Numéro d'ordre : 08-2016 Année 2016

Université Claude Bernard Lyon 1

Thèse présentée pour obtenir le grade de
Docteur de l'université Claude Bernard Lyon 1

Spécialité : Statistiques

Maximum de vraisemblance empirique pour la détection de changements dans un modèle avec un nombre faible ou très grand de variables.

Présentée par
Zahraa SALLOUM

Sous la direction de
Mme. Gabriela CIUPERCA

Après avis des rapporteurs

Jean Marc Bardet	Professeur	Université Paris 1
Cristina Butucea	Professeur	Université Marne-la-Vallée
Serguei Dachian	Professeur	Université Lille 1

Soutenu le 19 Janvier 2016 devant le jury d'examen composé de

Jean Marc Bardet	Professeur	Université Paris 1
Cristina Butucea	Professeur	Université Marne-la-Vallée
Gilles Celeux	Directeur de Recherche	INRIA Futurs-Orsay
Gabriela Ciuperca	Maitre de Conférences, HDR	Université Lyon 1
Sophie Dabo-Niang	Professeur	Université Lille 3
Serguei Dachian	Professeur	Université Lille 1
Céline Vial	Maitre de Conférences, HDR	Université Lyon 1

Numéro d'ordre : 08-2016 Année 2016

Université Claude Bernard Lyon 1

Thèse présentée pour obtenir le grade de
Docteur de l'université Claude Bernard Lyon 1

Spécialité : Statistiques

Maximum de vraisemblance empirique pour la détection de changements dans un modèle avec un nombre faible ou très grand de variables.

Présentée par
Zahraa SALLOUM

Sous la direction de
Mme. Gabriela CIUPERCA

Après avis des rapporteurs

Jean Marc Bardet	Professeur	Université Paris 1
Cristina Butucea	Professeur	Université Marne-la-Vallée
Serguei Dachian	Professeur	Université Lille 1

Soutenu le 19 Janvier 2016 devant le jury d'examen composé de

Jean Marc Bardet	Professeur	Université Paris 1
Cristina Butucea	Professeur	Université Marne-la-Vallée
Gilles Celeux	Directeur de Recherche	INRIA Futurs-Orsay
Gabriela Ciuperca	Maitre de Conférences, HDR	Université Lyon 1
Sophie Dabo-Niang	Professeur	Université Lille 3
Serguei Dachian	Professeur	Université Lille 1
Céline Vial	Maitre de Conférences, HDR	Université Lyon 1

À

ma mère et mon père,

Hassan et Ali,

Zaynab, Hanan, Hiba, Mohamad Ali, Soha

et aussi à Gabriela,

modeste témoignage de mon infinie amour et tendresse...

Merci !

Je resterai reconnaissante toute ma vie à plusieurs personnes qui m'ont aidée à remettre le pied dans les moments difficiles. Je ne peux qu'être d'accord avec Aristote[1] et dire "*Il y a parmi nous des magiciens et magiciennes, mais personne ne le sait...*". À ces magiciens et magiciennes, je veux dédier mes remerciements.

Trouver les bons mots pour dire "Merci" est une chose difficile, d'autant plus dans une langue qui n'est pas sa langue maternelle. Je tiens tout d'abord à manifester ma plus profonde et sincère reconnaissance envers une personne toute spéciale pour moi ma directrice de thèse Gabriela Ciuperca qui a encadré mon travail. Gabriela a toujours fait preuve d'une totale confiance, d'un soutien et d'une gentillesse permanente à mon égard. Malgré son emploi du temps chargé, elle a toujours su trouver du temps à m'accorder aussi bien lors de l'encadrement scientifique que dans ma vie quotidienne par ses conseils et suggestions, ce qui a fait nos rapports simples et chaleureux. Pour tout cela, Merci !

Je souhaite remercier très sincèrement les rapporteurs Jean Marc Bardet, Cristina Butucea et Serguei Dachian pour avoir accepté de consacrer de leur temps à la lecture de cette thèse. Merci également à Gilles Celeux et CélineVial d'avoir acceptée de faire partie de mon jury de thèse en tant qu'examinateurs. Je suis également très reconnaissante à Sophie Dabo-Niang qui a conduit mes premiers pas en France, pour son soutient durant ces quatre dernières années, en commençant avec mon stage de Master-2, ainsi que pour sa participation au jury de cette thèse en tant qu'examinateur.

