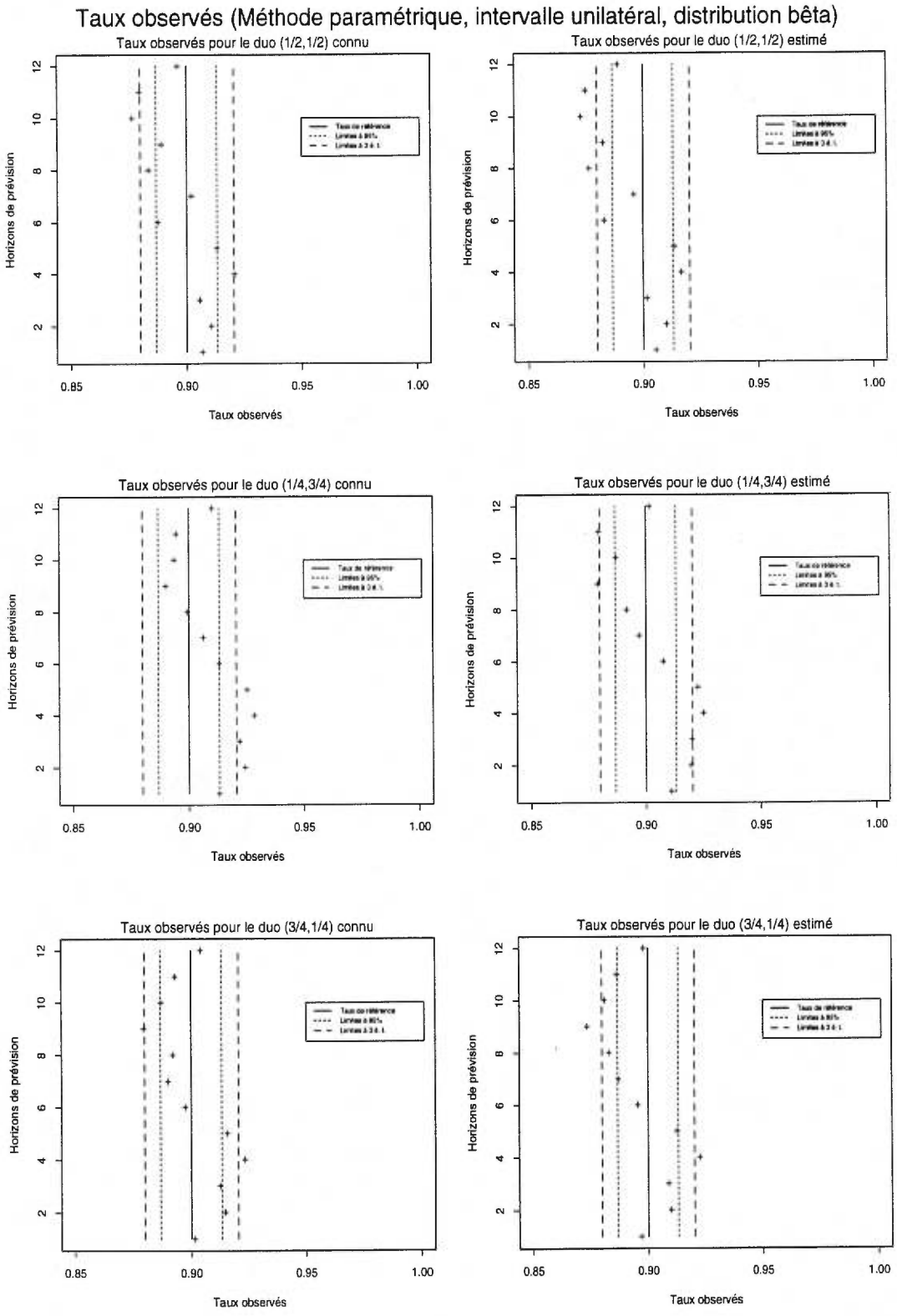

FIGURE 4.2.3. *Taux observés (Méthode paramétrique, intervalle unilatéral, $\epsilon_t \sim$ bêta modifiée)*

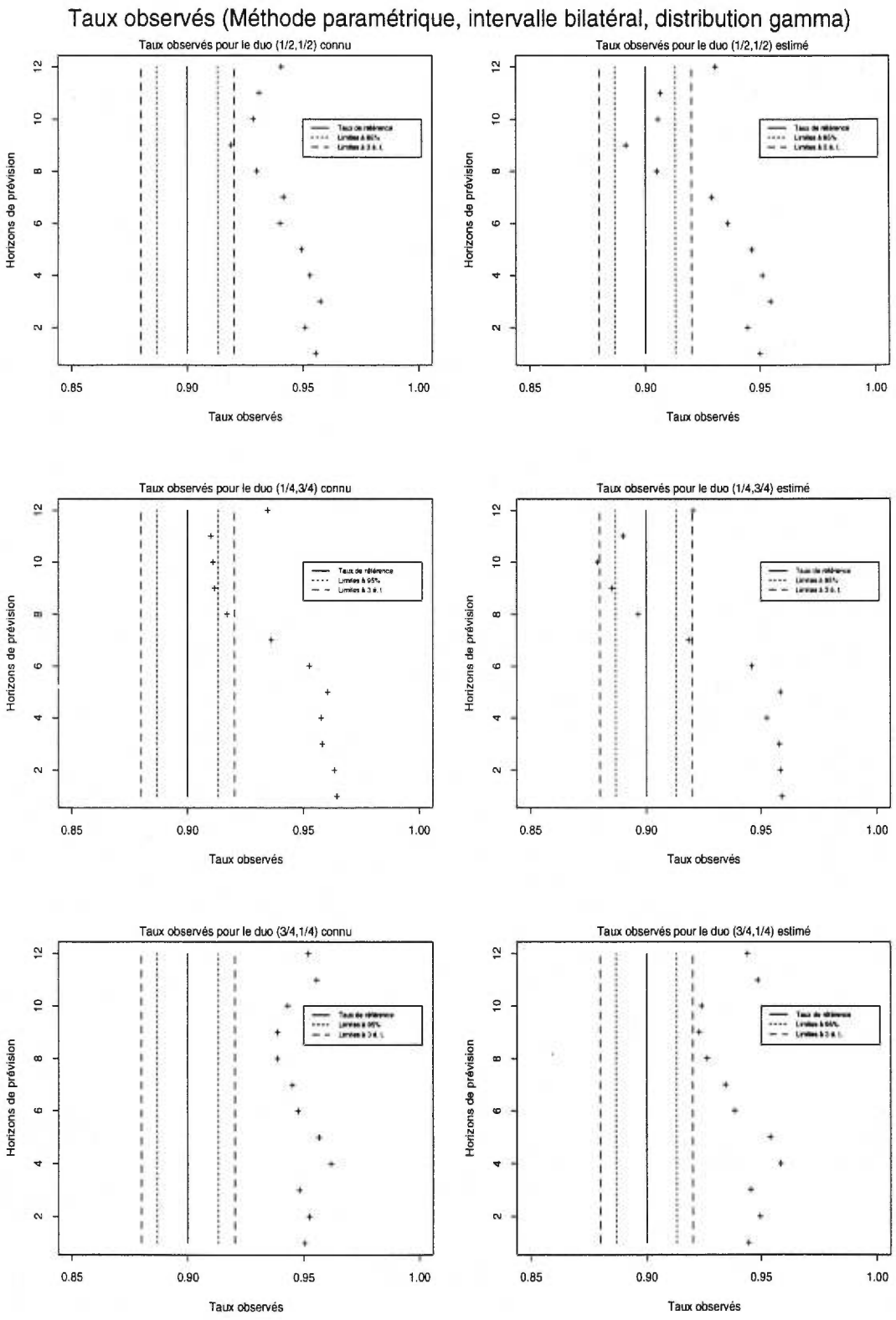

FIGURE 4.2.4. *Taux observés (Méthode paramétrique, intervalle bilatéral, $\epsilon_t \sim$ gamma)*

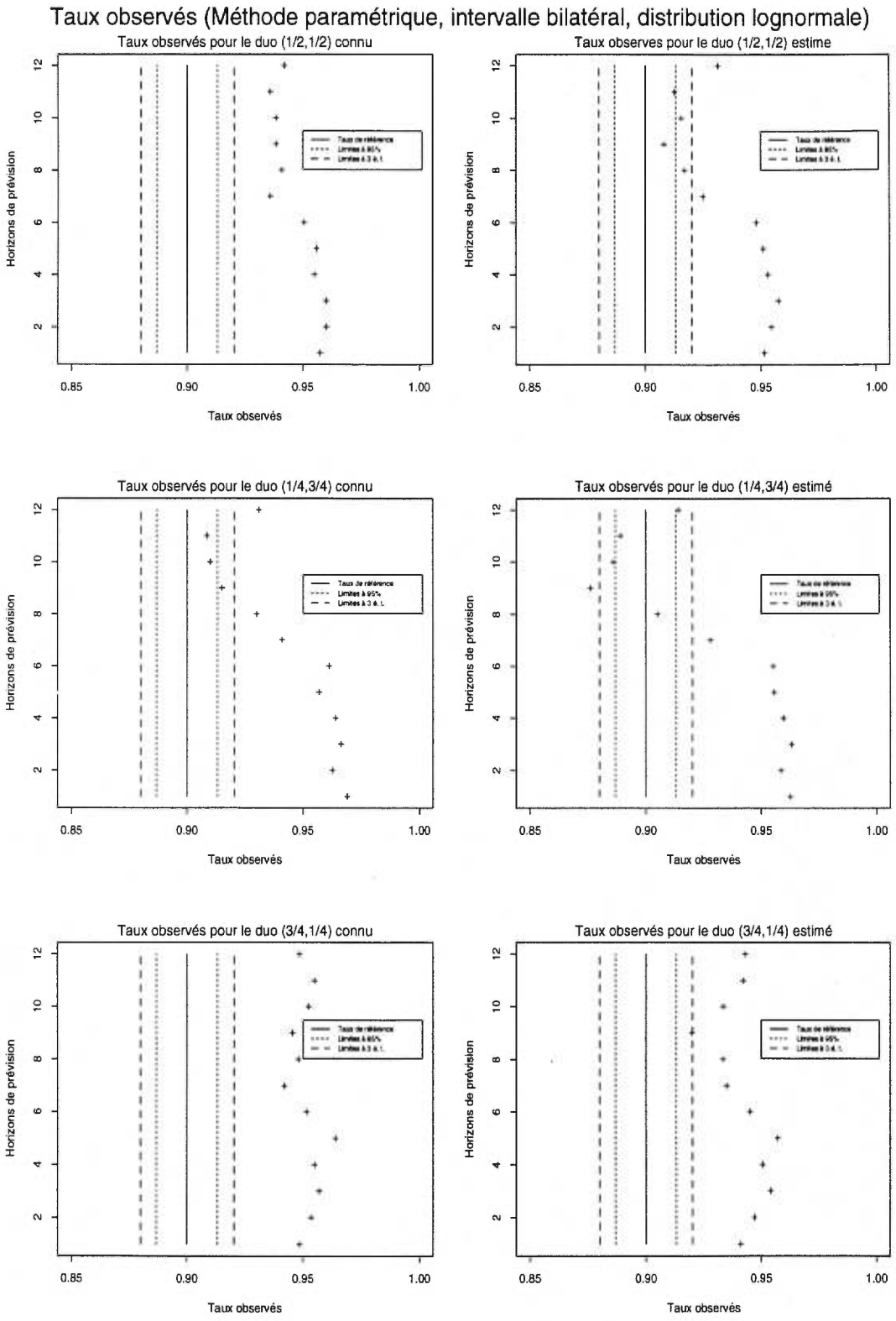

FIGURE 4.2.5. *Taux observés (Méthode paramétrique, intervalle bilatéral, $\epsilon_t \sim$ lognormale)*

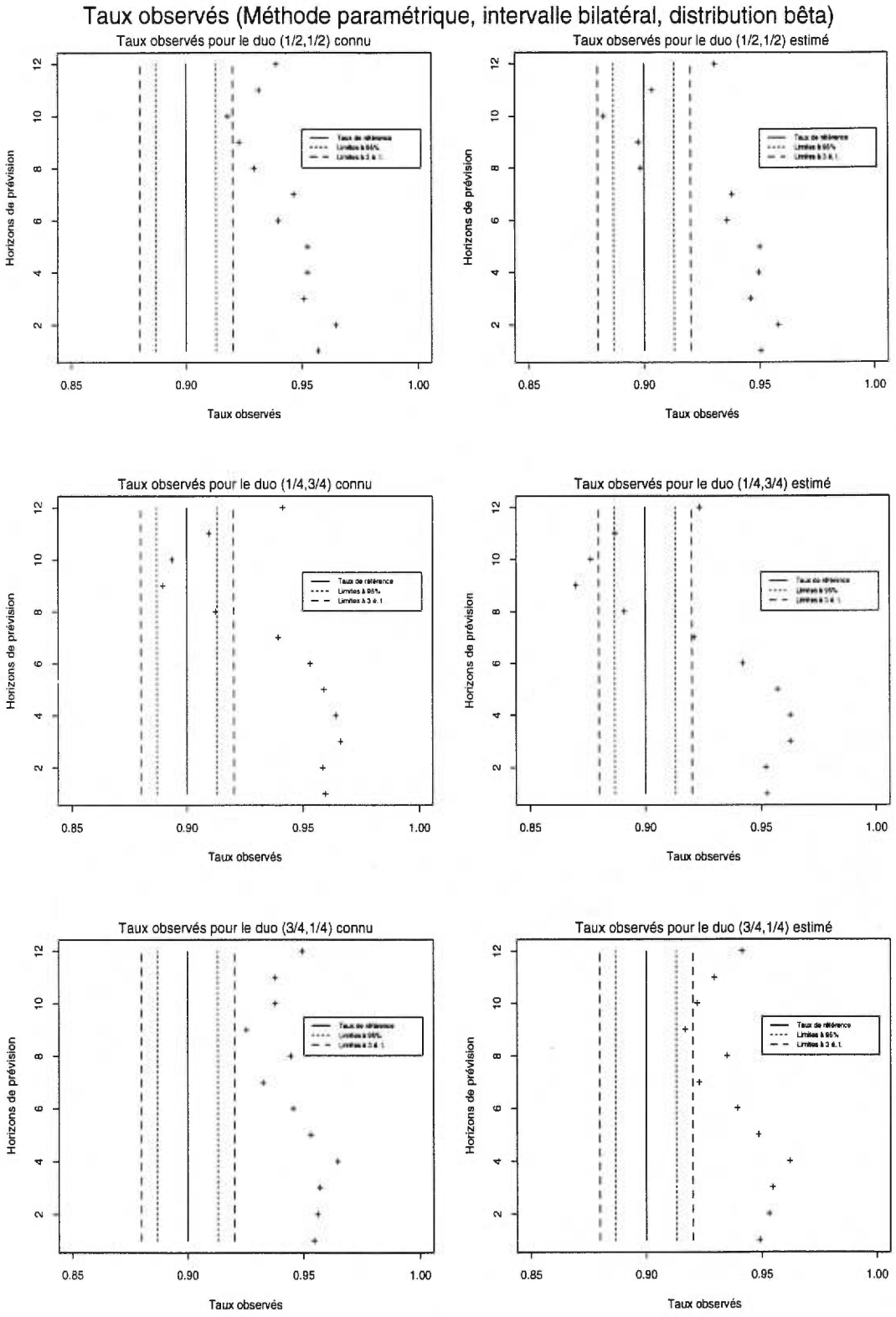

FIGURE 4.2.6. *Taux observés (Méthode paramétrique, intervalle bilatéral, $\epsilon_t \sim$ bêta modifiée)*

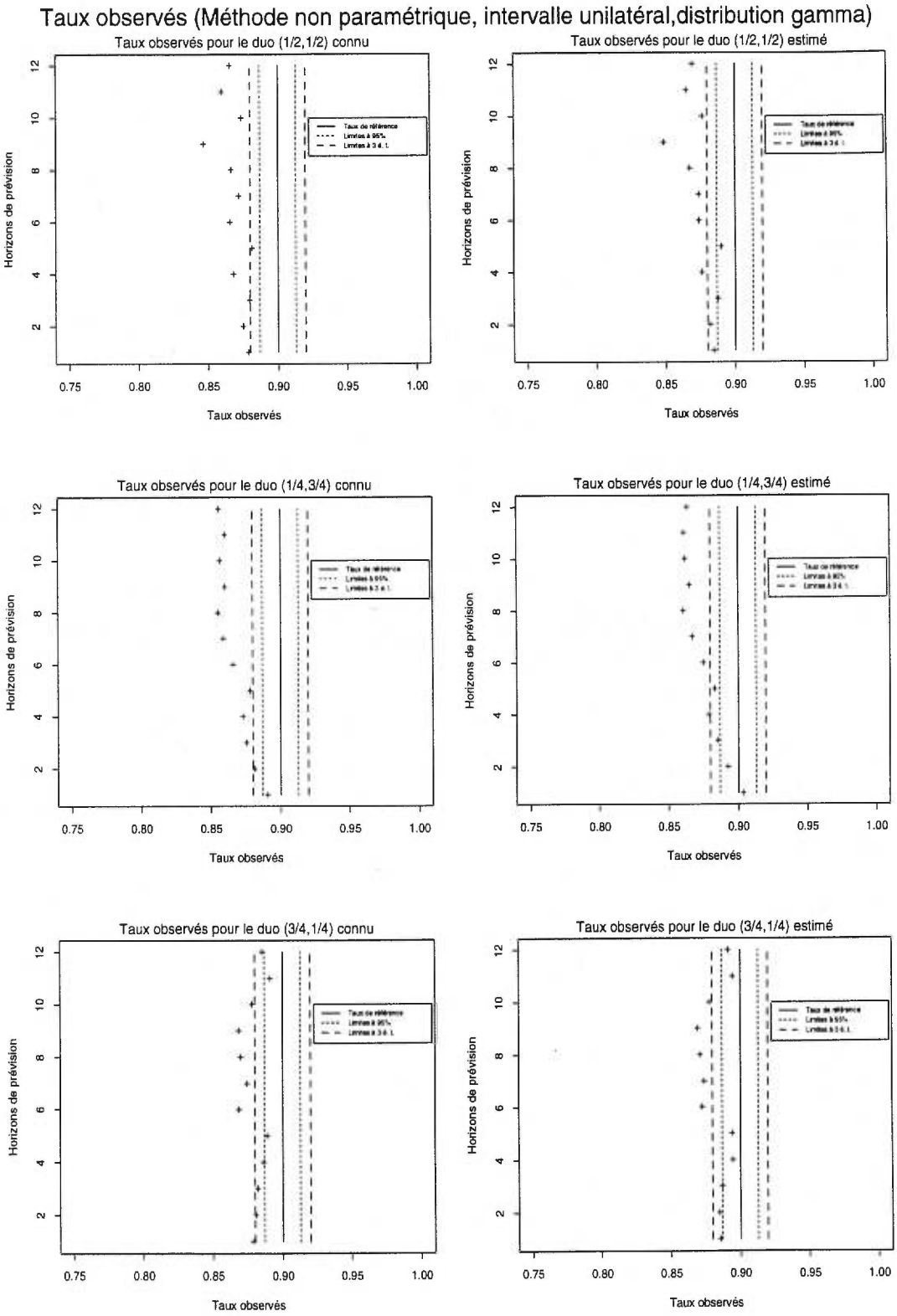

FIGURE 4.2.7. *Taux observés (Méthode non paramétrique, intervalle unilatéral, $\epsilon_t \sim$ gamma)*