Merci aussi aux doctorant-e-s de l'institut Camille Jordan (ICJ) de l'Université Lyon 1 avec qui j'ai pu discuter des joies et des malheurs de la vie de thésard et qui m'ont accueillie. Mes remerciments vont aussi aux membres de l'ICJ de Lyon qui ont su crée un environnement à la fois stimulant et agréable, pour mener au bien cette thèse.

"*Les anges existent mais des fois ils n'ont pas d'ailes... Nous les appelons alors amis*".
Un grand merci à ma belle Meriem AHMED BACHA, Mariouma tu as été une vraie amie pour moi, tu as su être présente à chaque fois que j'en avais besoin. Tu es une personne généreuse, reste comme tu es. Encore, un grand Merci à chaque membre du groupe du libanais de Saint-Etienne, pour les bons moments familiaux, pour toute la joie, la bonne ambiance et toutes les sorties ou soirées que nous avons passé ensemble.

1. Ludovico Ariosto dit L'Arioste, poète italien, né à Reggio d'Émilie le 8 septembre 1474.

Merci à vous toutes et vous tous pour votre présence.

"*La famille est un asile et un abri, un petit nid où l'on se sent en sécurité, protégé et aimé*[2]".
Je pense aujourd'hui avec beaucoup de tendresse à mes parents Sanna et Ahmad, à la confiance qu'ils m'accordent et qui, malgré la tristesse de voir leur fille partir à l'étranger, m'ont toujours soutenue et encouragée. Je veux les remercier de tout mon coeur, ainsi qu'à mes soeurs Zaynab, Hanan, Hiba, Soha et à mon frère Mohamad Ali. Sans votre soutien et votre amour sans faille, tout ceci n'aurai jamais pu aboutir et je vous serai éternellement reconnaissante d'avoir su me donner les moyens d'arriver jusqu'ici. Merci également à mes beaux frères Wissam, Ahmad, Ali et ma belle soeur Zahraa, mes niéces Laya et Yara et mes trois neveux Hadi, Ahmad et Haydar. Aucun mot ne saurait décrire ma pensée : Je vous aime beaucoup ! !

Je n'oublie pas de remercier ma deuxième famille : ma belle-mère Fatima, mon beau-père Mohamad, mon beau-frère Jawad, sa femme Zaynab et mes deux belles-soeurs Zaynab et Hanan.

"*Et si tu n'existais pas, dis-moi pourquoi j'existerais ? pour traîner dans un monde sans toi, sans espoir et sans regrets*[3]".
C'est le moment de remercier chaleureusement ma famille, mon mari Hassan et mon bébé Ali. Vous êtes ma plus belle réussite et ma plus grande fierté et sans vous, tout cela n'aurait jamais été possible.

Hassan, mon beau prince, Merci, pour ton amour, ta patience, ton soutien et surtout pour ta présence dans ma vie. Sans toi, je n'aurais pas surmonté les longues jours d'expatriation : Tu es ma patrie, Merci bien Habibi ! !

Ali, mon trésor, j'espère qu'un jour tu seras fière de ta maman comme je suis déjà si fière de toi mon coeur !

Enfin, merci à toutes les personnes que je ne cite pas ici mais qui mériteraient d'être citées.

---

2. Maxalexis, poéte français né en Josselin (Bretagne).
3. Chanson interprétée par Joe Dassin le 20 décembre 1975, lors de la Première émission en couleur de TF1.

Soon, when all is well, you're going to look back on this period of your life and be so glad that you never gave up. [4]

4. Brittany Burgunder.

Maximum de vraisemblance empirique pour la détection de changements dans un modèle avec un nombre faible ou très grand de variables.

Résumé : Cette thèse est consacrée à tester la présence de changements dans les paramètres d'un modèle de régression non-linéaire ainsi que dans un modèle de régression linéaire en très grande dimension.

Tout d'abord, nous proposons une méthode basée sur la vraisemblance empirique pour tester la présence de changements dans les paramètres d'un modèle de régression non-linéaire. Sous l'hypothèse nulle, nous prouvons la consistance et la vitesse de convergence des estimateurs des paramètres de régression. La loi asymptotique de la statistique de test sous l'hypothèse nulle nous permet de trouver la valeur critique asymptotique. D'autre part, nous prouvons que la puissance asymptotique de la statistique de test proposée est égale à 1. Le modèle épidémique avec deux points de rupture est également étudié.