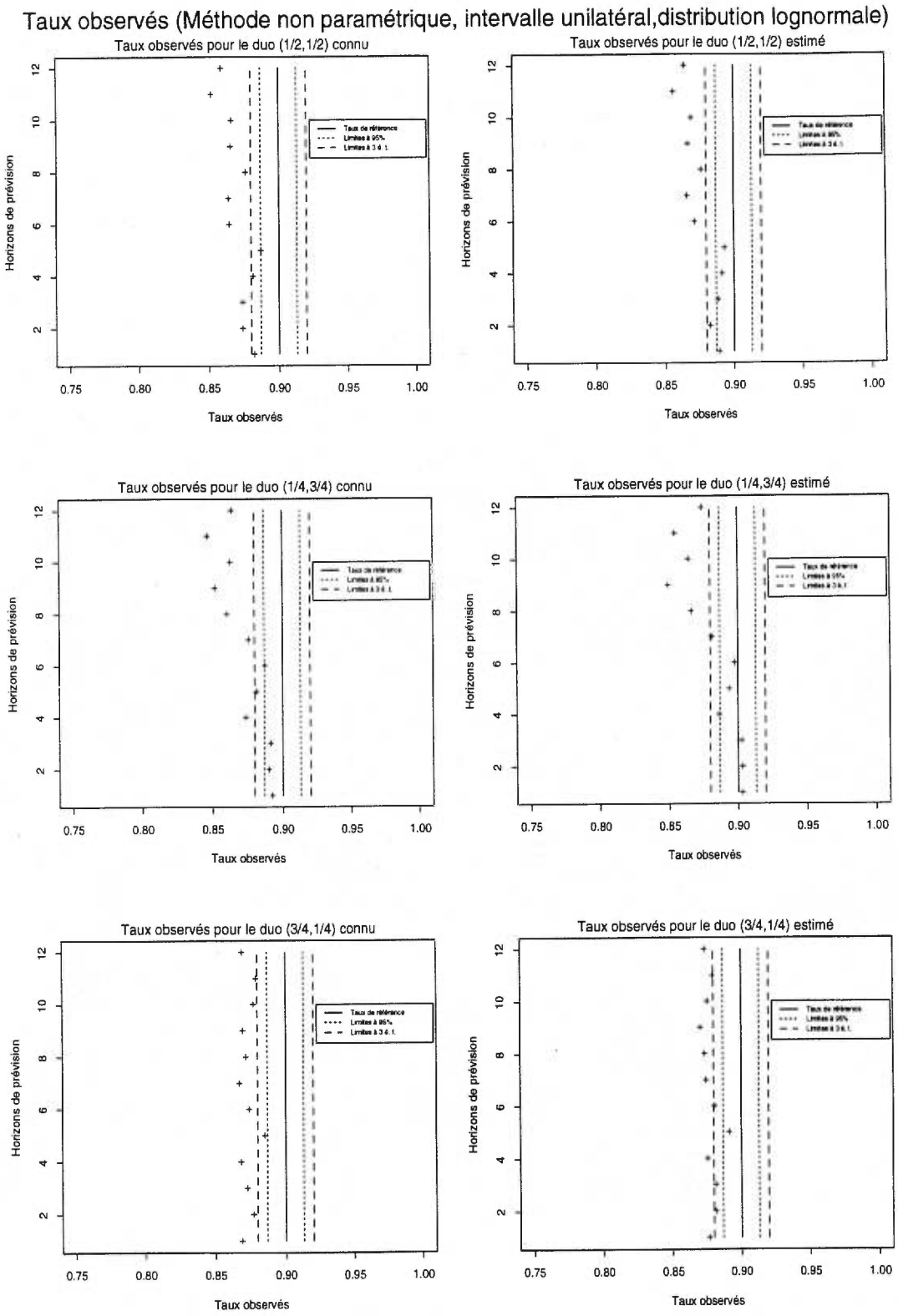

FIGURE 4.2.8. *Taux observés (Méthode non paramétrique, intervalle unilatéral, $\epsilon_t \sim$ lognormale)*

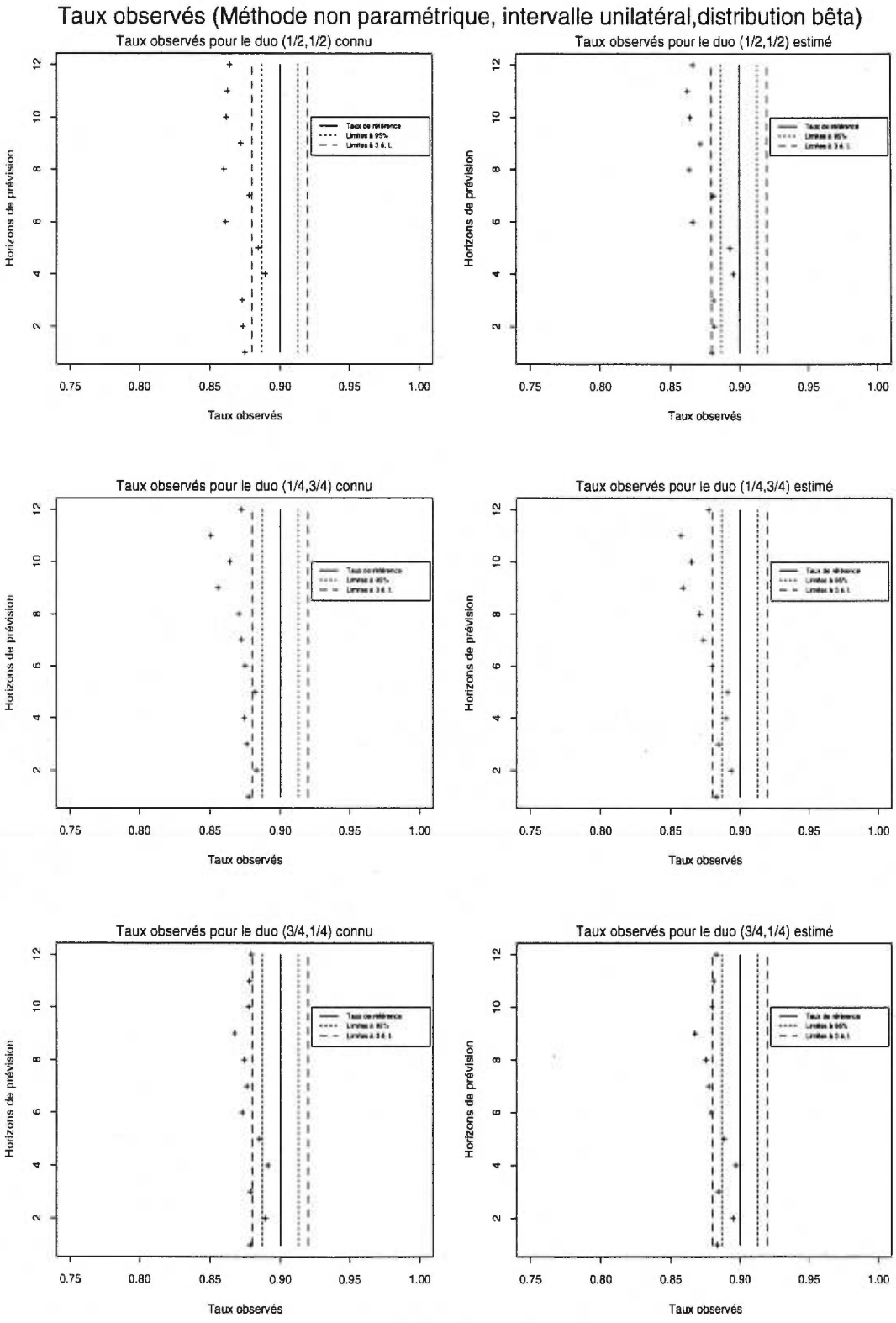

FIGURE 4.2.9. *Taux observés (Méthode non paramétrique, intervalle unilatéral, $\epsilon_t \sim$ bêta modifiée)*

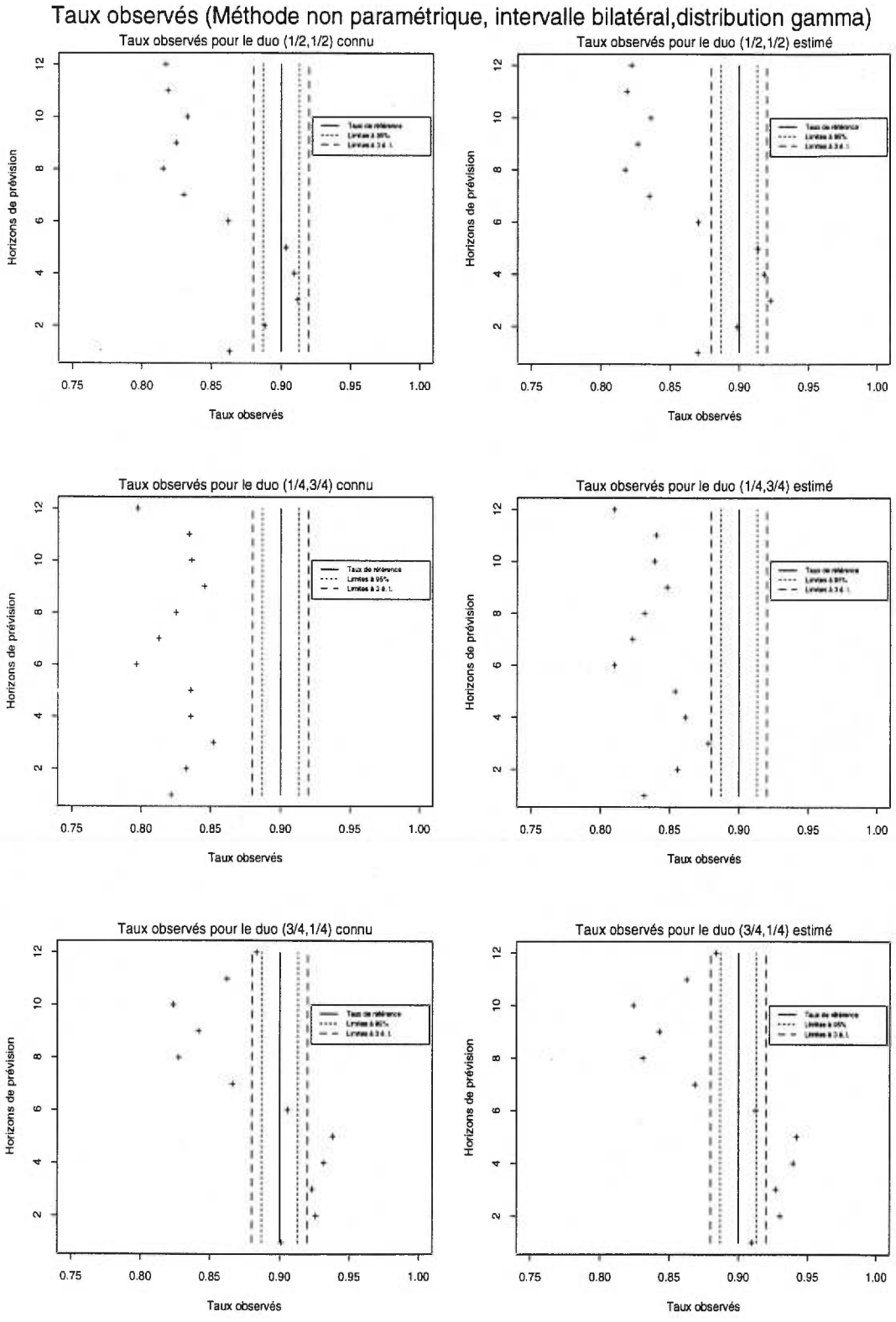

FIGURE 4.2.10. *Taux observés (Méthode non paramétrique, intervalle bilatéral, $\epsilon_t \sim$ gamma)*

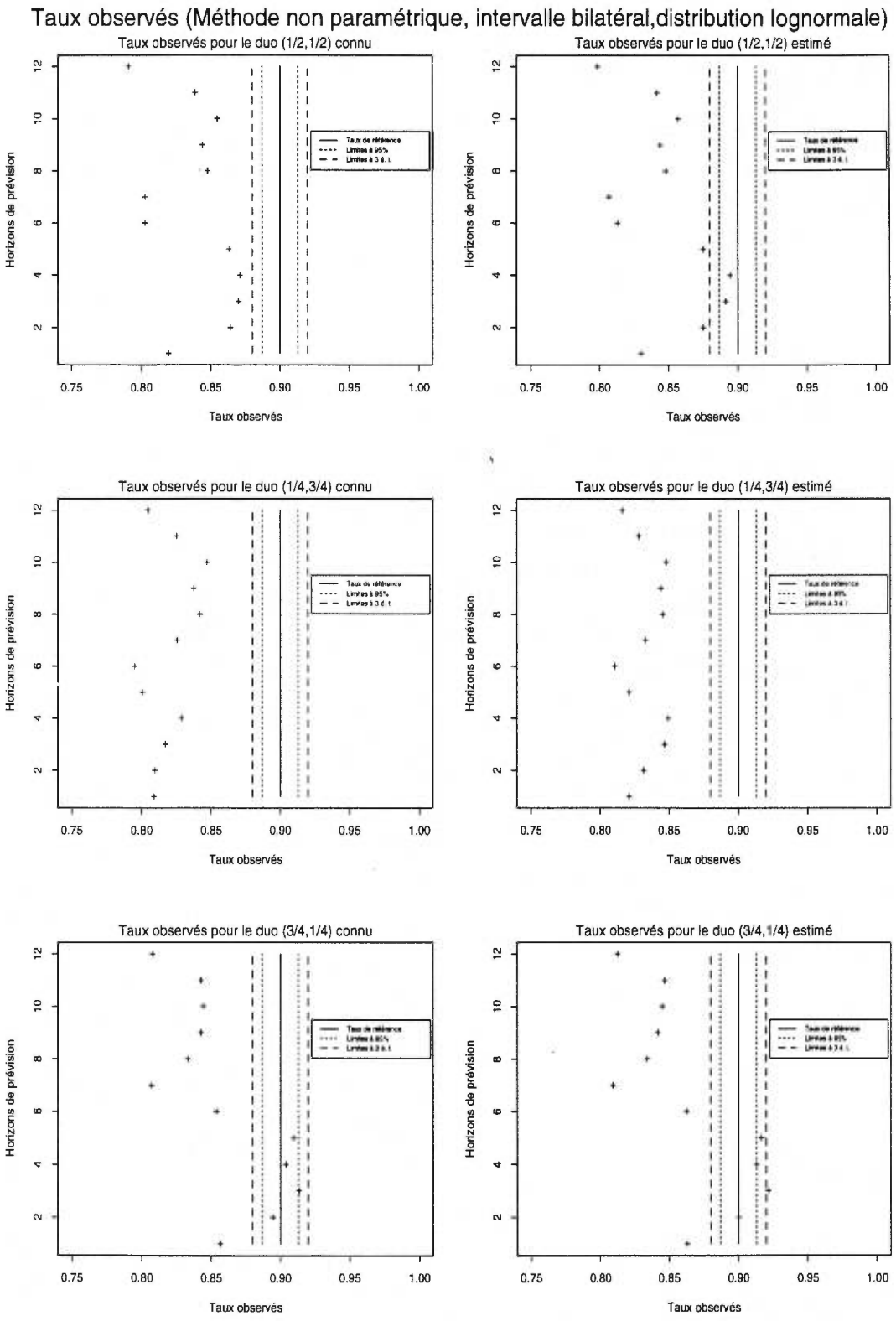

FIGURE 4.2.11. *Taux observés (Méthode non paramétrique, intervalle bilatéral, $\epsilon_t \sim$ lognormale)*

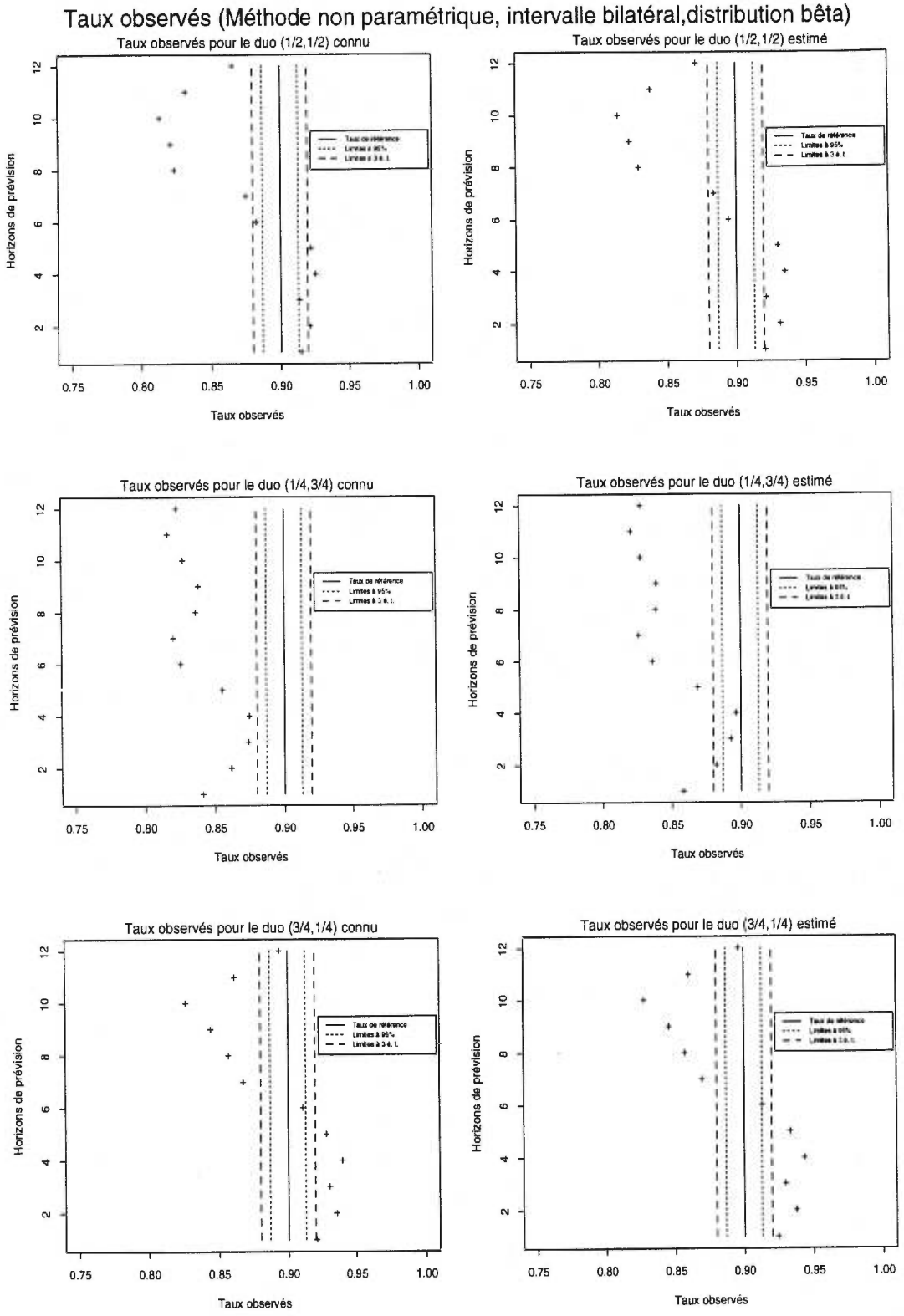

FIGURE 4.2.12. *Taux observés (Méthode non paramétrique, intervalle bilatéral, $\epsilon_t \sim$ bêta modifiée)*

4.3. Analyse des résultats et discussion

Afin de mieux analyser les résultats, nous décomposerons l'analyse des résultats en trois parties: l'analyse de la méthode paramétrique, l'analyse de la méthode non paramétrique et la comparaison des deux méthodes.