Ensuite, on s'intéresse à construire les régions de confiance asymptotiques pour la différence entre les paramètres de deux phases d'un modèle non-linéaire avec des régresseurs aléatoires en utilisant la méthode de vraisemblance empirique. On montre que le rapport de la vraisemblance empirique a une distribution asymptotique χ^2. La méthode de vraisemblance empirique est également utilisée pour construire les régions de confiance pour la différence entre les paramètres des deux phases d'un modèle non-linéaire avec des variables de réponse manquantes au hasard (Missing At Random (MAR)). Afin de construire les régions de confiance du paramètre en question, on propose trois statistiques de vraisemblance empirique : la vraisemblance empirique basée sur les données cas-complète, la vraisemblance empirique pondérée et la vraisemblance empirique par des valeurs imputées. On prouve que les trois rapports de vraisemblance empirique ont une distribution asymptotique χ^2.

Un autre but de cette thèse est de tester la présence d'un changement dans les coefficients d'un modèle linéaire en grande dimension, où le nombre des variables du modèle peut augmenter avec la taille de l'échantillon. Ce qui conduit à tester l'hypothèse nulle de non-changement contre l'hypothèse alternative d'un seul changement dans les coefficients de régression. Basée sur les comportements asymptotiques de la statistique de rapport de vraisemblance empirique, on propose une simple statistique de test qui sera utilisée facilement dans la pratique. La normalité asymptotique de la statistique de test proposée sous l'hypothèse nulle est prouvée. Sous l'hypothèse alternative, la statistique de test diverge.

L'efficacité du test de la vraisemblance empirique proposé pour chaque étude est illustrée par des simulations de Monte-Carlo en utilisant le logiciel statistique *R*.

Mots-clefs : Point de rupture. Modèle paramétrique non-linéaire. Test de la vraisemblance empirique. Intervalle de confiance. Données manquantes. Modèle linéaire en très grande dimension. Comportement asymptotique.

Maximum empirical likelihood for detecting the changes in a model with a low or very large number of variables.

Abstract : In this PHD thesis, we propose a nonparametric method based on the empirical likelihood for detecting the change in the parameters of nonlinear regression models and the change in the coefficient of linear regression models, when the number of model variables may increase as the sample size increases.

Firstly, we test the null hypothesis of no-change against the alternative of one change in the regression parameters. Under null hypothesis, the consistency and the convergence rate of the regression parameter estimators are proved. The asymptotic distribution of the test statistic under the null hypothesis is obtained, which allows to find the asymptotic critical value. On the other hand, we prove that the proposed test statistic has the asymptotic power equal to 1. The epidemic model, a particular case of model with two change-points, under the alternative hypothesis, is also studied.

Afterwards, we use the empirical likelihood method for constructing the confidence regions for the difference between the parameters of a two-phases nonlinear model with random design. We show that the empirical likelihood ratio has an asymptotic χ^2 distribution. Empirical likelihood method is also used to construct the confidence regions for the difference between the parameters of a two-phases nonlinear model with response variables missing at randoms (MAR). In order to construct the confidence regions of the parameter in question, we propose three empirical likelihood statistics : empirical likelihood based on complete-case data, weighted empirical likelihood and empirical likelihood with imputed values. We prove that all three empirical likelihood ratios have asymptotically χ^2 distributions.

An another aim for this thesis is to test the change in the coefficient of linear regression models for high-dimensional model. This amounts to testing the null hypothesis of no change against the alternative of one change in the regression coefficients. Based on the theoretical asymptotic behaviour of the empirical likelihood ratio statistic, we propose, for a deterministic design, a simpler test statistic, easier to use in practice. The asymptotic normality of the proposed test statistic under the null hypothesis is proved, a result which is different from the χ^2 law for a model with a fixed variable number. Under alternative hypothesis, the test statistic diverges.

For all studies, we show the performance of the proposed empirical likelihood tests by Monte-Carlo simulations. We use the statistical software *R*.

Key-words : Point de rupture. Nonlinear parametric model. Empirical likelihood test. Confidence region. Missing response. High-dimensional linear model. Asymptotic behaviour.

Table des matières

Liste des tableaux **21**

Notations et Abréviations **23**

1 Introduction **25**
1.1 Préliminaires 25
1.1.1 La vraisemblance empirique 28
1.1.2 Vraisemblance empirique et bootstrap 29
1.1.3 La vraisemblance empirique dans le cas de l'espérance 30
Définition 30
Théorème de convergence 32
1.2 Organisation de la thèse 32

I Test par la vraisemblance empirique pour de modèles non-linéaire avec points de rupture **35**

2 Test de la vraisemblance empirique pour un modèle non-linéaire avec points de rupture **37**
2.1 Introduction 38
2.2 Suppositions 40
2.3 Test avec un seul point de rupture 41
2.3.1 Hypothèses 41
2.3.2 Statistique de test 42
2.3.3 Comportement asymptotique de la statistique de test 47
Comportement asymptotique sous H_0 47
Comportement asymptotique de T_{nk} et $\tilde{T}_{nk}$ sous H_1 55
2.4 Test d'un modèle épidémique non-linéaire 63

2.5 Simulations . . . 65
2.5.1 Modèle avec un seul point de rupture . . . 66
2.5.2 Modèle épidémique . . . 68
2.6 Annexe . . . 69

3 Intervalle de confiance pour la différence entre les paramètres de deux phases d'un modèle non-linéaire 85
3.1 Introduction . . . 86
3.2 Suppositions, hypothèses, notations . . . 88
3.3 Modèle avec des données complètes . . . 90
3.3.1 Statistique de test . . . 90
3.3.2 Comportement asymptotique de la statistique $\mathcal{Z}_{nk}$ . . . 94
3.4 Modèle avec des données manquantes . . . 101
3.4.1 Statistique de test . . . 103
Vraisemblance empirique basée sur le cas complet . . . 103
Vraisemblance empirique pondérée . . . 106
Vraisemblance empirique avec des valeurs imputées . . . 107
3.4.2 Comportement asymptotique de $\mathcal{Z}_{nk,C}$, $\mathcal{Z}_{nk,W}$ et $\mathcal{Z}_{nk,R}$ . . . 109
3.5 Simulations . . . 110
3.5.1 Modèle avec des données complètes . . . 112
3.5.2 Modèle avec des données manquantes . . . 112
3.6 Annexe . . . 115

II Vraisemblance empirique pour un modèle à deux phases en grande dimension 121

4 Vraisemblance empirique pour un modèle à deux phases en grande dimension 123
4.1 Introduction . . . 124
4.2 Préliminaires . . . 127
4.2.1 Notations . . . 130
4.2.2 Suppositions . . . 131
4.3 Résultat principal . . . 132
4.4 Simulations . . . 157
4.4.1 L'importance des suppositions (A5) et (A6) . . . 157
Analyse du niveaux empiriques. . . . 157
Analyse de la puissance. . . . 159

Amélioration du taux de recouvrement 159
4.5 Annexe 162

Bibliographie **175**

Liste des tableaux

2.5.1 Valeur critique c_α, niveau empirique $\widehat{\alpha}$, puissance $\widehat{\pi}$, α=0.05 67
2.5.2 Statistiques descriptive pour les estimations du point de rupture. 68
2.5.3 Puissance empirique et niveau empirique pour le modèle épidémique, $\alpha = 0.05$, $n = 400$, $\varepsilon \sim \mathcal{N}(0, 1)$. 69

3.5.1 PR et LIC pour les modèles 1 and 2, avec les cas a, b et c. $M = 1000$ 113
3.5.2 PR et LIC, pour l'étude **1**, avec les cas **a**, **b** et **c**, $n = 1000$, $k = 600$. 114
3.5.3 PR et LIC, pour l'étude **2**, avec les cas **a**, **b** et **c**, $n = 1000$, $k = 600$. 114
3.5.4 PR et LIC, pour l'étude **3**, avec les cas **a**, **b** et **c**, $n = 1000$, $k = 600$. 115