4.3.1. Analyse de la méthode paramétrique

Dans le cadre d'étude choisi, la méthode paramétrique consiste à supposer que ϵ_t admet une distribution marginale gamma et à utiliser la distribution résultante de Y_t pour construire l'intervalle de prévision. Plusieurs éléments importants doivent être étudiés. Tout d'abord, nous vérifierons si la méthode permet de construire des intervalles de prévision de niveau 90%. Pour ce faire, nous confronterons les hypothèses:

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p \neq p_0 \tag{4.3.1}$$

Il s'agit en fait de faire le test d'une proportion de la façon classique. Il suffit de vérifier si le taux observé p est inclus dans l'intervalle

$$p_0 \pm 1,96\sqrt{p_0(1-p_0)/m}\,,$$

où $p_0 = 0.90$ est le taux de référence et m, le nombre de répétitions. Nous rappelons ici que nous avons utilisé $m = 2000$. L'intervalle de non-rejet est donc

$$(88,69\% \ , \ 91,31\%)$$

Pour fin de discussion, nous utiliserons également l'intervalle avec limites à 3 écarts types donné par

$$p_0 \pm 3\sqrt{p_0(1-p_0)/m}\,,$$

soit, dans notre cas, l'intervalle

$$(87,99\%\ ,\ 92,01\%)\,.$$

Par la suite, nous étudierons l'impact de l'estimation des paramètres de nuisance. Nous profiterons de ce contexte pour analyser l'effet des valeurs spécifiques de μ_t sur les taux observés. Finalement, nous nous intéresserons de plus près à l'effet causé par l'hypothèse «ϵ_t est marginalement gamma» lorsque celle-ci n'est pas vérifiée. Pour ce faire, nous confronterons, toutes choses étant égales par ailleurs,

$$\begin{aligned} H_0 &: \text{taux } (\epsilon_t \sim \text{gamma}) = \text{taux } (\epsilon_t \sim \text{lognormale}) \\ H_1 &: \neg H_0\,, \end{aligned} \tag{4.3.2}$$

de même que

$$\begin{aligned} H_0' &: \text{taux } (\epsilon_t \sim \text{gamma}) = \text{taux } (\epsilon_t \sim \text{bêta modifiée}) \\ H_1' &: \neg H_0\,. \end{aligned} \tag{4.3.3}$$

Il s'agit en fait du test classique de l'égalité de deux proportions. La région de rejet du test de niveau 5% est donnée par l'ensemble des valeurs de $\hat{p}_1$ et de $\hat{p}_2$ tel que

$$\left| \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/m_1 + 1/m_2)}} \right| \geq 1,96\,,$$

où $\hat{p}_1$ est le taux de couverture observé pour le cas où $\epsilon_t \sim$ gamma, $\hat{p}_2$, son homologue pour le cas où $\epsilon_t \sim$ lognormale (resp. bêta modifiée), $\hat{p}$, la proportion sous l'hypothèse $p_1 = p_2$ ($\hat{p} = \frac{m_1\hat{p}_1 + m_2\hat{p}_2}{m_1+m_2}$), m_1 et m_2, le nombre de simulations effectuées pour chaque distribution. Ici, $m_1 = m_2 = 2000$.

4.3.1.1. *Le cas unilatéral*

Débutons donc par l'analyse de la méthode paramétrique dans un contexte de construction d'intervalles de prévision unilatéraux.

Le cas $\epsilon_t \sim$ gamma.

Lorsque les paramètres de nuisance sont connus, les taux varient entre 87,80% et 91,85% pour le duo $(\sigma^2, \phi) = (1/2, 1/2)$, entre 88,85% et 93,60% pour le duo (1/4,3/4) et entre 88,80% et 92,20% pour le duo (3/4,1/4). Le nombre de taux observés situés à moins de 2 écarts types de 0,90 est de 7 sur 12 pour le duo (1/2,1/2), de 7 sur 12 pour le duo (1/4,3/4) et de 8 sur 12 pour le duo (3/4,1/4). À moins de 3 écarts types de 0,90, on retrouve 11 taux sur 12 pour le duo (1/2,1/2), 8 taux sur 12 pour le duo (1/4,3/4) et 9 taux sur 12 pour le duo (3/4,1/4). Un seul taux (87,80%, duo (1/2,1/2), horizon de prévision $l = 9$) est en deçà de 87,99% (90% - 3 écarts types). Quant aux taux situés au-delà de 92,01% (90% + 3 écarts types), ils sont tous très près de cette limite, à l'exception d'un seul (93,60%, duo (1/4,3/4), $l = 1$).

Lorsque les paramètres de nuisance sont estimés, les résultats sont similaires. Cette fois-ci, les taux observés varient entre 86,75% et 91,60% pour la combinaison $(\sigma^2, \phi) = (1/2, 1/2)$, entre 87,80% et 93,30% pour la combinaison (1/4,3/4) et entre 87,95% et 92,05% pour le duo (3/4,1/4). À moins de 2 écarts types de 0,90, on observe 8 taux sur 12 pour le duo (1/2,1/2), 4 taux sur 12 pour le duo (1/4,3/4) et 8 taux sur 12 pour le duo (3/4,1/4). Le nombre de taux observés situés à moins de 3 écarts types de 0,90 est de 11 taux sur 12 pour le duo (1/2,1/2), de 8 taux sur 12 pour le duo (1/4,3/4) et de 10 taux sur 12 pour le duo (3/4,1/4). On remarque des taux un peu plus faibles pour les horizons 8, 9, 10, et 11. Trois de ces horizons ont des taux inférieurs à 88,69% pour le duo (1/2,1/2), les quatre horizons ont des taux inférieurs à 88,69% pour le duo (1/4,3/4), alors que deux de ces horizons ont des taux inférieurs à 88,69% pour le duo (3/4,1/4). On peut également remarquer que 4 des 5 premiers horizons ont des taux supérieurs à 91,31% pour le duo (1/4,3/4). Un dernier élément intéressant à remarquer est le fait que chacun des taux observés est inférieur à son homologue dans le cas où les paramètres de nuisance sont connus, et ce, sans exception.

L'observation de quelques taux à plus de 3 écarts types de 0,90 est en partie due au fait que les taux sous-jacents ne sont pas exactement de 90%, la distribution de Y_t étant discrète. Les tableaux suivants indiquent les taux sous-jacents pour chaque horizon de prévision et pour chaque combinaison étudiée. Ces taux constituent les taux de couverture lorsque tous les paramètres, aussi bien les paramètres de nuisance que les paramètres de régression, sont connus et lorsque la distribution marginale de ϵ_t est effectivement gamma.

TABLEAU 4.3.1. *Taux de couverture sous-jacents de l'intervalle unilatéral, en pourcentage, pour les temps 101 à 106.*

(σ^2, ϕ)	101	102	103	104	105	106
$(1/2, 1/2)$	91,71	93,37	91,45	91,67	93,95	92,91
$(1/4, 3/4)$	94,00	95,17	93,08	93,30	95,68	90,04
$(3/4, 1/4)$	90,17	92,05	90,30	90,53	92,67	91,40

TABLEAU 4.3.2. *Taux de couverture sous-jacents de l'intervalle unilatéral, en pourcentage, pour les temps 107 à 112.*

(σ^2, ϕ)	107	108	109	110	111	112
$(1/2, 1/2)$	93,01	92,19	90,77	91,08	90,20	90,92
$(1/4, 3/4)$	92,16	92,36	91,19	91,54	93,26	93,53
$(3/4, 1/4)$	91,26	90,25	91,00	91,27	91,07	92,46

Ces taux permettent de mieux comprendre les taux supérieurs observés pour les 5 premiers horizons de prévision, particulièrement pour le duo (1/4,3/4), ainsi que les taux inférieurs observés pour les horizons 8,9,10 et 11, particulièrement pour les duos (1/2,1/2) et (3/4,1/4).

Le cas $\epsilon_t \sim$ lognormale.

Avec les paramètres de nuisance connus, les taux observés fluctuent entre 89,65% et 92,95% pour le duo (1/2,1/2), entre 89,00% et 93,75% pour le duo (1/4,3/4) et entre 90,25% et 93,15% pour le duo (3/4,1/4). À l'intérieur de l'intervalle avec limites à 2 écarts types de 0,90, on retrouve 6 taux sur 12 pour le duo (1/2,1/2), 5 taux sur 12 pour le duo (1/4,3/4) et 7 taux sur 12 pour le duo (3/4,1/4). À moins de 3 écarts types de 0,90, on compte 7 taux sur 12 pour la combinaison (1/2,1/2), 6 taux sur 12 pour la combinaison (1/4,3/4) et 10 taux sur 12 pour la combinaison (3/4,1/4). Le taux minimal observé étant de 89,00% (duo (1/4,3/4), $l = 9$), aucun taux observé n'est en deçà de 88,69% (90% - 2 écarts types). Les taux supérieurs sont principalement observés pour les 6 premiers horizons de prévision, particulièrement pour le duo (1/4,3/4). Il s'agit d'un comportement semblable à celui observé dans le cas gamma. Notons toutefois que, de façon générale, le taux observé dans le contexte lognormal est supérieur à son homologue du contexte gamma.

Lorsque les paramètres de nuisance sont estimés, les taux observés varient entre 88,90% et 92,80% pour la combinaison (1/2,1/2), entre 87,70% et 93,40% pour la combinaison (1/4,3/4) et entre 89,20% et 92,60% pour la combinaison (3/4,1/4). Le nombre de taux observés situés à moins de 2 écarts types de 0,90 est de 7 sur 12 pour le duo (1/2,1/2), de 3 sur 12 pour le duo (1/4,3/4) et de 9 sur 12 pour le duo (3/4,1/4). À moins de 3 écarts types, on observe 9 taux sur 12 pour le duo (1/2,1/2), 5 taux sur 12 pour le duo (1/4,3/4) et 10 taux sur 12 pour le duo (3/4,1/4). Un seul taux (87,70%, duo (1/4,3/4), $l = 9$) est en deçà de 87,99% (90% - 3 écarts types). Encore une fois, les taux supérieurs sont principalement observés pour les 6 premiers horizons de prévision, particulièrement pour le duo (1/4,3/4).

On peut également remarquer que chacun des taux observés est inférieur à son homologue dans le cas où les paramètres sont connus, à l'exception d'un seul (duo (1/4,3/4), $l = 6$). Finalement, on peut aussi constater que, de façon générale, les taux observés dans le contexte lognormal sont supérieurs à leurs homologues du contexte gamma.

Le cas $\epsilon_t \sim$ bêta modifiée.

Lorsque les paramètres de nuisance sont connus, les taux observés varient entre 87,65% et 92,05% pour le duo (1/2,1/2), entre 89,00% et 92,80% pour le duo (1/4,3/4) et entre 87,95% et 92,30% pour le duo (3/4,1/4). À moins de 2 écarts types de 0,90, on observe 7 taux sur 12 pour le duo (1/2,1/2), 8 taux sur 12 pour le duo (1/4,3/4) et 8 taux sur 12 pour le duo (3/4,1/4). À une distance inférieure à 3 écarts type de 0,90, on retrouve 9 taux sur 12 pour le duo (1/2,1/2), 8 taux sur 12 pour le duo (2/4,3/4) et 10 taux sur 12 pour le duo (3/4,1/4). Contrairement au cas lognormal, lorsqu'on compare les taux observés à leurs homologues du cas gamma, aucun comportement particulier ne peut être remarqué, les taux étant tantôt inférieurs et tantôt supérieurs.

Lorsque les paramètres de nuisance sont estimés, les taux fluctuent entre 87,30% et 91,65% pour le duo (1/2,1/2), entre 87,90% et 92,50% pour le duo (1/4,3/4) et entre 87,35% et 92,25% pour le duo (3/4,1/4). Le nombre de taux situés à moins de 2 écarts types de 0,90 est de 5 sur 12 pour le duo (1/2,1/2), de 6 sur 12 pour le duo (1/4,3/4) et de 8 sur 12 pour le duo (3/4,1/4). À moins de 3 écarts types de 0,90, on peut observer respectivement 9 taux sur 12, 8 taux sur 12 et 10 taux sur 12 selon les duos usuels. Encore une fois, chaque taux, à l'exception d'un seul (duo (1/2,1/2), $l = 5$), est inférieur à son homoloque du contexte où les

paramètres sont connus. On remarque une autre fois que les taux observés pour les 6 premiers horizons sont généralement supérieurs aux taux observés pour les 6 derniers horizons.

À la lumière de ces résultats, il semble que la méthode paramétrique performe relativement bien, dans un contexte unilatéral du moins. Les taux observés varient entre 87,65% et 93,75% lorsque les paramètres de nuisance sont connus (contexte connu) et entre 86,75% et 93,40% lorsque les paramètres de nuisance sont estimés (contexte estimé), soient des valeurs numériquement près de l'objectif visé de 90%. Les résultats sont également plutôt similaires, peu importe la distribution sous-jacente, bien que l'on ait remarqué une tendance à observer des taux légèrement supérieurs dans le cas lognormal. La méthode a mené à des taux non significativement différents de 90% (avec $m = 2000$) dans 59,3% des cas, dans le contexte connu, et dans 52,8% des cas, dans le contexte estimé. Relativement peu de résultats sont en deçà de la borne inférieure de l'intervalle de confiance de niveau 95%, soient 4,6% dans le contexte connu et 21,3% dans le contexte estimé, et encore moins sous 87,99% (90% - 3 écarts types), soient 3,7% dans le contexte connu et 9,3% dans le contexte estimé. On remarque davantage de résultats au-delà de la borne supérieure 91,31%, soient 36,1% dans le cas connu et 25,9% dans le cas estimé, de même qu'au-delà de 92,01%, soient 24,1% dans le contexte connu et 16,7% dans le contexte estimé.

Tel que mentionné précédemment, ces résultats sont en partie dus au fait que les taux sous-jacents ne sont pas exactement de 90%, la distribution de Y_t étant discrète. Il est également important de jeter un coup d'oeil sur l'impact

des paramètres de nuisance et les paramètres de régression. Ces derniers ont un impact majeur sur la distribution de Y_{t+l}. Cet impact est amplifié par la méthode paramétrique car, dans le cadre de celle-ci, ils déterminent complètement la distribution de Y_{t+l}. La difficulté d'établir des intervalles de prévision de niveau fixe pour des variables discrètes réside dans le fait que les masses de probabilité pour les différentes valeurs sont généralement importantes, particulièrement lorsque la variance est petite. Cette situation peut résulter, par exemple, dans le fait que le quantile d'ordre 90% soit également le quantile d'ordre 95%. Lorsque la variance de Y_t est plus grande, les masses de probabilité ont tendance à être moindres en chaque valeur et les quantiles d'ordre distinct ont tendance à se dissocier. C'est ce qui explique, en partie du moins, la différence observée entre les taux des horizons de prévision 1 à 6 et les taux des horizons 7 à 12, μ_t étant inférieur à 2,25 pour les six premiers horizons et supérieur à 2,75 pour les six derniers. C'est ce qui explique aussi le fait qu'on dénote davantage de taux supérieurs pour le duo (1/4,3/4) que pour les deux autres duos.

Nous avons également souligné le fait que l'estimation des paramètres de nuisance avait pour effet la diminution des taux observés. Cette tendance est reliée à la sous-estimation de σ^2 résultant de l'utilisation de l'estimateur des moments, provoquant ainsi une sous-estimation de la variance de Y_t et, par le fait même, la construction d'intervalles de prévision un peu plus courts.

4.3.1.2. *Le cas bilatéral*

Nous poursuivons l'analyse de la méthode paramétrique, cette fois dans un contexte de construction d'intervalles de prévision bilatéraux.

Le cas $\epsilon_t \sim$ gamma.

Lorsque les paramètres de nuisance sont connus, les taux observés varient entre 91,90% et 95,75% pour le duo (1/2,1/2), entre 91,00% et 96,45% pour le duo (1/4,3/4) et entre 93,85% et 96,20% pour le duo (3/4,1/4). Peu de taux sont donc situés à moins de 2 ou 3 écarts types de 0,90.

La situation est semblable dans le cas où les paramètres de nuisance sont estimés, même si cette fois-ci, les taux varient entre 89,15% et 95,45% pour le duo (1/2,1/2), entre 87,90% et 95,90% pour le duo (1/4,3/4) et entre 92,30% et 95,85% pour le duo (3/4,1/4). Les taux inférieurs à 91,31% (0,90 + 2 écarts types) sont observés pour les horizons 8,9,10 et 11 pour les duos (1/2,1/2) et (1/4,3/4). Les autres taux sont supérieurs à 92,01% (0,90 + 3 écarts types), à l'exception d'un seul (duo (1/4,3/4), $l = 7$). Tous les taux sont inférieurs à leurs homologues observés lorsque les paramètres de nuisance sont connus.

Sans pousser plus loin l'analyse, on pourrait être tenté de croire que la méthode paramétrique semble moins propice à la construction d'intervalles bilatéraux. En effet, les intervalles ainsi construits semblent trop conservateurs car les taux s'éloignent de l'objectif visé, soit un taux de couverture de 90%. Toutefois, tel n'est pas le cas. Cette différence trouve sa source dans le choix des valeurs de μ_t. En effet, pour les valeurs choisies, les masses de probabilité en zéro sont importantes et il en résulte que les quantiles d'ordre 5% sont presque tous égaux à zéro. On observe ainsi des taux gonflés de 5%. Les deux tableaux qui suivent indiquent les taux sous-jacents pour chaque horizon de prévision et pour chaque combinaison étudiée. Rappelons encore une fois que ces taux constituent les taux de couverture lorsque tous les paramètres, aussi bien les paramètres de nuisance

que les paramètres de régression, sont connus et lorsque la distribution marginale de ϵ_t est effectivement gamma.

TABLEAU 4.3.3. *Taux de couverture sous-jacents de l'intervalle bilatéral, en pourcentage, selon la méthode paramétrique, pour les temps 101 à 106.*

(σ^2, ϕ)	101	102	103	104	105	106
$(1/2, 1/2)$	95,09	96,61	95,99	96,13	96,97	95,92
$(1/4, 3/4)$	97,02	95,17	97,34	97,44	95,68	95,05
$(3/4, 1/4)$	95,85	95,45	97,33	95,04	95,87	96,54

TABLEAU 4.3.4. *Taux de couverture sous-jacents de l'intervalle bilatéral, en pourcentage, selon la méthode paramétrique, pour les temps 107 à 112.*

(σ^2, ϕ)	107	108	109	110	111	112
$(1/2, 1/2)$	93,01	92,19	90,77	91,08	90,20	90,92
$(1/4, 3/4)$	95,53	95,14	91,78	91,86	95,79	96,44
$(3/4, 1/4)$	95,57	95,55	95,42	95,60	96,04	96,32

Pour des valeurs de μ_t plus grandes, il est fort probable que des taux similaires auraient été observés dans le cas unilatéral et dans le cas bilatéral. Cette affirmation est d'ailleurs confirmée par les taux observés aux horizons 8, 9, 10 et 11 pour les duos (1/2,1/2) et (1/4,3/4). Pour ces cas, les masses de probabilité en zéro varient entre 4% et 11%, alors que dans les autres cas, elles sont presque toutes supérieures à 15%, certaines atteignant même 40%.

Le cas $\epsilon_t \sim$ lognormale.

Lorsque les paramètres de nuisance sont connus, les taux observés varient entre 93,60% et 96,00% pour le duo (1/2,1/2), entre 90,85% et 96,90% pour le duo (1/4,3/4) et entre 94,20% et 96,40% pour le duo (3/4,1/4). Nous n'identifierons pas le nombre de taux significativement différents de 0,90, la discussion précédente justifiant la non-pertinence d'une telle démarche. Remarquons simplement que les taux observés pour les horizons 8,9,10 et 11 pour le duo (1/4,3/4) sont similaires à ceux obtenus dans le contexte unilatéral. Notons finalement que la tendance observée dans le contexte unilatéral (des taux observés supérieurs lorsque $\epsilon_t \sim$ lognormale par rapport aux taux observés lorsque $\epsilon_t \sim$ gamma) est moins présente dans le contexte bilatéral.

Lorsque les paramètres de nuisance sont estimés, les taux observés varient entre 90,80% et 95,75% pour la combinaison (1/2,1/2), entre 87,60% et 96,30% pour la combinaison (1/4,3/4) et entre 91,70% et 96,20% pour la combinaison (3/4,1/4). On peut encore constater que, pour les horizons 8, 9, 10 et 11 des duos (1/2,1/2) et (1/4,3/4), les taux observés sont similaires à ceux observés dans le contexte unilatéral. Finalement, notons que tous les taux sont inférieurs à leurs homologues observés lorsque les paramètres de nuisance sont connus.

Le cas $\epsilon_t \sim$ bêta modifiée.

Avec les paramètres de nuisance connus, les taux observés fluctuent entre 91,80% et 96,45% pour le duo (1/2,1/2), entre 88,95% et 96,60% pour le duo (1/4,3/4) et entre 92,55% et 96,45% pour le duo (3/4,1/4). Encore une fois,

les taux observés pour les horizons 8, 9, 10 et 11 pour le duo (1/4,3/4) sont comparables à ceux observés dans le contexte unilatéral.

Lorsque les paramètres de nuisance sont estimés, les taux observés varient entre 88,25% et 95,80% pour le duo (1/2,1/2), entre 87,00% et 96,25% pour le duo (1/4,3/4) et entre 91,70% et 96,20% pour le duo (3/4,1/4). Les taux observés pour les horizons 8, 9, 10 et 11 des duos (1/2,1/2) et (1/4,3/4) sont similaires aux taux observés dans le contexte unilatéral. On remarque aussi que tous les taux sont inférieurs à leurs homologues observés lorsque les paramètres de nuisance sont connus.