4.4.1 Niveaux empiriques avec des erreurs exponentielles et gaussiennes. 158
4.4.2 Niveaux empiriques avec des erreurs exponentielles et gaussiennes, M= 2000, $\theta = 1/2$. 159
4.4.3 Valeur critique empirique $\hat{c}_{1-\alpha/2}$, niveaux empirique $(\hat{\alpha})$, puissances empirique $(\hat{\pi})$, pour des erreurs exponentielles et gaussiennes, M=2000, $p = 50$. 160

Notations et Abréviations

$\mathbb{E}[\mathbf{X}]$	L'espérance de $\mathbf{X}$
$\mathbb{V}ar$	Matrice de variance-covariance
$\mathbb{P}$	Probabilité
$\mathbf{v} = (v_1, \ldots, v_m)^t$	Vecteur de m éléments
$\mathbf{v}^t$	Transposé du vecteur
$\parallel \mathbf{v} \parallel_1$	La norme L_1 du vecteur $\mathbf{v}$, $\parallel \mathbf{v} \parallel_1 = \sum_{j=1}^{m} \mid v_j \mid$.
$\parallel \mathbf{v} \parallel$	La norme euclidienne du vecteur $\mathbf{v}$, $\parallel \mathbf{v} \parallel = (\sum_{j=1}^{m} v_j^2)^{1/2}$.
$\mathbf{A} = (a_{ij})_{\substack{1 \leqslant i \leqslant m_1 \\ 1 \leqslant j \leqslant m_2}}$	Matrice de terme général a_{ij}
$\gamma_p(A)$	La p-ème valeur propre de la matrice $\mathbf{A}$
$tr(\mathbf{A})$	Trace de la matrice $\mathbf{A}$
$\parallel \mathbf{A} \parallel_1$	La norme matricielle subordonnée à la norme vectorielle, $\mathbf{A} = (a_{ij})_{\substack{1 \leqslant i \leqslant m_1 \\ 1 \leqslant j \leqslant m_2}}$, $\parallel \mathbf{A} \parallel_1 = \max_{j=1,\ldots,m_2} (\sum_{i=1}^{m_1} \mid a_{ij} \mid)$
$\mathbf{0}_d$	Vecteur dont tous ces d éléments sont zero
$\mathbf{1}_d$	Vecteur dont tous ces d éléments sont 1
$\mathbb{1}_M$	La fonction indicatrice de M
$\mathbf{I}_d$	Matrice identité de dimension $d \times d$
$\xrightarrow{\mathcal{L}}$	La convergence en loi
$\xrightarrow{\mathbb{P}}$	La convergence en probabilité
$\xrightarrow{p.s.}$	La convergence presque sûr
$\mathcal{N}(0,1)$	La loi normale centrée réduite
$\mathcal{E}xp(d)$	La loi exponentielle de paramètre d
$\chi^2(d)$	La loi de chi-deux à d degrés de liberté

$t(d)$	La loi de student à d degrés de liberté
C	Constante positive qui peut prendre différentes valeurs dans les formules différentes ou même dans différentes parties de la même formule
CUSUM	Somme Cumulée (cumulative sum)
i.i.d	Indépendant et identiquement distribué
MAR	Missing At Random
MC	Moindre carré
LASSO	Least Absolute Shrinkage and Selection Operator
SCAD	Smoothly Clipped Absolute Deviation

Chapitre 1

Introduction

Sommaire

1.1 Préliminaires 25
1.1.1 La vraisemblance empirique 28
1.1.2 Vraisemblance empirique et bootstrap 29
1.1.3 La vraisemblance empirique dans le cas de l'espérance 30
1.2 Organisation de la thèse 32

1.1 Préliminaires

Le problème de points de rupture est le problème d'identification des possibles changements à des instants inconnus et d'estimer la localisation de ces changements dans un processus stochastique. Le problème de points de rupture a été introduit dans le cadre de contrôle de la qualité, puis il a été développé pour devenir un problème fondamental dans différents domaines tels que : la stationnarité d'un processus stochastique, l'estimation de la position actuelle d'une série temporelle, le test et l'estimation du changement dans les paramètres d'un modèle de régression et récemment dans la comparaison et l'appariement de séquences d'ADN dans l'analyse de données de microarray.