À la lumière de ces résultats, il semble que la méthode paramétrique de construction d'intervalles de prévision dans le contexte bilatéral performe aussi bien que dans le contexte unilatéral, dans la mesure où les masses de probabilité pour les petites valeurs de Y_t sont suffisamment petites pour que le quantile d'ordre 5% soit différent de 0 et qu'il puisse être distingué du quantile d'ordre 1%. Malheureusement, pour les valeurs choisies pour μ_t, de telles conditions ont été rarement rencontrées. Remarquons finalement l'estimation des paramètres de nuisance a eu pour effet la diminution des taux observés. Cette tendance est également reliée à la sous-estimation de σ^2 provenant de l'utilisation de l'estimateur des moments, provoquant ainsi la construction d'intervalles de prévision un peu plus courts.

4.3.1.3. *Analyse de l'hypothèse «ϵ_t est marginalement gamma »*

Ici, nous cherchons à pousser un peu plus loin notre quête de l'effet causé par l'hypothèse «ϵ_t est marginalement gamma ». Nous chercherons donc à savoir s'il existe une différence significative entre le taux observé lorsque ϵ_t est effectivement gamma et le taux observé lorsque ϵ_t est lognormale ou bêta modifiée, toute chose étant égale par ailleurs, à l'aide du test (4.3.2) et du test (4.3.3).

Le cas unilatéral.

Il semble plutôt rare d'observer une différence significative entre les deux taux. Aussi bien dans le cas où les paramètres de nuisance sont connus que dans le cas où ils sont estimés, aussi bien pour ϵ_t lognormal que pour ϵ_t bêta modifiée, les résultats du test (4.3.2) et du test (4.3.3) démontrent qu'un seul horizon de prévision sur 12 ou qu'aucun horizon de prévision sur 12 selon le cas n'a permis d'observer des taux significativement différents au niveau 5%.

Le cas bilatéral.

Les différences significatives ne sont pas plus fréquentes. Aussi bien dans le cas où les paramètres de nuisance sont connus que dans le cas où ils sont estimés, aussi bien pour ϵ_t lognormal que pour ϵ_t bêta modifiée, les résultats du test (4.3.2) et du test (4.3.3) démontrent qu'un seul horizon de prévision sur 12 ou qu'aucun horizon de prévision sur 12 selon le cas n'a permis d'observer des taux significativement différents au niveau 5%. En une seule occasion, soit pour la combinaison $(1/2, 1/2)$ estimée et pour $\epsilon_t \sim$ bêta modifiée, deux différences significatives sur 12 ont pu être enregistrées par le test (4.3.3).

Ces résultats sont très encourageants. Ils viennent éliminer le doute subsistant concernant les taux observés avec $\epsilon_t \sim$ lognormale. Ils permettent de croire que l'hypothèse à la base de la méthode paramétrique, soit «ϵ_t est marginalement gamma », n'est pas restrictive. Elle est suffisamment souple pour permettre à la méthode d'obtenir de bons résultats même lorsqu'elle n'est pas vérifiée. Bizarrement, une des raisons de cette souplesse est le grand nombre de contraintes dans le modèle. Même si ϵ_t n'admet pas marginalement la même distribution, les contraintes font en sorte que les distributions sont similaires. Les graphiques de la page suivante sont révélateurs à cet effet. Et comme l'impact de ϵ_t sur Y_t est dilué par le fait que $Y_t|\epsilon_t$ soit aléatoire, et non pas déterministe, les petites différences observées au niveau des distributions sont moindres de conséquences.

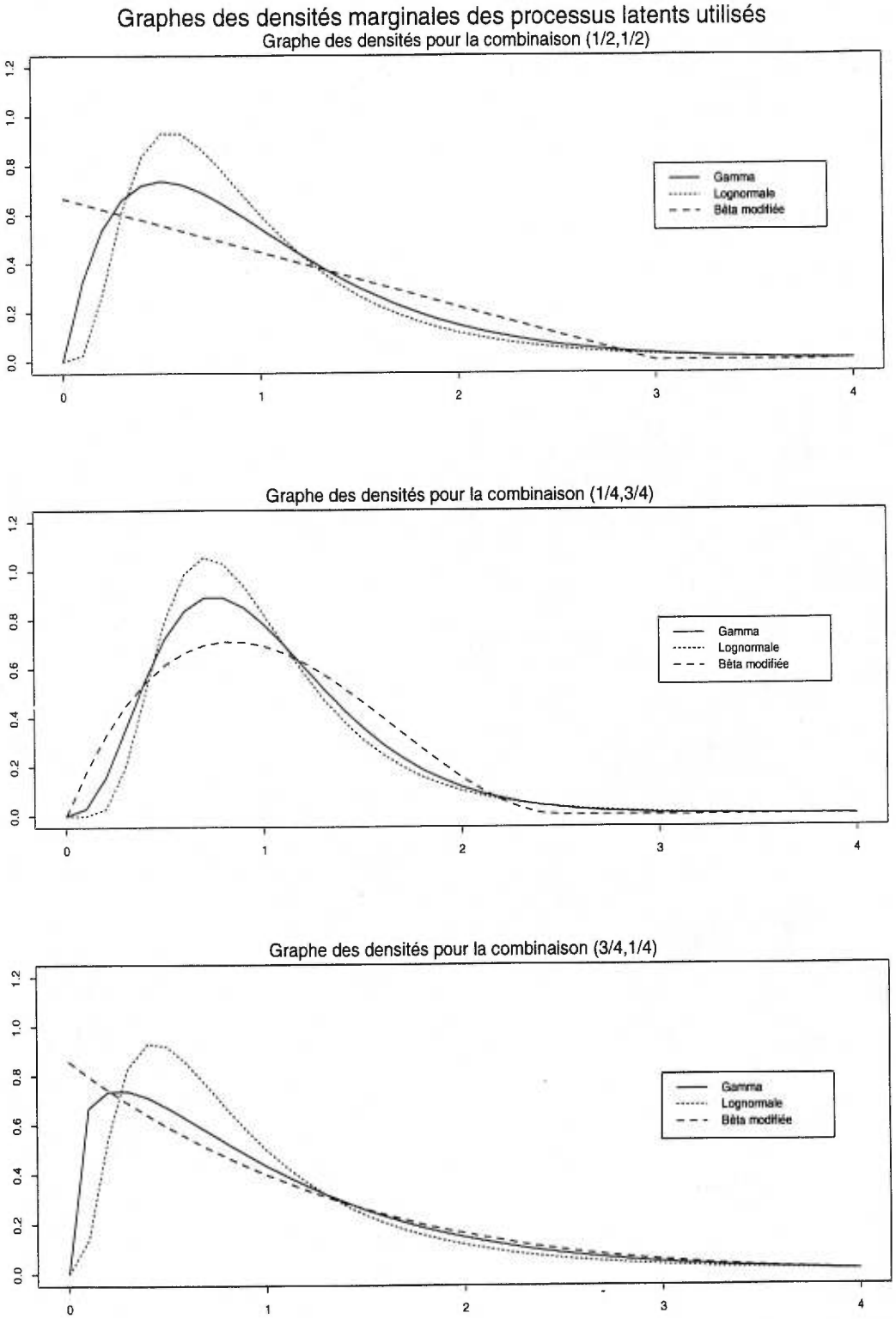

FIGURE 4.3.1. *Graphiques comparatifs des densités marginales pour* $\{\epsilon_t\}$ *selon la combinaison* (σ^2, ϕ).

4.3.2. Analyse de la méthode non paramétrique

Rappelons d'abord que la méthode paramétrique consiste à appliquer une transformation stabilisant la variance du processus $\{Y_t\}$ et à utiliser les différences $\delta_t = T(y_t) - T(\hat{\mu}_t)$. Sous les hypothèses du chapitre 2, le processus $\{\delta_t\}$ est approximativement stationnaire en moyenne et en variance et les δ_t sont approximativement identiquement distribués. Ceci laisse croire que les intervalles de prévision pour Y_{n+l} donnés par

$$Y_{n+l} \in \left(T^{-1}(q_{\delta;0.05} + T(\hat{\mu}_{n+l})), T^{-1}(q_{\delta;0.95} + T(\hat{\mu}_{n+l}))\right) ,$$

dans le cadre bilatéral, et par

$$Y_{n+l} \in \left[0 , T^{-1}(q_{\delta;0.90} + T(\hat{\mu}_{n+l}))\right) ,$$

dans le cadre unilatéral, sont de niveau approximatif 90%. Pour vérifier l'efficacité de la méthode, nous confronterons les hypothèses

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p \neq p_0 ,$$

où $p_0 = 0,90$. Il s'agit encore une fois du test classique d'une proportion. Nous utiliserons donc le test (4.3.1) ainsi que les bornes présentées à la section précédente.

4.3.2.1. *Le cas unilatéral*

Débutons par l'analyse de la méthode non paramétrique dans le cadre de la construction d'intervalles de prévision unilatéraux. Puisqu'il s'agit d'une méthode non-paramétrique, nous n'analyserons pas les résultats de façon séparée selon les

distributions. Nous décomposerons plutôt l'analyse en deux temps, selon que les paramètres de nuisance sont connus ou estimés.

Paramètres de nuisance connus.

Dans ce contexte, les taux observés varient entre 84,65% et 88,95% pour le duo (1/2,1/2), entre 84,65% et 89,25% pour le duo (1/4,3/4) et entre 86,70% et 89,10% pour le duo (3/4,1/4). Bien que plusieurs taux soient situés à plus de 2 ou 3 écarts types de 0,90, les taux observés sont tout de même numériquement près de l'objectif visé de 90%, le taux minimal observé étant de 84,65%. Un fait intéressant à noter est la grande similitude entre les taux par rapport aux horizons de prévision. Pour le duo (3/4,1/4), l'écart entre le taux maximal et le taux minimal est de moins de 2,5%. Pour les duos (1/2,1/2) et (1/4,3/4), l'écart est de l'ordre de 4,5%.

Paramètres de nuisance estimés.

Dans le contexte où les paramètres de nuisance sont estimés, les résultats sont très semblables, bien que légèrement supérieurs. Cette fois-ci, les taux fluctuent entre 84,90% et 89,55% pour la combinaison (1/2,1/2), entre 84,95% et 90,35% pour la combinaison (1/4,3/4) et entre 86,75% et 89,45% pour la combinaison (3/4,1/4). Le nombre de taux situés à moins de 3 écarts types de 0,90 varient de 4 à 8 sur 12 selon les duos. Lorsque ce nombre est de 4 ou 5, il s'agit généralement du duo (1/4,3/4). Lorsque ce nombre est de 7 ou 8, il s'agit plutôt du duo (3/4,1/4). On remarque aussi à peu près la même tendance que dans le cas où les paramètres de nuisance sont connus, soient une grande similitude des taux par rapport aux horizons pour le duo (3/4,1/4) et des taux un peu plus épars pour les duos (1/2,1/2) et (1/4,3/4). Un autre élément important à noter est le fait que, dans

ce contexte, les taux ont tendance à être supérieurs ou égaux à leurs homologues du contexte où les paramètres de nuisance sont connus.

Bien que la plupart des taux observés soient inférieurs à 90%, la méthode non paramétrique performe relativement bien, du moins dans le cadre unilatéral. Dans le contexte où les paramètres de nuisance sont estimés, c'est-à-dire dans la situation réelle, les taux observés varient entre 84,90% et 90,35%, des taux numériquement près de l'objectif visé de 90%. Compte tenu des approximations et des hypothèses faites lors de l'établissement de la méthode, il s'agit de résultats intéressants. En ce qui a trait aux paramètres de régression, leur impact est moindre que dans le cadre de la méthode paramétrique car la transformation T qui stabilise la variance de Y_t permet de tenir compte de μ_t. Ceci explique en partie la similitude entre les taux par rapport aux horizons de prévision. Quant aux paramètres de nuisance, leur impact est relativement important. Toutefois, ici, c'est le coefficient de corrélation ϕ qui a davantage d'impact que la variance, car la transformation T utilise σ^2 comme paramètre, ce qui en réduit grandement l'influence. Nous avons déjà soulevé le fait que les différences δ_t étaient corrélés et que leur structure de corrélation était semblable à celle des observations (Voir équation (2.2.1)). Lorsque la corrélation entre les observations $Y_1, \ldots, Y_n$ est faible, la corrélation entre les différences δ_t l'est également. On observe alors des taux similaires et près de 90%, comme pour le duo (3/4,1/4). Lorsque la corrélation est plus forte, les différences δ_t sont également plus fortement corrélés et on peut alors observer des taux un peu moins bons et légèrement moins similaires, comme pour le duo (1/4,3/4). Finalement, en ce qui concerne les taux supérieurs observés dans le contexte où les paramètres de nuisance sont estimés, ce phénomène n'est pas attribuable à la sous-estimation de la variance, comme

dans le cas paramétrique, mais plutôt à une plus grande variance dans l'estimation des paramètres de régression et, par le fait même, dans les quantiles q_δ, le tout résultant dans la formation d'intervalles un peu plus larges.

4.3.2.2. *Le cas bilatéral*

Nous poursuivons l'analyse de la méthode non paramétrique dans le cadre de la construction d'intervalles de prévision bilatéraux.

Paramètres de nuisance connus.

À première vue, les résultats semblent moins bons que dans le contexte unilatéral. On remarque d'abord que les taux varient beaucoup plus dans ce contexte. En fait, les taux fluctuent entre 79,05% et 92,55% pour le duo (1/2,1/2), entre 79,50% et 87,45% pour le duo (1/4,3/4) et entre 80,70% et 93,95% pour le duo (3/4,1/4). On peut également remarquer qu'on observe des taux plus élevés pour les 6 premiers horizons de prévision et des taux moins élevés pour les 6 derniers horizons, particulièrement pour les duos (1/2,1/2) et (3/4,1/4). Ce comportement est visible bien que moins net pour le duo (1/4,3/4).

Paramètres de nuisance estimés.

Lorsque les paramètres de nuisance sont estimés, les taux fluctuent tout autant. Cette fois, les taux observés varient entre 79,85% et 93,50% pour le duo (1/2,1/2), entre 81,00% et 89,65% pour le duo (1/4,3/4) et entre 80,90% et 94,35% pour le duo (3/4,1/4). On peut encore remarquer des taux généralement plus élevés pour les 6 premiers horizons et moins élevés pour les 6 derniers. À l'exception de trois d'entre eux, tous les taux sont supérieurs à leurs homologues du contexte où les paramètres de nuisance sont connus.

Les résultats de la méthode non paramétrique sont plus décevants dans le cadre bilatéral. Les taux varient énormément: l'écart entre le taux minimal et le taux maximal est de l'ordre de 10% à 12%. La plupart des taux observés sont à plus de 2 ou 3 écarts types de 0,90. La principale cause de cette performance plus mitigée est plutôt complexe. Pour en faciliter l'explication, nous nous servirons d'un seul cas: $\epsilon_t \sim$ gamma, duo (1/2,1/2), paramètres de nuisance estimés. Il est important de noter toutefois que le comportement décrit ci-après est similaire aux comportements observés dans les autres cas. Les graphiques des pages suivantes présentent les diagrammes en boîte des différences δ_t ainsi que l'histogramme du compte des provenances des quantiles d'ordre 5% $q_{\delta;0.05}$ selon la famille d'appartenance pour les 2000 séries simulées dans le cas choisi. La famille d'appartenance est constituée de tous les entiers entre 1 et 100 partageant le même modulo lorsque divisé par 12. Elles ont été identifiées par les valeurs de t de 101 à 112 pour faciliter le lien avec les horizons de prévision et les taux. En jetant un coup d'oeil aux diagrammes en boîte, on dénote la grande instabilité de δ_t par rapport à la famille d'appartenance au niveau des petites valeurs. On y constate en effet que les petites valeurs sont très différentes selon la famille. Alors que pour les familles 102, 103, 104 et 105, les valeurs inférieures à -3 sont presque inexistantes, de nombreuses différences de cet ordre sont observées pour les familles 108, 109, 110 et 111. Quant à l'histogramme, il nous apprend que la majorité des quantiles d'ordre 5% proviennent des familles 101 à 106. Voilà donc en partie pourquoi on observe des taux bien différents selon l'horizon de prévision, et plus particulièrement des taux plus élevés pour les horizons 2, 3, 4 et 5 et des taux moins élevés pour les horizons 8, 9, 10 et 11.

Parmi les autres causes, notons les approximations faites pour établir la méthode ainsi que l'influence de la corrélation qui joue un rôle important, tout comme dans le contexte unilatéral. Finalement, en ce qui concerne les taux supérieurs observés dans le contexte où les paramètres de nuisance sont estimés, par rapport aux taux homologues du contexte où les paramètres de nuisance sont connus, ce phénomène s'explique de la même façon que dans le cas unilatéral, soit par une plus grande variance dans l'estimation des paramètres de régression et, par le fait même, dans les quantiles q_{δ_t}, le tout résultant dans la formation d'intervalles un peu plus larges.

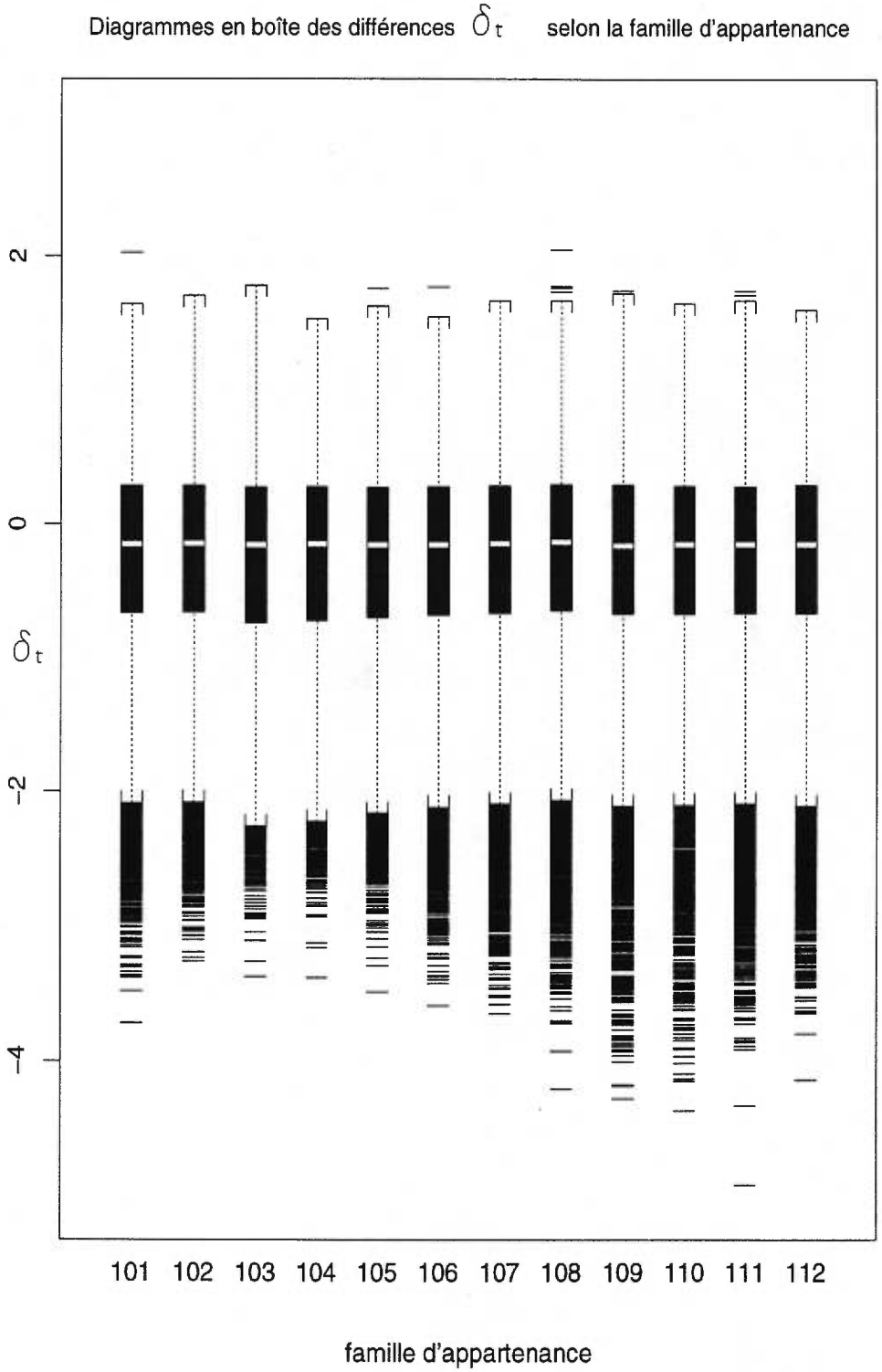

FIGURE 4.3.2. *Diagrammes en boîte des différences δ_t selon la famille d'appartenance pour les 2000 séries simulées pour le cas $\epsilon_t \sim$ gamma, duo (1/2,1/2) estimé.*