On considère des modèles de la forme :

$$Y_i = f(X_i, t_i) + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{1.1}$$

La fonction de régression f comportant $N, N \geq 0$ points de rupture :

$$f(x,t) = h_1(x)\mathbb{1}_{t\leq\tau_1} + h_2(x)\mathbb{1}_{\tau_1<t\leq\tau_2} + \ldots + h_{N+1}(x)\mathbb{1}_{\tau_N<t}, \text{pour } x \in \mathbb{R}^p, t \in \mathbb{R}, \qquad (1.2)$$

où les fonctions $h_1, \ldots, h_{N+1} : \mathbb{R}^p \to \mathbb{R}$ font partie d'une même famille de fonctions $\mathcal{H}$, avec $p \geq 1$. Pour chaque observation i, Y_i est la variable aléatoire modélisée, $\mathbf{X}_i$ est le vecteur des régresseurs, t_i est le moment de mesure et ε_i est l'erreur.

On dit que le modèle (1.2) a $N + 1$ phases. L'intervalle $[\tau_r, \tau_{r+1}]$ pour $r = 0, \ldots, N$ avec $\tau_0 = 1$ et $\tau_{N+1} = n$, s'appelle le r-ème segment et $\tau_1, \ldots, \tau_N$ sont les points de rupture.

Le problème de points de rupture a fait l'objet de recherches intensives dans le dernier demi-siècle. Le problème standard des points de rupture dans les modèles de régression consiste à tester l'hypothèse nulle qu'aucun changement n'a eu lieu dans le modèle contre l'alternative que les observations ont générés deux (ou plus) modèles différents. [Bai and Perron (1998)] considèrent un test d'hypothèse basé sur le rapport de vraisemblance, dans un modèle linéaire $h_k(x) = \beta_k^t x$, pour tester si le modèle contient N ou $N + 1$ points de rupture, sous la suppositions que les erreurs ε_i sont gaussiennes i.i.d. Par ailleurs, [Lombard (1987)], [Mia et al. (1988)] proposent des procédures basées sur des statistiques de rang pour tester s'il y a un ou plusieurs changements. [Horváth et al. (2004)] détectent les changements dans un modèle linéaire par la méthode du CUSUM des résidus. Les rèsultats ont été généralisés par [Ciuperca (2013a)] pour un modèle non-linéaire. [Antoch et al. (2004)] teste la possibilité d'avoir un point de rupture dans un modèle linéaire généralisé, en utilisant une statistique de test basée sur le score maximal. Ensuite, [Antoch et al. (2007)] testent la continuité d'une fonction de régression contre la présence d'un point de saut, sur la base d'estimateurs non-paramétriques localement polynomiaux. Si les variables aléatoires ne sont pas indépendantes, [Huskova et al. (2007)] étudient le problème de la détection d'un changement dans les paramètres d'une série chronologique autorégressive. Toujours, pour des séries chronologiques, [Bardet and Kengne (2014)] ont proposé une nouvelle procédure séquentielle pour détecter les changements dans les paramètres d'un processus $X = (X_t)_{t\in\mathbb{Z}}$ appartenant à une grande classe de modèles causaux.

Pour estimer le nombre de points de rupture dans la moyenne d'une loi normale, [Yao and Au (1989)] proposent un critère de type Schwarz. Avec ce même critère, [Serbinowska (1996)] et [Kühn (2001)] montrent la consistance de l'estimateur du nombre de points de rupture dans le cas des observations indépendantes. Ainsi, [Serbinowska (1996)] considère des variables aléatoires binomiale indépendantes ($Y_i = \mathcal{B}(m_i, q_i)$ avec $q_i = p_r$ pour $\tau_{r-1} < i \leq \tau_r$, $r = 1, \ldots, N, \tau_0 = 1$, $\tau_{N+1} = n$) et [Kühn (2001)] étudie l'estimateur

du nombre de sauts dans un processus stochastique avec saut. Si les fonctions h_r de la relation (1.2) sont constantes, $h_r = \mu_r$ pour tout $r = 1, \ldots, N+1$, [Lavielle (1999)] propose une fonction de contraste pénalisée pour détecter les changements dans une suite de variables aléatoires dépendantes. En supposant que les erreurs ε_i sont soit indépendantes, soit ϕ-mixing, soit α-mixing, [Lavielle (1999)] montre que l'estimateur de N obtenu par la méthode des contrastes est faiblement consistant.