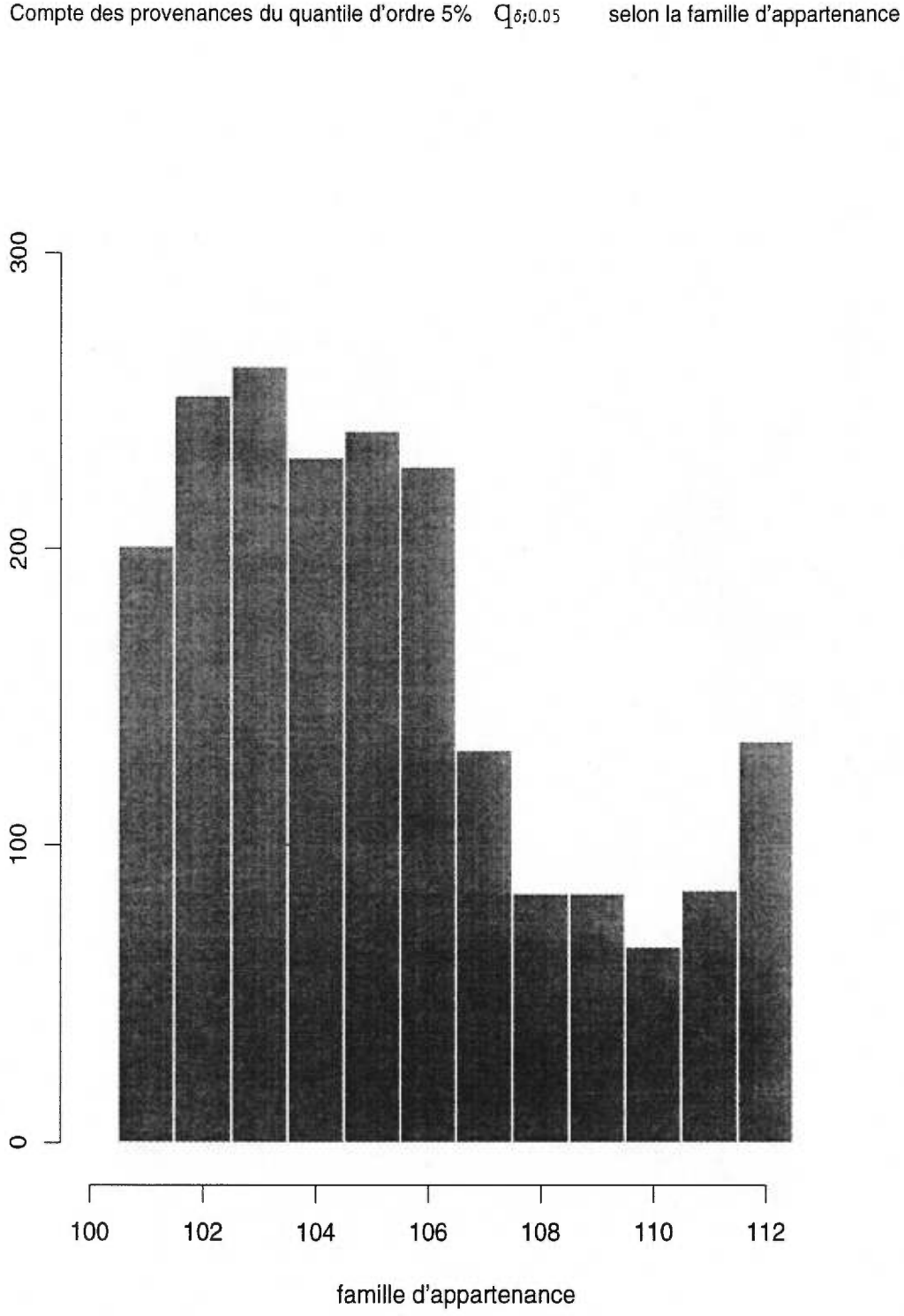

FIGURE 4.3.3. *Histogramme du compte des provenances du quantile $q_{\delta_t;0.05}$ selon la famille d'appartenance pour les 2000 séries simulées pour le cas $\epsilon_t \sim$ gamma, duo (1/2,1/2) estimé.*

4.4. ANALYSE COMPARATIVE

Cette section est consacrée à l'analyse comparative des deux méthodes de construction d'intervalles de prévision proposées au chapitre 2. Compte tenu des résultats plus mitigés de la méthode non paramétrique dans le contexte bilatéral, nous limiterons cette étude au contexte unilatéral. Une façon intéressante de comparer les deux méthodes est d'analyser le rapport de la longueur de l'intervalle paramétrique à la longueur de l'intervalle non paramétrique. Puisqu'il s'agit d'intervalles unilatéraux, ce rapport est également celui de la borne supérieure de l'intervalle paramétrique à la borne supérieure de l'intervalle non paramétrique. Nous avons choisi d'étudier ce rapport pour l'horizon de prévision 1, étant donné la similitude relative des taux observés selon les horizons et étant donné que c'est l'horizon le plus souvent sollicité. Nous avons également restreint l'étude comparative au cas où les paramètres sont estimés car il s'agit de la situation rencontrée lors de l'analyse de séries réelles.

Les graphiques de la page 88 présentent les diagrammes en boîte du rapport de la longueur de l'intervalle paramétrique à la longueur de l'intervalle non paramétrique pour l'horizon de prévision 1 selon la distribution marginale de ϵ_t et selon les duos (σ^2, ϕ) estimés. On peut d'abord y remarquer que les médianes sont très près de 1 et que les distances interquartiles (les hauteurs des boîtes) sont petites. On peut en conclure que, dans la majorité des cas, les intervalles de prévision coïncident ou sont d'une longueur similaire. On peut également noter qu'un plus grand nombre de valeurs à l'écart est observé au-dessus de la moustache supérieure qu'en-dessous de la moustache inférieure. Les intervalles unilatéraux paramétriques sont donc occasionnellement plus longs que les intervalles unilatéraux non paramétriques, alors que l'inverse est très rare. Voilà sans

doute pourquoi on a observé des taux légèrement supérieurs à ceux de la méthode non paramétrique avec la méthode paramétrique. Il n'en demeure pas moins que, de façon générale, les deux méthodes établissent des bornes comparables.

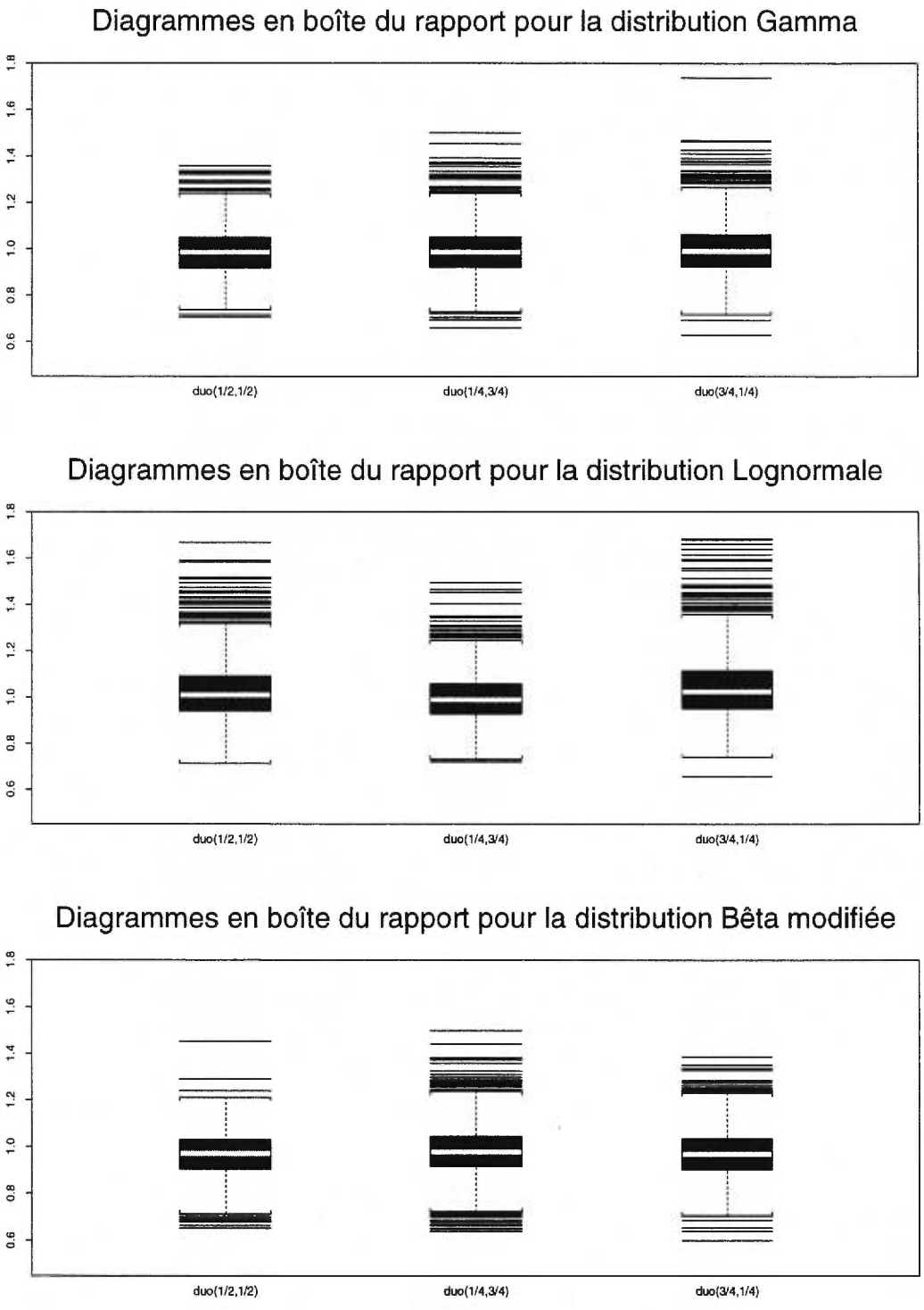

FIGURE 4.4.1. *Diagrammes en boîte du rapport de la longueur de l'intervalle paramétrique à la longueur de l'intervalle non paramétrique pour l'horizon 1 selon la distribution marginale de ϵ_t et selon les duos (σ^2, ϕ) estimés.*

Chapitre 5

APPLICATIONS À DES SÉRIES RÉELLES

Dans ce chapitre, nous illustrerons l'utilité des méthodes développées précédemment en les appliquant à un domaine particulier: la surveillance de maladies infectueuses. L'un des objectifs de la surveillance de maladies infectueuses est la détection des périodes épidémiques. Lorsqu'une incidence élevée est observée, il est important de pouvoir déterminer s'il s'agit du résultat d'une véritable épidémie ou simplement d'une fluctuation aléatoire. Les deux techniques de construction d'intervalles de prévision unilatéraux fournissent un point de comparaison, auquel est associé un degré de confiance. Il semble donc s'agir d'outils tout désignés à cette fin.

Nous analyserons donc trois séries d'incidence de maladies infectueuses. La première est devenue une série classique dans le domaine relativement récent des modèles dictés par les états. Il s'agit de la série d'incidence mensuelle de la poliomyélite aux États-Unis entre 1970 et 1983, étudiée d'abord par Zeger (1988). Nous profiterons de cette série largement étudiée pour comparer les résultats obtenus par notre méthode d'estimation des paramètres de régression à ceux obtenus par d'autres méthodes. Ensuite, nous analyserons des séries d'incidence de maladies infectueuses, plus particulièrement l'hépatite B et la coqueluche. Ces données proviennent du Bureau régional des maladies infectueuses de la région de Montréal

et concernent la période s'écoulant du 1^{er} janvier 1986 au 31 décembre 1993. Elles ont déjà été étudiées par Cardinal (1995), avec des modèles INAR toutefois. Pour ces deux maladies, une série $\{Y_t\}$ a été construite en considérant le nombre de cas déclarés par période de 28 jours. Puisque 13 périodes de 28 jours comportent 364 jours, le dernier jour (les deux derniers dans le cas d'une année bissextile) a(ont) été assigné(s) à la dernière période. Les séries ainsi construites sont donc de longueur 104. Nous utiliserons les 91 premières observations pour l'estimation et nous nous servirons des treize dernières pour évaluer les performances des deux méthodes dans un contexte réel.

5.1. Série d'incidence de la poliomyélite

Dans le domaine des modèles dictés par les états, la série de référence est la série d'incidence mensuelle de la poliomyélite aux États-Unis entre 1970 et 1983. Elle fut étudiée en premier lieu par Zeger (1988). Afin de modéliser adéquatement cette série par l'approche des modèles dictés par les états, Zeger (1988) a proposé le modèle suivant:

$$\log(\mu_t) = \beta_0 + \beta_1 \frac{t}{1000} + \beta_2 cos(\frac{2\pi t}{12}) + \beta_3 sin(\frac{2\pi t}{12}) + \beta_4 cos(\frac{4\pi t}{12}) + \beta_5 sin(\frac{4\pi t}{12}) . \tag{5.1.1}$$

Depuis, plusieurs chercheurs ont étudié cette série en proposant différentes méthodes d'estimation. Dans le tableau qui suit, nous présentons quelques-unes de ces études, en rapportant les valeurs estimées pour les paramètres de régression ainsi que l'erreur type associée. Il est important de noter qu'il ne s'agit pas d'une liste exhaustive.

TABLEAU 5.1.1. *Paramètres de régression et erreur type associées à la série d'incidence mensuelle de la poliomyélite selon la méthode utilisée.*

	Chercheurs	Zeger	Chan et Ledolter	Kuk et Cheng	Jorgensen et al	Lafortune et Roy
	Méthode	Équ. d'estimation approx.	Algorithme EM type Monte Carlo	Newton-Raphson type Monte Carlo	Algorithme EM avec Itérations de Kalman	Équ. d'estimation
PARAMÈTRES	β_0	0,17 (0,13)	0,42 (0.125)	0,243 (0,278)	0,5601 (0,0011)	0,210 (0,133)
	β_1	-4,35 (2,68)	-4,62 (1.38)	-3,81 (2,83)	-1,614 (0,018)	-3,828 (2,663)
	β_2	-0,11 (0,16)	0,15 (0,09)	0,161 (0,145)	0,1328 (0,000052)	-0,134 (0,165)
	β_3	-0,48 (0,17)	-0,50 (0,12)	-0,481 (0,165)	-0,5205 (0,000075)	-0,487 (0,173)
	β_4	0,20 (0,14)	0,44 (0,10)	0,413 (0,127)	0,4525 (0,000057)	0,172 (0,144)
	β_5	-0,41 (0,14)	-0,04 (0,10)	-0,011 (0,125)	-0,0652 (0,000057)	-0,414 (0,146)

De façon générale, on peut constater que la plupart des résultats sont relativement similaires, à l'exception des erreurs types fournies par Jorgensen, Lundbye-Christensen, Song et Sun (1995) qui semblent très petites lorsqu'on les compare aux résultats obtenus par les autres méthodes. On peut également remarquer que nos résultats sont très similaires à ceux obtenues par Zeger (1988). Il s'agit d'une similitude peu surprenante compte tenu que, dans les deux cas, la méthode d'équations d'estimation a été utilisée. La différence réside dans le fait que Zeger (1988) a utilisé une approximation de la matrice de covariance $\mathbf{V}_n$ plutôt que la matrice $\mathbf{V}_n$ elle-même.

Dans son article, Zeger (1988) mentionne que l'observation du mois de novembre 1972 (14) constitue une valeur à l'écart, mais qu'une indicatrice n'a pas été incluse au modèle compte tenu du peu d'impact de cette observation sur les résultats. Nous avons voulu vérifier cette affirmation. Nous avons donc ajouter une variable indicatrice $I_{Nov.1972}$ au modèle (5.1.1). Les paramètres de régression et de nuisance résultant de cet ajout sont présentés dans le tableau qui suit.

TABLEAU 5.1.2. *Paramètres de régression et de nuisance pour le modèle (5.1.1) augmenté de l'indicatrice* $I_{Nov.1972}$.

β_0	β_1	β_2	β_3	β_4	β_5	$\beta_{Nov.1972}$	σ^2	$\rho_\epsilon(1)$
0,179	-2,581	-0.155	-0.424	0.215	-0.342	1,701	0,367	0,756
(0,148)	(2,883)	(0,129)	(0,140)	(0,112)	(0,113)	(0,554)		

Effectivement, peu de changements sont notables au niveau des paramètres de régression. Seul le coefficient de la tendance linéaire (β_1) subit une modification relativement importante, passant de -3,828 à -2,581. L'introduction de la variable indicatrice dans le modèle a toutefois un impact majeur sur les estimations des

paramètres de nuisance. Alors que dans le modèle (5.1.1), les estimations respectives de σ^2 et de $\rho_\epsilon(1)$ obtenus par notre méthode étaient de 0,807 et de 0,419, elles sont respectivement de 0,367 et de 0,756 dans le modèle avec indicatrice. Il faudrait donc faire preuve d'un peu plus de prudence lorsqu'on affirme que l'observation du mois de novembre 1972 a peu d'impact sur les résultats. On pourrait plutôt mentionner que cette observation a peu d'impact sur les résultats des estimations des paramètres de régression. Nous avons jugé bon d'inclure, à la page suivante, les graphiques de la série de polio, de notre modélisation initiale et de notre modélisation avec indicatrice.

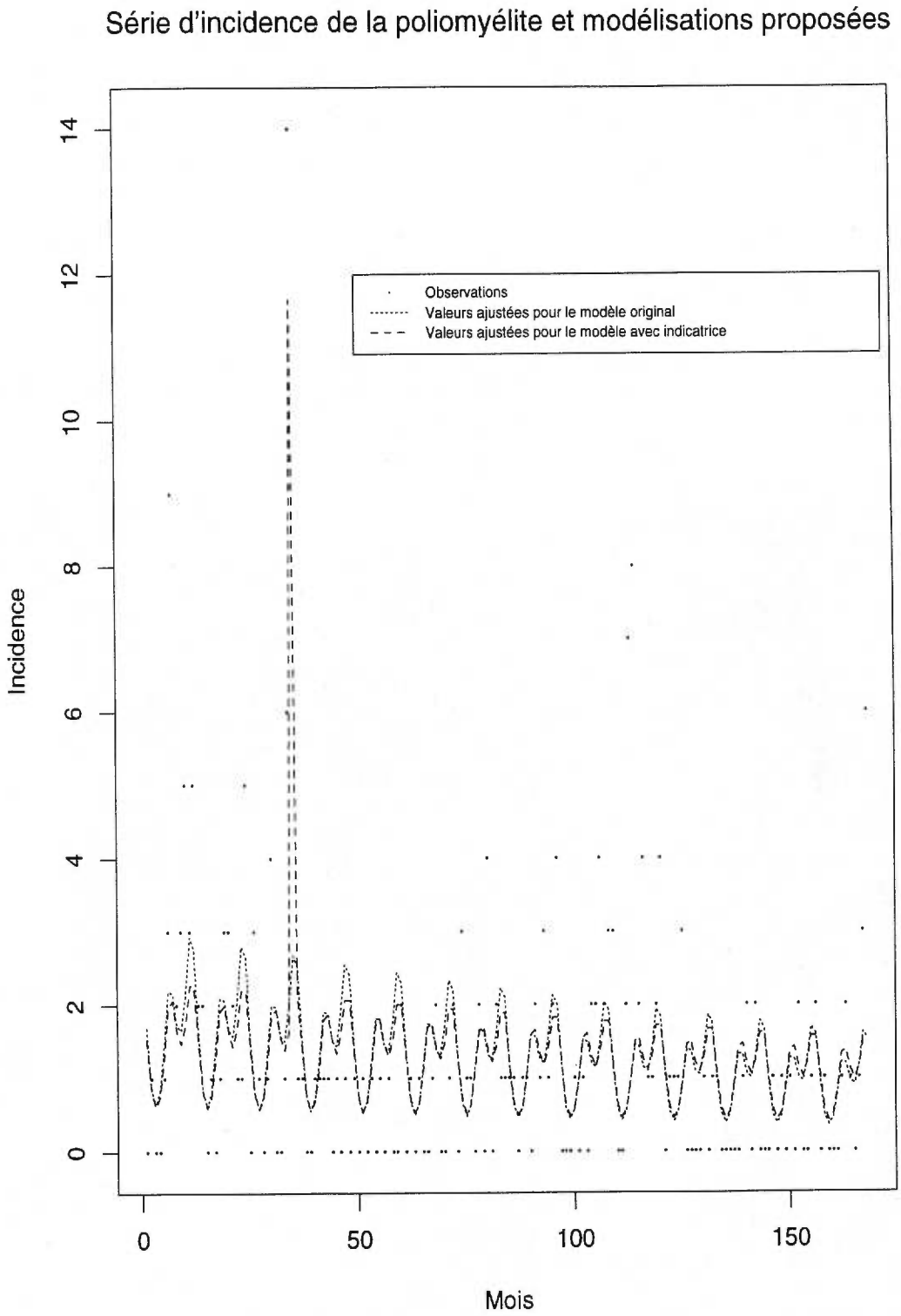

FIGURE 5.1.1. *Graphique de la série d'incidence de la poliomyélite, ainsi que les modélisations selon le modèle (5.1.1) et selon le modèle (5.1.1) augmenté de l'indicatrice* $I_{Nov.1972}$.

5.2. Série d'incidence de l'hépatite B

La série d'incidence de l'hépatite B est illustré à la page 98. On peut y remarquer la présence de deux périodes possiblement épidémiques. La première période, elle-même constituée de deux petites périodes, a eu lieu au cours de l'année 1987 ($t = 20, 23, 24, 25$). La deuxième période, plus courte, s'est produite au début de l'année 1989 ($t = 40, 41$). Afin de modéliser cette série, nous avons choisi d'inclure les composantes usuelles: un terme constant, pour prendre en considération la moyenne de la série, une tendance linéaire ($t/91$), car on observe une tendance à la hausse, et des composantes annuelles et semi-annuelles car la propagation des maladies est souvent reliée aux saisons. Nous avons également choisi d'inclure sept variables indicatrices: $I_t^*, I_{t_{20}}, I_{t_{23}}, I_{t_{24}}, I_{t_{25}}, I_{t_{40}}, I_{t_{41}}$. La première indicatrice, définie par

$$I_t^* = \begin{cases} 1 & , \text{ si } t \geq 53, \\ 0 & , \text{ sinon,} \end{cases}$$

a été introduite dans le modèle pour tenir compte d'une définition plus stricte des cas et d'un changement dans la collecte des données à partir de janvier 1989 (Voir Cardinal (1995), p. 58). Les six autres indicatrices ont la forme suivante:

$$I_{t_n} = \begin{cases} 1 & , \text{ si } t = n, \\ 0 & , \text{ sinon.} \end{cases}$$

Elles ont pour but de prendre en considération les valeurs extrêmes observées aux temps possiblement épidémiques ($t = 20, 23, 24, 25, 40, 41$) et d'éviter que ces données ne faussent le modèle. Elles ont été préférées à deux indicatrices de période épidémique, étant donné la grande variation des valeurs à l'intérieur d'une même période épidémique.

Le modèle proposé a donc l'allure suivante:

$$\log(\mu_t) = \beta_0 + \beta_1 \frac{t}{91} + \beta_2 cos(\frac{2\pi t}{13}) + \beta_3 sin(\frac{2\pi t}{13}) + \beta_4 cos(\frac{4\pi t}{13}) + \beta_5 sin(\frac{4\pi t}{13}) + \beta_6 I_t^* + \beta_7 I_{t_{20}} + \beta_8 I_{t_{23}} + \beta_9 I_{t_{24}} + \beta_{10} I_{t_{25}} + \beta_{11} I_{t_{40}} + \beta_{12} I_{t_{41}} . \tag{5.2.1}$$

Le tableau suivant présente les estimations des paramètres de régression et de nuisance pour le modèle (5.2.1) ajusté à la série d'incidence de l'hépatite B. Il est important de rappeler que les estimations sont celles obtenues à partir des 91 premières observations, les 13 dernières étant réservées à l'appréciation des deux méthodes de construction d'intervalles de prévision dans un contexte réel.

TABLEAU 5.2.1. *Paramètres de régression et de nuisance pour le modèle (5.2.1) ajusté à la série d'incidence de l'hépatite B.*

β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
3,353	0,754	-0,072	-0,064	-0,068	-0,035	-1,067	0,539
(0,067)	(0,191)	(0,041)	(0,041)	(0,040)	(0,041)	(0,112)	(0,223)
β_8	β_9	β_{10}	β_{11}	β_{12}	σ^2	$\rho_\epsilon(1)$	
0,330	0,405	0,507	0,980	0,533	0,028	0,035	
(0,226)	(0,227)	(0,226)	(0,226)	(0,213)			

Le graphique de la série et des valeurs ajustées est présenté à la page 98, en compagnie des graphiques des résidus de Pearson du modèle et des autocorrélations de ces mêmes résidus. Rappelons simplement que le résidu de Pearson correspondant à la t^e observation est défini par

$$r_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\hat{v}_{tt}}},$$

où $\hat{v}_{tt}$ est la variance estimée de Y_t. Grâce aux différents graphiques, on peut constater que ces résidus varient environ entre -3 et 3, de part et d'autre de zéro et qu'ils se comportent comme un bruit blanc. Le modèle ajusté semble donc adéquat.

Le graphique des valeurs futures $t = 92, \ldots, 104$, associées aux horizons de prévision $l = 1, \ldots, 13$, des prévisions et des limites des intervalles de prévision paramétriques et non paramétriques est présenté à la page 99. On peut y constater qu'aucune période d'épidémie ne s'est produite au cours de l'année 1993 ($l = 1, \ldots, 13$). Quant aux valeurs futures elles-même, elles sont toutes inférieures aux bornes respectives des intervalles de prévision paramétriques et non paramétriques, à l'exception d'une seule. En effet, pour l'horizon $l = 12$ ($t = 103$), la valeur future observée est 30. La borne de l'intervalle paramétrique est de 30 également, alors que la borne de l'intervalle non paramétrique est 29,87. Ainsi construit, notre modèle aurait donc correctement conclu à la non-présence de périodes épidémiques au cours de l'année 1993.

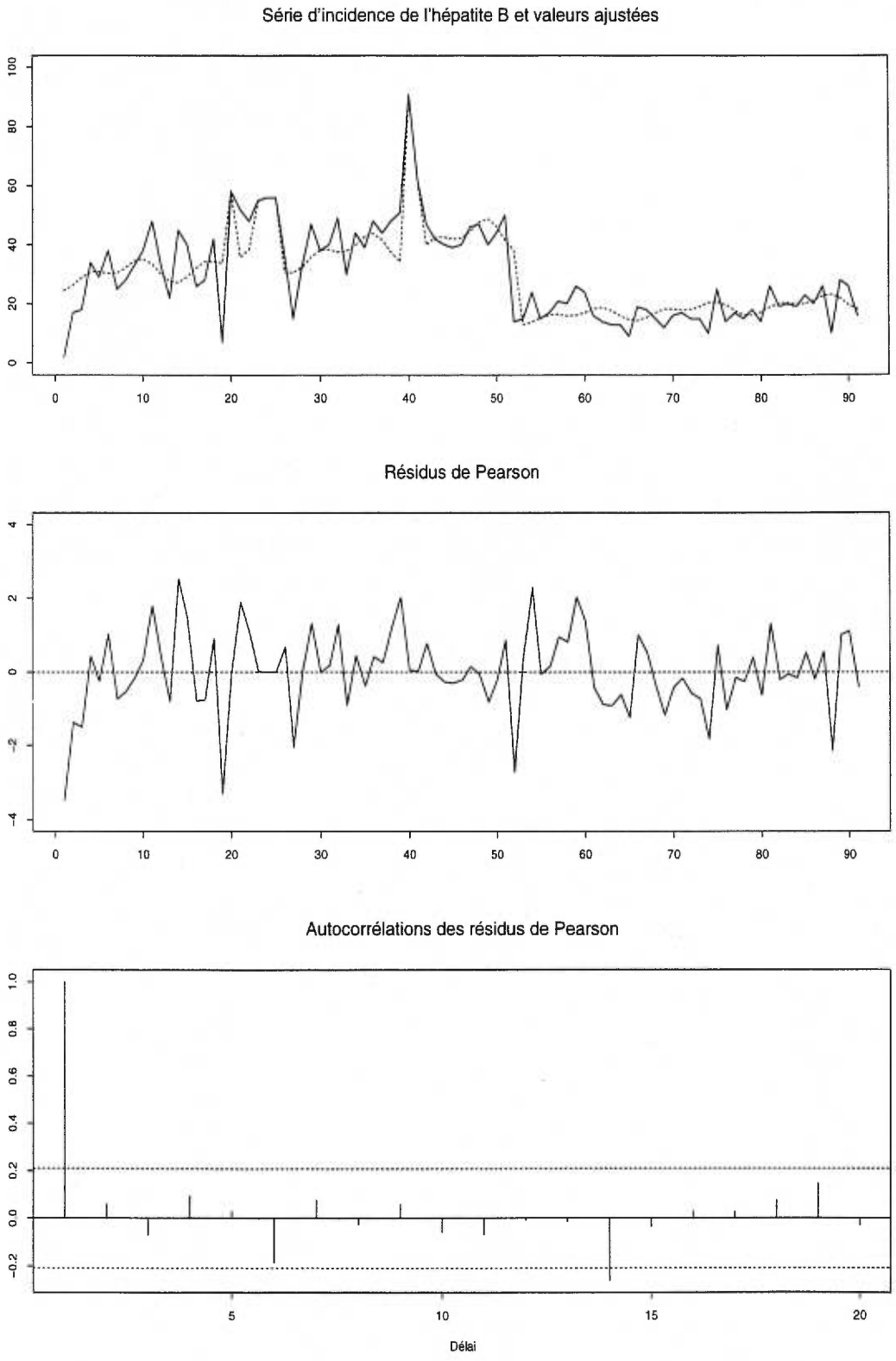

FIGURE 5.2.1. *Graphique de la série d'incidence de l'hépatite B et des valeurs ajustées, des résidus de Pearson et des autocorrélations des résidus de Pearson.*

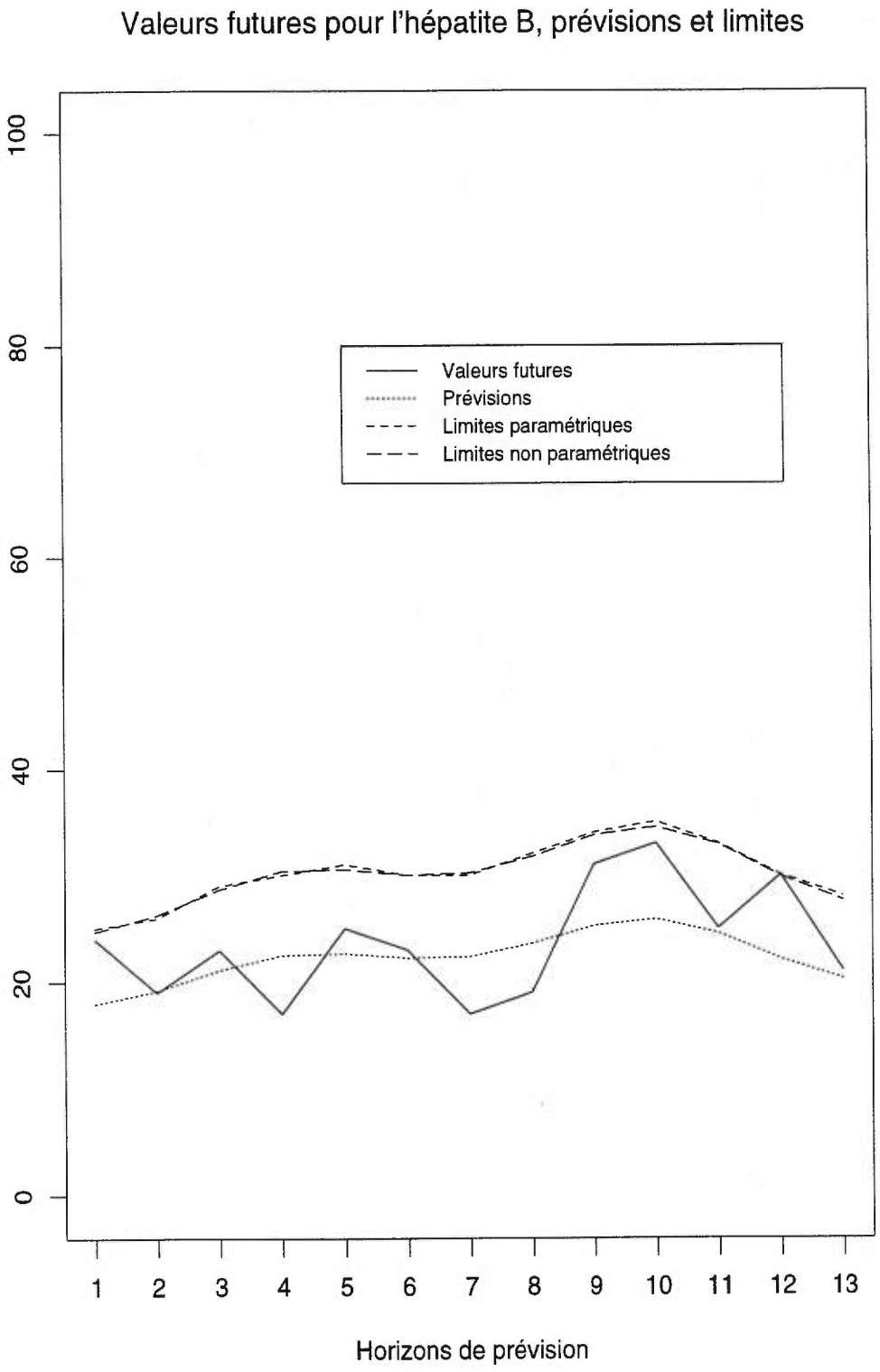

FIGURE 5.2.2. *Graphique des valeurs futures pour la série d'incidence de l'hépatite B, des prévisions et des limites des intervalles de prévision.*

5.3. Série d'incidence de la coqueluche

La série d'incidence de la coqueluche est illustrée à la page 102. Tout comme dans le cas de l'hépatite B, on y constate la présence de deux périodes possiblement épidémiques. La première a eu lieu à la fin de l'année 1990 ($t = 62, 63, 64, 65$), alors que la deuxième s'est produite à la fin de l'année 1992 ($t = 88, 89, 90$). Encore une fois, nous avons choisi d'inclure dans notre modèle un terme constant, une tendance linéaire et des composantes annuelles et semi-annuelles. Nous avons également choisi d'inclure 8 variables indicatrices. La première, définie par

$$I_t^* = \begin{cases} 1 & , \text{ si } t \geq 53, \\ 0 & , \text{ sinon,} \end{cases}$$

est la même que celle incluse pour l'hépatite B et remplit la même fonction. Les sept autres indicatrices ont la forme

$$I_{t_n} = \begin{cases} 1 & , \text{ si } t = n, \\ 0 & , \text{ sinon.} \end{cases}$$

Elles ont pour but de prendre en considération les valeurs extrêmes observées aux temps possiblement épidémiques ($t = 62, 63, 64, 65, 88, 89, 90$) et d'éviter que ces données ne faussent le modèle. Elles ont été préférées à deux indicatrices de

période épidémique, étant donné la grande variation des valeurs à l'intérieur d'une même période épidémique.

Le modèle suggéré est donc le suivant:

$$\log(\mu_t) = \beta_0 + \beta_1 \frac{t}{91} + \beta_2 cos(\frac{2\pi t}{13}) + \beta_3 sin(\frac{2\pi t}{13}) + \beta_4 cos(\frac{4\pi t}{13}) + \beta_5 sin(\frac{4\pi t}{13}) + \beta_6 I_t^* + \beta_7 I_{t_{62}} + \beta_8 I_{t_{63}} + \beta_9 I_{t_{64}} + \beta_{10} I_{t_{65}} + \beta_{11} I_{t_{88}} + \beta_{12} I_{t_{89}} + \beta_{13} I_{t_{90}} . \tag{5.3.1}$$

Les estimations des paramètres de régression et de nuisance sont présentées dans le tableau qui suit. Rappelons, encore une fois, que les estimations sont celles obtenues à partir des 91 premières observations, les 13 dernières étant réservées pour l'appréciation des deux méthodes de construction d'intervalles de prévision dans un contexte réel.

TABLEAU 5.3.1. *Paramètres de régression et de nuisance pour le modèle (5.3.1) ajusté à la série d'incidence de la coqueluche.*

β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
1,767	1,423	0,134	-0,520	-0,074	-0,012	-0,240	1,270
(0,250)	(0,526)	(0,060)	(0,063)	(0,051)	(0,052)	(0,262)	(0,205)
β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	σ^2	$\rho_\epsilon(1)$
1,467	1,304	1,085	0,744	0,341	0,681	0,075	0,920
(0,218)	(0,231)	(0,241)	(0,201)	(0,234)	(0,227)		

Le graphique de la série et des valeurs ajustées est présenté à la page suivante, en compagnie des graphiques des résidus de Pearson du modèle et des autocorrélations de ces mêmes résidus.

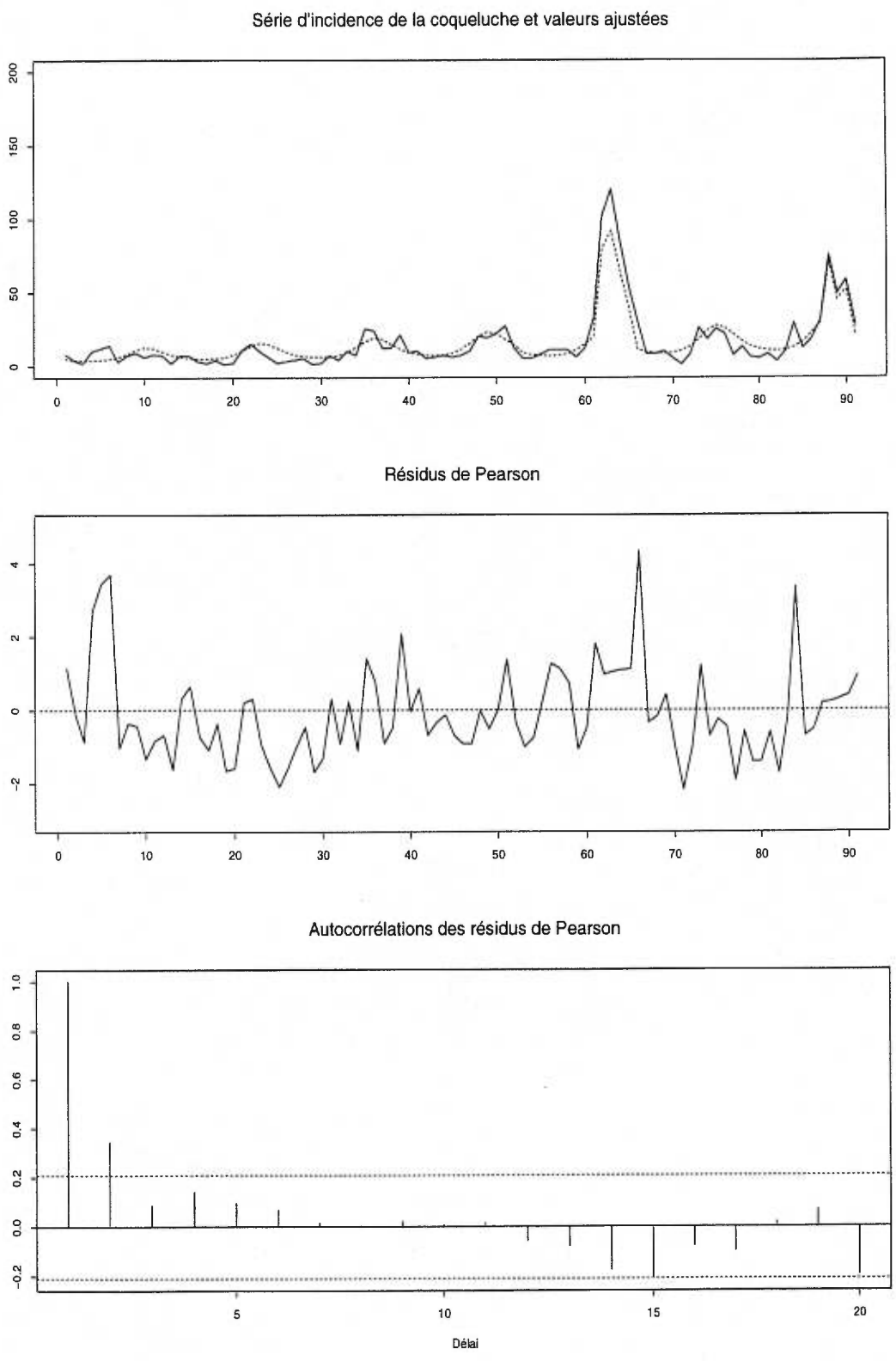

FIGURE 5.3.1. *Graphique de la série d'incidence de la coqueluche et des valeurs ajustées, des résidus de Pearson et des autocorrélations des résidus de Pearson.*

Bien que les résidus varient entre -2 et 4 , on constate que la majorité d'entre eux sont négatifs. De plus, l'alternance entre les résidus négatifs et les résidus positifs est rare. Cette situation est due au fait que la corrélation de délai 1 est relativement grande. En fait, elle est estimée à 0,344. Une bonne façon de régler le problème serait de proposer un processus latent AR(2).