Si le nombre de points de rupture est fixé, l'étape suivante est de les localiser : $\tau_1, \ldots, \tau_N$. Une possible approche est d'utiliser une méthode non-paramétrique d'estimation. [Facer and Muller (2003)] proposent deux estimateurs : un en utilisant une méthode à noyau et un deuxième basé sur la méthode des moindres carrés locale. [Dachian and Yang (2015)] ont considéré le problème de points de rupture dans la fonction de l'intensité, dans les réalisations d'un processus de poisson non-homogène. Ils étudient deux cas : le premier cas quand la taille du saut converge vers une limite non nulle, ensuite quand cette limite est nulle. Les lois limites des rapports de vraisemblance associés à ces deux cas sont très différents.

Dans ce travail, on suppose que le modèle (1.1) est paramétrique et a la forme suivante

$$Y_i = f(X_i, \boldsymbol{\beta}) + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{1.3}$$

On s'intéresse à tester l'hypothèse nulle de non-changement dans les paramètres de régression, contre l'hypothèse alternative d'un seul changement dans ces paramètres (N=1).

Il existe deux types de détection de changements dans un modèle. Le premier type est dite "a posteriori" : une fois que les n observations ont été réalisées, on étudie si à un certain moment $k \in \{2, \ldots, n-1\}$, le modèle (les paramètres $\boldsymbol{\beta}$, pour être plus précis) a changé :

$$Y_i = \begin{cases} f(\mathbf{X}_i; \boldsymbol{\beta}_1) + \varepsilon_i & i = 1, \ldots, k \\ f(\mathbf{X}_i; \boldsymbol{\beta}_2) + \varepsilon_i & i = k+1, \ldots, n. \end{cases} \tag{1.4}$$

Le deuxième type de détection de rupture dans le modèle est dite séquentiel (a priori), où la détection du changement est réalisée en temps réel. Si aucun changement n'a eu lieu dans les k premières observations, on teste le non-changement à l'observation $k+1$. Alors, on teste l'hypothèse que aucun changement n'a eu lieu dans les $k+1$ observations, contre l'hypothèse que le modèle change à l'observation $k+1$ et a la forme suivante

$$\begin{aligned} Y_i &= f(\mathbf{X}_i; \boldsymbol{\beta}) + \varepsilon_i && \text{pour } i = 1, \ldots, k \\ Y_{k+1} &= f(\mathbf{X}_{k+1}; \boldsymbol{\beta}^*) + \varepsilon_{k+1}, \end{aligned} \tag{1.5}$$

avec $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$.

Dans notre travail, on considère le problème de point de rupture a posteriori.

Pour les deux types de problèmes, le nombres de publications dans les dernières années est très vaste. Citons quelques références pour le problème de point de rupture séquentiel. Si la fonction f est de type linéaire, $f(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{x}^t\boldsymbol{\beta}$, [Horváth et al. (2004)] et [Hušková and Kirch (2012)] ont utilisé la technique CUSUM pour trouver une statistique de test qui détecte l'existence du point de rupture. On peut citer aussi les papiers de [Lai and Xing (2010)], [Mei (2006)] et de [Neumeyer and al. (2009)] pour la détection séquentielle de point de rupture.

Pour le problème "a posteriori" de point des rupture, le test d'hypothèse du modèle (1.3) contre le modèle (1.4) ne peut pas être réalisé par les techniques classiques de test d'hypothèse, puisque le modèle (1.4) sous l'hypothèse nulle n'est pas identifiable.

La plupart des articles de la littérature considèrent alors des critères pour détecter s'il y a de point de rupture dans le modèle : voir par exemple les travaux de [Nosek (2010)] et de [Wu (2008)]. Divers tests d'hypothèses ont été proposés uniquement dans le cas d'un modèle linéaire. [Bai (1999)] et [Lee et al. (2015)] ont utilisé la méthode du test du rapport de vraisemblance. [Hall and Sen (1999)] ont considéré une approche non paramétrique basée sur la vraisemblance empirique pour tester le changement dans un modèle linéaire. Toujours en utilisant la méthode de vraisemblance empirique, [Wei et al. (2013)] et [Zi et al. (2010)] ont construit la région de confiance pour la différence entre les coefficients d'un modèle linéaire à deux phases. Pour le modèle de régression quantile, [Qu (2008)] a proposé deux types de statistiques : une statistique basée sur le sous-gradient et une autre sur la base de la statistique de Wald. Pour les modèles linéaires généralisés, [Antoch et al. (2004)] a utilisé une méthode basée sur le score maximal pour tester le changement dans les paramètres de régression.