Pour ce faire, on doit modifier quelque peu les estimateurs. À partir de la relation

$$\mathrm{Cov}(Y_t, Y_{t+\tau}) = \mu_t \mu_{t+\tau} \sigma^2 \rho_\epsilon(\tau), \quad \tau > 0,$$

on peut proposer, à l'aide de la méthode des moments, les estimateurs suivants pour $\rho_\epsilon(1)$ et $\rho_\epsilon(2)$:

$$\hat{\rho}_\epsilon(1) = \frac{\sum_{t=1}^{n-1} (y_t - \hat{\mu}_t)(y_{t+1} - \hat{\mu}_{t+1})}{\hat{\sigma^2} \sum_{t=1}^{n-1} \hat{\mu}_t \hat{\mu}_{t+1}},$$

$$\hat{\rho}_\epsilon(2) = \frac{\sum_{t=1}^{n-2} (y_t - \hat{\mu}_t)(y_{t+2} - \hat{\mu}_{t+2})}{\hat{\sigma^2} \sum_{t=1}^{n-2} \hat{\mu}_t \hat{\mu}_{t+2}}.$$

À partir des relations bien connues unissant ρ_1 et ρ_2 à ϕ_1 et ϕ_2 dans le cadre d'un processus stationnaire AR(2), on propose naturellement les estimateurs suivants, aussi appelés estimateurs de Yule-Walker, pour ϕ_1 et ϕ_2 :

$$\hat{\phi}_1 = \frac{\hat{\rho}_\epsilon(1)(1-\hat{\rho}_\epsilon(2))}{1-\hat{\rho}_\epsilon(1)^2}\,,$$
$$\hat{\phi}_2 = \hat{\rho}_\epsilon(2) - \hat{\phi}_1\hat{\rho}_\epsilon(1)\,.$$

Aussi, étant donné que les autocorrélations d'un processus AR(2) sont caractérisées par l'équation

$$\rho_\epsilon(j) = \phi_1\rho_\epsilon(j-1) + \phi_2\rho_\epsilon(j-2)\,, \qquad j \geq 3\,,$$

nous proposons d'estimer $\rho_\epsilon(j)$ par

$$\hat{\rho}_\epsilon(j) = \hat{\phi}_1\hat{\rho}_\epsilon(j-1) + \hat{\phi}_2\hat{\rho}_\epsilon(j-2)\,, \qquad j \geq 3\,.$$

Quant à la matrice de covariance $\mathbf{V}_n$, elle peut être estimée par

$$\hat{\mathbf{V}} = (\hat{v}_{st})_{n\times n}\,,$$

où

$$\hat{v}_{st} = \begin{cases} \hat{\mu}_t + \hat{\sigma^2}(\hat{\mu}_t)^2 & , \quad s = t\,, \\ \hat{\mu}_s\hat{\mu}_t\hat{\sigma^2}\hat{\rho}_\epsilon(|s-t|) & , \quad s \neq t\,. \end{cases}$$

Malheureusement, l'implantation de ces estimateurs dans l'algorithme utilisé résulte en un problème d'inversion impossible dû à une matrice apparemment singulière, du moins dans le cas présent. Nous conserverons donc le modèle (5.3.1) et les résultats présentés dans le tableau 5.3.1, bien qu'apparemment il puisse être amélioré.

Le graphique des valeurs futures $t = 92, \ldots, 104$, associées aux horizons de prévision $l = 1, \ldots, 13$, des prévisions et des limites des intervalles de prévision paramétriques et non paramétriques est présenté à la page suivante. On peut y constater qu'une période épidémique débute au temps $t = 100$ ($l = 9$) et se poursuit jusqu'au temps $t = 104$ ($l = 13$). Pendant toute cette période, les limites des intervalles de prévision sont inférieures aux valeurs de la série. On aurait donc conclu correctement à la présence d'une épidémie dès le temps $t = 100$. On remarque également qu'en deux autres occasions, soient aux temps $t = 94$ ($l = 3$) et $t = 96$ ($l = 5$), les valeurs futures sont supérieurs à la limite des intervalles de prévision, alors qu'il n'y a pas épidémie. Notons toutefois que, dans ces deux cas, les limites fournies sont relativement près des valeurs de la série.

A la lumière de ces résultats, il semble que les deux techniques développées constituent un outil valable pour la détection de périodes épidémiques. Il est évident qu'il s'agit d'une modélisation mathématique d'un phénomène médical et que cet outil ne doit pas constituer l'unique base de décision. Il peut toutefois s'avérer un excellent point de référence.

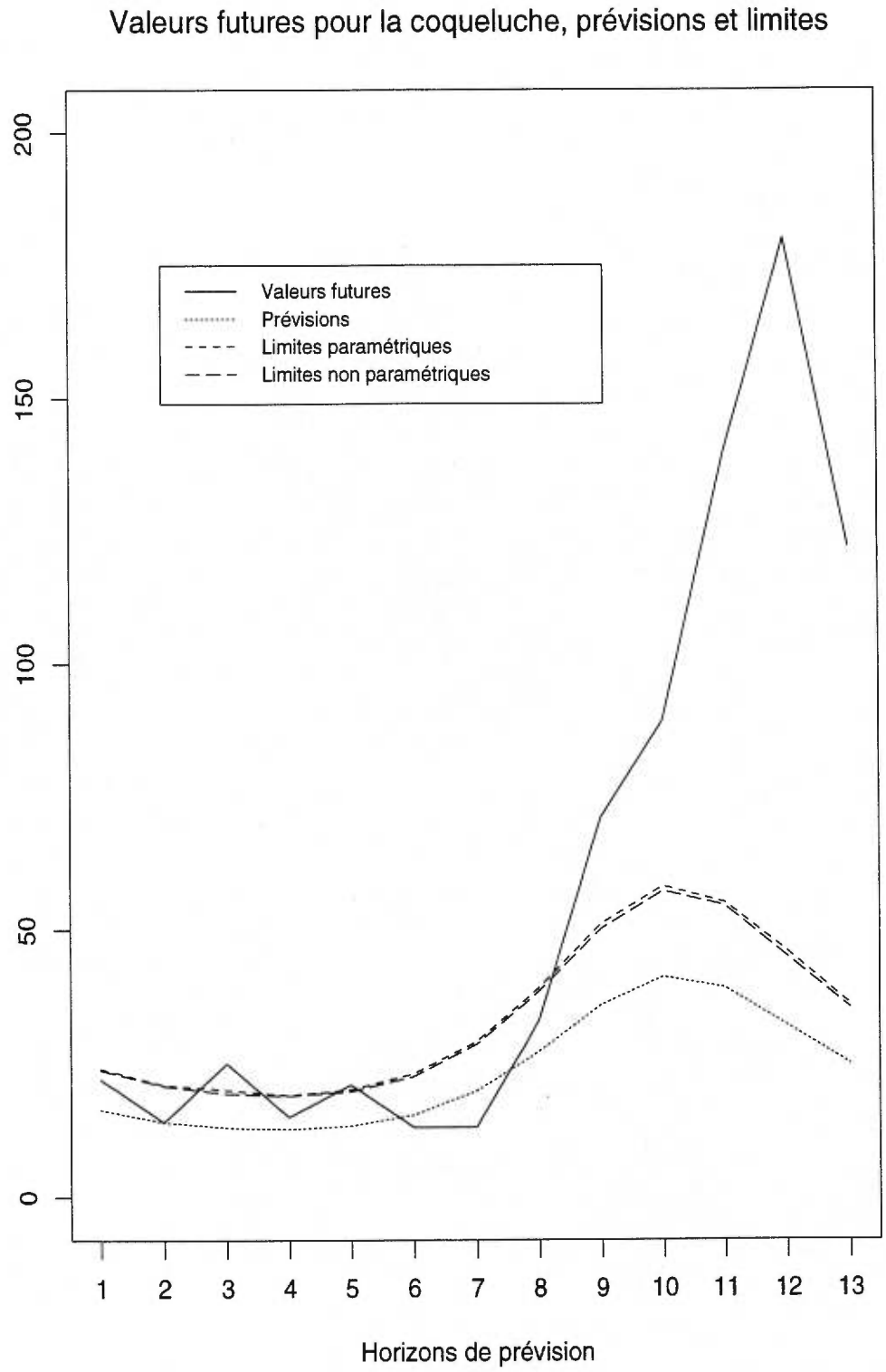

FIGURE 5.3.2. *Graphique de la série d'incidence de la coqueluche, des prévisions et des limites des intervalles de prévision.*

CONCLUSION

Ce mémoire avait pour but de proposer des solutions permettant l'obtention d'intervalles de prévision pour les valeurs futures d'une série chronologique à valeurs entières modélisée à l'aide de la classe de modèles dictés par les états développés par Zeger (1988) et par Blais, MacGibbon et Roy (1997). Deux méthodes originales ont été proposées. La première, paramétrique, repose sur la spécification de la distribution marginale du processus latent $\{\epsilon_t\}$. La deuxième méthode, non paramétrique, est basée sur l'utilisation d'une transformation T qui stabilise la variance de la série et sur l'utilisation des différences $\delta_t = T(y_t) - T(\hat{\mu}_t)$.

Des simulations ont permis d'évaluer les performances de ces deux méthodes. Dans le cas de la méthode paramétrique, on a pu constater que les taux de couverture observés, bien que généralement supérieurs à l'objectif visé, étant donné la distribution discrète de Y_t, sont près de l'objectif, et ce, aussi bien dans le cas unilatéral que dans le cas bilatéral. Des taux similaires ont également été observés, peu importe la vraie distribution marginale du processus latent $\{\epsilon_t\}$. Dans le cas de la méthode non paramétrique, nous avons pu constater que les taux de couverture observés ont un comportement différent selon qu'il s'agit d'intervalles unilatéraux ou bilatéraux. Dans le cadre unilatéral, les taux observés sont très près de l'objectif visé, bien que généralement légèrement inférieurs. Dans le cadre bilatéral, les taux observés sont beaucoup plus variables. D'après nos analyses, cette variabilité serait due à une instabilité au niveau des petites valeurs de $\delta_t = T(y_t) - T(\hat{\mu}_t)$.

Nous avons finalement appliqué nos méthodes à des séries réelles du domaine de l'épidémiologie. Les résultats positifs permettent de penser que les deux techniques développées constituent des outils valables pour la détection de périodes d'épidémie.

Les deux méthodes proposées ouvrent la voie à plusieurs sujets de recherche. La méthode non paramétrique semble prometteuse mais il y a place à amélioration dans le cas bilatéral. Aussi, la validité, d'un point de vue théorique, mérite d'être étudiée plus en détail. Il serait également intéressant d'effectuer des recherches sur l'estimation des paramètres de nuisance. Une méthode plus efficace que la méthode des moments pour estimer ces derniers devrait améliorer la modélisation et avoir un impact sur les intervalles de prévision.

Annexe A

ANNEXE A

A.1. Tableaux des résultats des simulations

Dans cette section, nous présentons sous la forme de tableaux les résultats obtenus lors des simulations. Les résultats sont présentés dans l'ordre selon lequel ils ont été analysés. Les résultats concernant la méthode paramétrique précèdent donc les résultats non paramétriques, alors qu'à l'intérieur de chaque méthode, les taux de couverture unilatéraux précèdent leurs homologues bilatéraux.

TABLEAU A.1.1. *Taux de couverture de l'intervalle unilatéral, en pourcentage, selon la méthode paramétrique, pour les temps 101 à 112.*

Mét.	par.	(σ^2, ϕ)	Dist.	101	102	103	104	105	106	107	108	109	110	111	112
		$(1/2, 1/2)$	Taux	92.00	93.05	90.65	91.95	94.65	92.45	92.35	91.90	90.70	91.15	89.95	91.00
		$(1/4, 3/4)$	de	95.05	95.70	93.00	92.30	96.05	89.35	92.50	92.75	91.45	91.45	92.85	93.25
		$(3/4, 1/4)$	référence	89.80	92.00	90.30	91.35	93.20	91.05	90.15	89.95	90.30	92.20	92.45	93.05
PARAMÉTRIQUE	CONNUS	$(1/2, 1/2)$	Gamma	91.45	91.00	91.80	91.35	91.85	90.25	89.55	88.95	87.80	89.70	88.75	89.40
			LogN.	92.50	92.10	92.80	92.70	92.95	91.50	89.80	90.85	89.65	90.95	89.85	89.70
			Beta	90.65	91.05	90.55	92.05	91.30	88.75	90.20	88.35	88.90	87.65	87.95	89.60
		$(1/4, 3/4)$	Gamma	93.60	92.40	92.30	91.75	92.55	91.10	89.90	89.10	89.50	88.85	89.65	89.25
			LogN.	93.55	93.40	93.75	92.90	93.05	92.80	91.35	90.55	89.00	89.80	89.45	89.85
			Beta	91.30	92.40	92.15	92.80	92.50	91.30	90.60	89.95	89.00	89.35	89.45	91.00
		$(3/4, 1/4)$	Gamma	90.35	91.40	91.25	92.20	92.10	90.35	90.20	88.80	89.00	89.70	92.20	91.05
			LogN.	91.10	91.50	92.70	91.80	93.15	91.00	90.85	91.10	90.25	90.90	91.50	91.20
			Beta	90.15	91.45	91.25	92.30	91.55	89.75	89.00	89.20	87.95	88.70	89.30	90.40
	ESTIMES	$(1/2, 1/2)$	Gamma	91.10	90.40	91.25	90.95	91.60	89.70	88.90	88.30	86.75	88.80	88.15	88.85
			LogN.	91.70	91.85	92.30	92.60	92.80	90.80	89.15	89.95	88.95	89.50	88.90	89.00
			Beta	90.55	91.00	90.15	91.65	91.35	88.30	89.60	87.65	88.25	87.30	87.50	88.90
		$(1/4, 3/4)$	Gamma	93.30	92.15	91.90	91.25	92.35	90.45	89.40	88.35	88.45	87.80	88.25	89.00
			LogN.	92.80	93.10	93.40	92.60	92.85	92.95	90.40	89.45	87.70	88.35	88.65	89.35
			Beta	91.10	91.95	92.00	92.50	92.25	90.75	89.70	89.15	87.90	88.70	87.95	90.15
		$(3/4, 1/4)$	Gamma	90.15	91.25	91.20	92.05	91.85	89.70	89.50	87.95	88.25	89.15	91.30	90.40
			LogN.	90.25	91.05	92.30	91.55	92.60	90.40	90.00	89.60	89.20	89.95	90.40	89.80
			Beta	89.70	91.00	90.90	92.25	91.25	89.55	88.70	88.30	87.35	88.10	88.65	89.80

TABLEAU A.1.2. *Taux de couverture de l'intervalle bilatéral, en pourcentage, selon la méthode paramétrique, pour les temps 101 à 112.*

Mét.	par.	(σ^2, ϕ)	Dist.	101	102	103	104	105	106	107	108	109	110	111	112
		$(1/2, 1/2)$	Taux	95.35	96.45	95.85	95.40	97.40	96.10	95.25	95.70	96.25	96.35	96.65	96.35
		$(1/4, 3/4)$	de	97.90	95.70	97.25	97.05	96.05	95.10	95.50	95.45	91.45	91.20	95.40	95.90
		$(3/4, 1/4)$	référence	95.70	95.35	96.80	95.20	95.80	96.65	95.55	95.40	95.45	96.10	96.75	96.70
PARAMÉTRIQUE	CONNUS	$(1/2, 1/2)$	Gamma	95.55	95.10	95.75	95.30	94.95	94.00	94.15	93.00	91.90	92.85	93.10	94.05
			LogN.	95.75	96.00	96.00	95.50	95.60	95.05	93.60	94.10	93.85	93.85	93.60	94.20
			Beta	95.70	96.45	95.10	95.25	95.25	94.00	94.65	92.95	92.30	91.80	93.15	93.90
		$(1/4, 3/4)$	Gamma	96.45	96.35	95.80	95.75	96.05	95.25	93.60	91.70	91.15	91.10	91.00	93.45
			LogN.	96.90	96.25	96.65	96.40	95.70	96.10	94.10	93.00	91.50	91.00	90.85	93.10
			Beta	95.95	95.85	96.60	96.40	95.90	95.30	93.95	91.25	88.95	89.35	90.95	94.15
		$(3/4, 1/4)$	Gamma	95.05	95.25	94.80	96.20	95.65	94.75	94.50	93.85	93.85	94.30	95.55	95.20
			LogN.	94.85	95.35	95.70	95.50	96.40	95.15	94.20	94.80	94.55	95.25	95.50	94.85
			Beta	95.45	95.60	95.70	96.45	95.30	94.55	93.25	94.45	92.55	93.80	93.80	94.95
	ESTIMÉS	$(1/2, 1/2)$	Gamma	95.00	94.45	95.45	95.10	94.65	93.60	92.90	90.50	89.15	90.55	90.65	93.05
			LogN.	95.15	95.45	95.75	95.30	95.10	94.80	92.50	91.70	90.80	91.55	91.25	93.15
			Beta	95.05	95.80	94.60	94.95	95.00	93.60	93.80	89.85	89.75	88.25	90.35	93.05
		$(1/4, 3/4)$	Gamma	95.90	95.85	95.80	95.25	95.85	94.60	91.85	89.65	88.50	87.90	89.00	92.05
			LogN.	96.25	95.85	96.30	95.95	95.55	95.50	92.80	90.50	87.60	88.60	88.90	91.40
			Beta	95.25	95.20	96.25	96.25	95.70	94.20	92.10	89.10	87.00	87.65	88.70	92.35
		$(3/4, 1/4)$	Gamma	94.45	94.95	94.55	95.85	95.40	93.85	93.45	92.65	92.30	92.40	94.85	94.40
			LogN.	94.10	94.70	95.40	95.05	95.70	94.50	93.50	93.35	92.00	93.35	94.20	94.30
			Beta	94.90	95.30	95.45	96.20	94.85	93.95	92.30	93.50	91.70	92.20	92.95	94.15

TABLEAU A.1.3. *Taux de couverture de l'intervalle unilatéral, en pourcentage, selon la méthode non paramétrique pour les temps 101 à 112.*

Mét.	par.	(σ^2, ϕ)	Dist.	101	102	103	104	105	106	107	108	109	110	111	112
			Gamma	87.85	87.50	87.95	86.80	88.15	86.55	87.20	86.65	84.65	87.40	85.95	86.55
N	C	(1/2, 1/2)	LogN.	88.20	87.40	87.40	88.10	88.70	86.45	86.40	87.60	86.55	86.60	85.20	85.90
O	O		Beta	87.50	87.35	87.30	88.95	88.45	86.10	87.80	86.00	87.20	86.15	86.25	86.40
N	N		Gamma	89.05	88.10	87.55	87.30	87.80	86.60	85.90	85.50	86.00	85.65	86.00	85.55
	N	(1/4, 3/4)	LogN.	89.25	89.00	89.15	87.35	88.10	88.75	87.55	86.05	85.20	86.30	84.65	86.40
P	U		Beta	87.75	88.30	87.65	87.45	88.20	87.50	87.25	87.05	85.55	86.40	85.05	87.25
A	S		Gamma	87.90	88.05	88.20	88.60	88.85	86.85	87.45	87.00	86.85	87.80	89.10	88.55
R		(3/4, 1/4)	LogN.	86.85	87.70	87.25	86.80	88.50	87.40	86.70	87.15	86.95	87.70	87.90	86.90
A			Beta	87.85	88.95	87.90	89.10	88.50	87.30	87.65	87.45	86.75	87.75	87.80	87.95
M			Gamma	88.45	88.20	88.75	87.60	89.00	87.40	87.40	86.75	84.90	87.65	86.50	86.95
É	E	(1/2, 1/2)	LogN.	88.90	88.25	88.80	89.15	89.30	87.10	86.55	87.60	86.65	86.90	85.60	86.40
T	S		Beta	88.00	88.20	88.15	89.55	89.30	86.65	88.15	86.40	87.20	86.45	86.25	86.65
R	T		Gamma	90.35	89.25	88.55	87.90	88.30	87.50	86.70	86.05	86.50	86.20	86.10	86.35
I	I	(1/4, 3/4)	LogN.	90.30	90.30	90.25	88.65	89.35	89.75	88.10	86.60	84.95	86.40	85.45	87.40
Q	M		Beta	88.30	89.35	88.45	89.00	89.10	88.00	87.35	87.10	85.90	86.50	85.75	87.75
U	E		Gamma	88.55	88.50	88.70	89.45	89.40	87.25	87.40	87.10	86.95	87.80	89.45	89.15
E	S	(3/4, 1/4)	LogN.	87.65	88.15	88.15	87.55	89.15	88.05	87.45	87.35	87.05	87.55	87.95	87.35
			Beta	88.35	89.50	88.45	89.70	88.85	87.95	87.75	87.55	86.75	88.00	88.10	88.30

TABLEAU A.1.4. *Taux de couverture de l'intervalle bilatéral, en pourcentage, selon la méthode non paramétrique pour les temps 101 à 112.*

Mét.	par.	(σ^2, ϕ)	Dist.	101	102	103	104	105	106	107	108	109	110	111	112
			Gamma	86.30	88.80	91.20	90.95	90.35	86.20	83. 00	81.50	82.45	83.25	81.85	81.70
N	C	(1/2, 1/2)	LogN.	81.95	86.40	87.00	87.10	86.30	80.25	80.25	84.75	84.35	85.45	83.85	79.05
O	O		Beta	91.50	92.15	91.40	92.55	92.25	88.25	87.50	82.35	82.10	81.30	83.20	86.55
N	N		Gamma	82.15	83.25	85.20	83.55	83.55	79.65	81.25	82.50	84.55	83.60	83.45	79.75
	N	(1/4, 3/4)	LogN.	80.85	80.95	81.70	82.85	80.05	79.50	82.55	84.20	83.75	84.70	82.50	80.45
P	U		Beta	84.10	86.15	87.40	87.45	85.50	82.55	82.00	83.60	83.80	82.70	81.60	82.25
A	S		Gamma	90.10	92.60	92.35	93.15	93.80	90.55	86.60	82.75	84.15	82.35	86.20	88.30
R		(3/4, 1/4)	LogN.	85.65	89.45	91.35	90.40	90.95	85.35	80.70	83.30	84.25	84.40	84.25	80.75
A			Beta	92.10	93.55	93.05	93.95	92.80	91.10	86.75	85.70	84.45	82.65	86.15	89.40
M			Gamma	87.00	89.85	92.25	91.80	91.35	87.05	83.50	81.75	82.65	83.60	81.85	82.20
É	E	(1/2, 1/2)	LogN.	83.00	87.50	89.15	89.45	87.50	81.30	80.70	84.80	84.40	85.70	84.20	79.85
T	S		Beta	92.05	93.15	92.15	93.50	93.00	89.45	88.35	82.90	82.25	81.40	83.80	87.10
R	T		Gamma	83.15	85.55	87.80	86.10	85.40	81.00	82.30	83.15	84.80	83.90	84.05	81.00
I	I	(1/4, 3/4)	LogN.	82.10	83.15	84.65	84.95	82.10	81.05	83.30	84.55	84.40	84.80	82.80	81.60
Q	M		Beta	85.80	88.20	89.25	89.65	86.90	83.60	82.60	83.85	83.90	82.75	82.05	82.75
U	E		Gamma	90.95	93.00	92.70	93.95	94.20	91.25	86.85	83.10	84.30	82.40	86.25	88.40
E	S	(3/4, 1/4)	LogN.	86.30	90.00	92.15	91.30	91.65	86.25	80.90	83.40	84.20	84.50	84.65	81.25
			Beta	92.45	93.75	93.00	94.35	93.35	91.35	86.95	85.70	84.55	82.75	86.00	89.60

A.2. Section informatique

Dans cette section, nous présentons diverses fonctions et commandes informatiques utilisées dans le cadre de la réalisation de ce mémoire.

Commandes utilisées pour générer le processus latent et les observations. Étant donné la similitude des commandes, seul les commandes du duo (3/4,1/4) sont reproduites ci-bas.

```
#INITIALISATIONS#
vuni31_NULL
vvn31_NULL
vexp31_NULL
vserie31_NULL
seriesEn31_matrix(0,nrow=2000,ncol=112)
seriesYt31_matrix(0,nrow=2000,ncol=112)

#Génération de processus latents avec distribution marginale Gamma#
#selon la méthode présentée au chapitre 3#
for (i in 1:2000)
{  vuni31_runif(212,min=0,max=1)
   vvn31_vuni31^3
   vexp31_rexp(212,rate=4/3)
   vserie31[1]_1
   for(j in 2:212)
   {  vserie31[j]_vvn31[j]*vserie31[j-1]+vexp31[j]  }
   seriesEn31[i,]_vserie31[-c(1:100)]
```

```
   if (i%%100==0) {cat(i,")",date(),"\n")}     }

#Génération des observations à partir des processus latents#
for(i in 1:2000)
{  seriesYt31[i,]_rpois(112,VraiMUt*seriesEn31[i,])  }

#INITIALISATIONS#
vnor32_NULL
vserie32_NULL
seriesEn32_matrix(0,nrow=2000,ncol=112)
rhodelta3_log(19/16)/log(7/4)
seriesYt32_matrix(0,nrow=2000,ncol=112)

#Génération de processus latents avec distribution marginale#
#lognormale selon la méthode présentée à la section 4.1#
for (i in 1:2000)
{  vnor32_rnorm(212,mean=0,sd=sqrt((1-rhodelta3^2)*log(7/4)))
   vserie32[1]_-log(7/4)/2
   for(j in 2:212)
   {  vserie32[j]_(1-rhodelta3)*(-log(7/4)/2)
                 +rhodelta3*vserie32[j-1]+vnor32[j]  }
   seriesEn32[i,]_exp(vserie32[-c(1:100)])
   if (i%%100==0) {cat(i,")",date(),"\n")}   }

#Génération des observations à partir des processus latents#
for(i in 1:2000)
```

```
{  seriesYt32[i,]_rpois(112,VraiMUt*seriesEn32[i,])  }

#INITIALISATIONS#
vbeta33_NULL
vserie33_NULL
serieUn33_NULL
serieVn33_NULL
seriesEn33_matrix(0,nrow=2000,ncol=112)
c33_7
seriesYt33_matrix(0,nrow=2000,ncol=112)

#Génération de processus latents avec distribution marginale#
#bêta modifiée selon la méthode présentée à la section 4.1#
for (i in 1:2000)
{  vbeta33_rbeta(212,1,6)
   vserie33[1]_1/7
   serieUn33_rbeta(212,6,18/25)
   serieVn33_rbeta(212,7/25,18/25)
   for(j in 2:212)
   {  vserie33[j]_1-serieUn33[j]*(1-serieVn33[j]*vserie33[j-1])  }
   seriesEn33[i,]_c33*vserie33[-c(1:100)]
   if (i%%100==0) {cat(i,")",date(),"\n")}  }

#Generation des observations à partir des processus latents#
for(i in 1:2000)
{  seriesYt33[i,]_rpois(112,VraiMUt*seriesEn33[i,])  }
```

La prochaine fonction est utilisée pour l'estimation des paramètres de régression et de nuisance lorsque les paramètres de nuisance sont inconnus. Elle a été écrite par Blais (1996). Elle est reprise ici, ayant été légèrement transformée. La fonction utilisée lorsque les paramètres de nuisance sont connus étant très similaire, elle ne sera pas reproduite ci-bas.

```
quasi.nuis.estimes_function(obs, X, beta0, FON=exp, DER=exp, prec=0.01)
{

# Initialisation #

  X <- as.matrix(X)
  obs <- as.vector(obs)
  beta <- as.vector(beta0)
  if(is.character(FON))
     FON <- get(FON, mode = "function")
  else if(mode(FON) != "function")
  {
     farg <- substitute(FON)
     if(mode(farg) == "name")
        FON <- get(farg, mode = "function")
     else stop(paste("\"", farg, "\" is not a function", sep = ""))
  }
  if(is.character(DER))
```

```
   DER <- get(DER, mode = "function")
else if(mode(DER) != "function")
{
   farg <- substitute(DER)
   if(mode(farg) == "name")
      DER <- get(farg, mode = "function")
   else stop(paste("\"", farg, "\" is not a function", sep = ""))
}
mu <- FON(as.vector(X %*% beta))
n <- length(obs)
err <- mean((obs - mu)^2)
sigma <- sum((obs - mu)^2 - mu)/sum(mu^2)
if(is.na(sigma) | sigma < 0.01)
   sigma <- 0.01
rho <- sum((obs - mu)[2:n] * (obs - mu)[1:(n - 1)])/
       sum(mu[2:n] * mu[1:(n - 1)])/sigma
if(is.na(rho))
   rho <- 0
if(abs(rho) > 0.99)
   rho <- 0.99 * sign(rho)
D <- X * DER(as.vector(X %*% beta))
rha <- rho^(1:(n - 1))
rha <- c(rep(c(1, rha, 0), n - 1), 1)
R <- matrix(rha, n, n)
R[col(R) > row(R)] <- 0
R <- R + t(R) - diag(n)
```

```
  V <- mu * t(mu * R) * sigma + diag(mu)
  IO <- solve(V)
  if(is.na(sum(IO)))
     break
  VR <- solve(t(D) %*% IO %*% D)
  if(is.na(sum(VR)))
     break

#Fin de l'initialisation#

#Itérations pour estimer les paramètres de régression et les  #
#paramètres de nuisance.  La méthode d'équations d'estimation #
#est utilisée. La convergence est atteinte lorsque la somme en#
#valeurs absolues des différences entre les composantes des   #
#paramètres de régression de l'itération actuelle et de       #
#l'itération précédente est inférieure à 0,01.                #

  j <- 0
  compteur <- 0
  bb <- rep(0, length(beta))
  while(sum(abs(bb - beta)) > prec & j <= 25)
  {
      j <- j + 1
      bb <- beta
      beta <- beta + as.vector(VR %*% t(D) %*% IO %*% (obs - mu))
      mu <- FON(as.vector(X %*% beta))
```

```
sigma <- sum((obs - mu)^2 - mu)/sum(mu^2)
if(is.na(sigma) | sigma < 0.01)
   sigma <- 0.01
rho <- sum((obs - mu)[2:n] * (obs - mu)[1:(n - 1)])/
       sum(mu[2:n] * mu[1:(n - 1)])/sigma
if(is.na(rho))
   rho <- 0
if(abs(rho) > 0.99)
   rho <- 0.99 * sign(rho)
D <- X * DER(as.vector(X %*% beta))
rha <- rho^(1:(n - 1))
rha <- c(rep(c(1, rha, 0), n - 1), 1)
R <- matrix(rha, n, n)
R[col(R) > row(R)] <- 0
R <- R + t(R) - diag(n)
V <- mu * t(mu * R) * sigma + diag(mu)
if(is.na(sum(V)))
   break
I0 <- solve(V)
if(is.na(sum(I0)))
   break
VR <- solve(t(D) %*% I0 %*% D)
if(is.na(sum(VR)))
   break
if(j == 26)
   {compteur <- compteur + 1}
```

```
  }

#Fin des iterations#

#Calculs finaux#

  erreur <- mean((obs - mu)^2)
  A <- diag((mu + sigma * mu^2)^(-0.5))
  rec <- obs - mu
  res <- as.vector(A %*% rec)
  sss <- sqrt(diag(VR))
  corr <- diag(1/sss) %*% VR %*% diag(1/sss)
  reponse <- list(coefficients = beta, res.pearson = res,
             res.classique = rec, erreur = erreur,
             resul = list(estimation = matrix(c(beta, sss, beta/sss),
             ncol = 3), correlation = corr, sigma.deux = sigma,
             phi = rho, conv. = compteur))
  reponse
}
```

Les deux prochaines fonctions sont utilisées dans le cadre de la méthode paramétrique afin de déterminer les quantiles de la distribution résultante.

```
dnbinomgen_function(a, b, mut, yt)
{
 (gamma(a + yt)/(gamma(a) * gamma(yt + 1))) *
 (b/(b + mut))^a * (mut/(b + mut))^yt
}
```

```
qnbinomgen_function(a, b, c, x)
{
  j <- 0
  if(dnbinomgen(a, b, c, j) >= x)
     { j }
  else
  {  while(sum(dnbinomgen(a, b, c, 0:j)) < x)
     { j <- j + 1 }
     j
  }
}
```

Finalement, les commandes suivantes ont été utilisées pour déterminer les bornes et calculer les taux. Encore une fois, les commandes étant très similaires, nous ne présenterons qu'un seul cas (duo (3/4,1/4) estimé, $\epsilon_t \sim$ gamma).

```
#initialisation#
coefficientsbeta312_matrix(0,nrow=2000,ncol=11)
bornes3122_matrix(0,nrow=2000,ncol=36)
niveau3122bi_c(1:12)
niveau3122uni_c(1:12)
residus3124_matrix(0,nrow=2000,ncol=100)
bornes3124_matrix(0,nrow=2000,ncol=36)
niveau3124bi_c(1:12)
niveau3124uni_c(1:12)

#estimation des paramètres de nuisance et de régression#
for(i in 1:2000)
{   tempo_quasi.nuis.estimes(seriesYt31[i,1:100],Xt[1:100,],
          glm(seriesYt31[i,1:100]~Xt[1:100,2]+Xt[1:100,3]+Xt[1:100,4],
          family="poisson")$coef,exp,exp,0.01)
    coefficientsbeta312[i,]_c(tempo$coef,tempo$resul$estimation[,2],
                              tempo$resul$sigma.deux,tempo$resul$phi,
                              tempo$resul$conv)
}
```

```
#calcul des bornes et des taux pour la méthode paramétrique#

b312_(1/coefficientsbeta312[,9])
a312_b312
MUt312_exp(coefficientsbeta312[1:2000,1:4]%*%t(Xt[1:112,]))
for(i in 1:2000)
{   for(j in 1:12)
    { bornes3122[i,c(j,j+12,j+24)]_c(qnbinomgen(a312[i],b312[i],
      MUt312[i,100+j],0.05),qnbinomgen(a312[i],b312[i],MUt312[i,100+j],
      0.95),qnbinomgen(a312[i],b312[i],MUt312[i,100+j],0.90))
    }
    if (i %% 100 == 0) cat(i,")",date(),"\n")
}

for(i in 1:12)
{   niveau3122bi[i]_(2000-sum(seriesYt31[,100+i]<bornes3122[,i])
                    -sum(seriesYt31[,100+i]>bornes3122[,12+i]))/2000
}

for(i in 1:12)
{   niveau3122uni[i]_(2000-sum(seriesYt31[,100+i]>
                      bornes3122[,24+i]))/2000
}
```

```
#calcul des bornes et des taux pour la méthode non paramétrique#

var312_coefficientsbeta312[,9]
for(i in 1:2000)
{   for(j in 1:100)
    { residus3124[i,j]_acosh(2*var312[i]*seriesYt31[i,j]+1)
                        -acosh(2*var312[i]*MUt312[i,j]+1) }
    if (i %% 100 == 0) cat(i,")",date(),"\n")
}

delta3124_t(apply(residus3124,1,quantile,c(0.05,0.95,0.90)))
for(i in 1:2000)
{   for(j in 1:12)
    { bornes3124[i,j]_(cosh(delta3124[i,1]+acosh(2*var312[i]*
                        MUt312[i,100+j]+1))-1)/(2*var312[i])
      if (delta3124[i,1]+acosh(2*var312[i]*MUt312[i,100+j]+1)<0)
          {bornes3124[i,j]_0}
    }
}

for(i in 1:2000)
{   for(j in 13:24)
    { bornes3124[i,j]_(cosh(delta3124[i,2]+acosh(2*var312[i]*
                        MUt312[i,100+j-12]+1))-1)/(2*var312[i]) }
}
```

```
for(i in 1:2000)
{  for(j in 25:36)
   { bornes3124[i,j]_(cosh(delta3124[i,3]+acosh(2*var312[i]*
                        MUt312[i,100+j-24]+1))-1)/(2*var312[i]) }
}

for(i in 1:12)
{ niveau3124bina[i]_(2000-sum(seriesYt31[,100+i]<bornes3124[,i])
                     -sum(seriesYt31[,100+i]>bornes3124[,12+i]))/2000
}

for(i in 1:12)
{ niveau3124unina[i]_(2000-sum(seriesYt31[,100+i]>
                        bornes3124[,24+i]))/2000
}
```

BIBLIOGRAPHIE

Blais, M. (1996), *Modèle de régression pour des séries chronologiques à valeurs entières*, Mémoire de maîtrise, Département de mathématiques et de statistique, Université de Montréal.

Blais, M., MacGibbon, B. et Roy, R. (1997), Inference in generalized linear models for time series of counts, *Rapport technique #G-97-61*, GERAD.

Bradley, E. (1973), The equivalence of maximum likelihood and weighted least squares estimates in the exponential family, *Journal of the American Statistical Association* **68**, 199–200.

Brockwell, P. J. et Davis, R. A. (1996), *Introduction to Time Series and Forecasting*, New-York: Springer-Verlag.

Cardinal, M. (1995), *Modélisation temporelle d'incidences de maladies*, Mémoire de maîtrise, Département de médecine sociale et préventive, Université de Montréal.

Cox, D. R. (1981), Statistical analysis of time series: some recent developments, *Scandinavian Journal of Statistics* **8**, 93–115.

Davis, R. A., Dunsmuir, W. T. M. et Wang, Y. (1997), Modeling time series of count data, *Rapport technique*, Department of Statistics, Colorado State University.

Jorgensen, B., Lundbye-Christensen, S., Song, X.-K. et Sun, L. (1995), A state space model for multivariate longitudinal count data, *Rapport technique # 148*, Department of Statistics, University of British Columbia.

McCullagh, P. et Nelder, J. (1989), *Generalized Linear Models*, 2e édition, New-York: Chapman & Hall.

McKenzie, E. (1985), An autoregressive process for beta random variables, *Management Science* **31**, 988–997.

Sim, C.-H. (1986), Simulation of weibull and gamma autoregressive stationary process, *Communications in Statistics - Simulation and Computation* **15**, 1141–1146.

Spiegel, M. R. (1989), *Formules et tables de mathématiques*, Montréal: McGraw-Hill.

Wedderburn, R. (1974), Quasi-likelihood functions, generalized linear models, and the gauss newton method, *Biometrika* **61**, 439–447.

Wei, W. W. S. (1989), *Time Series Analysis: Univariate and Multivariate Methods*, New-York: Addison-Wesley.

Zeger, S. L. (1988), A regression model for time series of counts, *Biometrika* **75**, 621–629.