1.1.1 La vraisemblance empirique

La vraisemblance empirique est une méthode d'estimation inspirée de la méthode du maximum de vraisemblance usuelle. La méthode de vraisemblance empirique a été principalement introduite par [Owen (1988)], [Owen (1990)] et [Owen (2001)], bien qu'on puisse la considérer comme une extension des méthodes de calage utilisées depuis de nombreuses années en sondage (voir [Deville and Sarndal (1992)] et [Hartley and Rao (1968)]). Cette

méthode de type non-paramétrique consiste à maximiser la vraisemblance d'une loi ne chargeant que les données, sous des contraintes satisfaites par le modèle.

[Owen (1990)] a montré que l'on pouvait obtenir une version non-paramétrique du théorème de Wilks. Ce théorème établit la convergence du rapport de vraisemblance empirique vers une loi du χ^2, permettant ainsi de réaliser des tests ou de construire des régions de confiance. Cette méthode proposée a été généralisée à de nombreux modèles, lorsque le paramètre d'intérêt est défini à partir de contraintes de moments ([Qin and Lawless (1994)], [Newey and Smith (2004)]).

Plusieurs auteurs ont utilisé la méthodologie de la vraisemblance empirique dans plusieurs situations statistiques. Pour construire les régions de confiance pour les coefficients dans un modèle de régression linéaire, [Chen (1994)] a proposé une méthode non paramétrique basé sur la vraisemblance empirique. [Qin and al. (2009)], [Xue (2009)] et [Ciuperca (2013a)] ont considéré ce même problème mais pour les modèles avec des donnés manquantes. [Kolaczyk (1994)] a montré que la vraisemblance empirique est justifié comme une méthode de l'inférence pour une classe de modèle linéaire, et en particulier, il a montré comment la vraisemblance empirique peut s'utiliser pour des modèles linéaires généralisés. Pour les modèles avec des points de rupture, on cite les papiers de [Kim and Siegmund (1989)] et de [Liu et al. (2008)]. Pour des modèles épidémiques, [Ning and al. (2012)] ont proposé une méthode basée sur la vraisemblance empirique pour détecter les changements dans la moyenne d'une loi. Toujours avec la vraisemblance empirique, [Zi et al. (2010)] ont construit les régions de confiance pour la différence entre les coefficients de régression de deux phases pour un modèle linéaire avec des données complètes, [Wei et al. (2013)] ont généralisé le même travail pour un modèle linéaire avec des donnés manquantes.

1.1.2 Vraisemblance empirique et bootstrap

Comparant la méthode de vraisemblance empirique avec celle de bootstrap, les avantages principaux de la vraisemblance empirique découlent de l'utilisation de la fonction de vraisemblance.

Contrairement à la méthode bootstrap, [DiCiccio et al. (1991)] a montré que la vraisemblance empirique est corrigeable au sens de Bartlett, ceci signifie que l'on peut corriger la statistique de test d'un facteur qui permet de faire disparaître le biais du premier ordre. Ceci assure un meilleur comportement à distance finie, puisque la différence avec le comportement asymptotique est réduit de $O(n^{-1})$ à $O(n^{-2})$. De plus, la vraisemblance

permet de combiner des données qui proviennent de plusieurs sources avec des différents plans d'échantillonnage. En général, la vraisemblance empirique est facile à calculer par un problème d'optimisation convexe pour lequel l'optimum global unique peut être facilement trouvé par les moindres carrés itérées.

Le bootstrap et la vraisemblance empirique peuvent être combinés de manière efficace. Une manière consiste à utiliser la vraisemblance empirique pour déterminer la famille imbriquée des régions de confiance, et le bootstrap pour choisir lequel utiliser pour un niveau de confiance donné. Les vraisemblances paramétriques et empiriques peuvent être également combiné afin d'avoir des solutions pour des certains problèmes.

1.1.3 La vraisemblance empirique dans le cas de l'espérance

Dans la sous-section suivante, on présente une revue de la littérature traitant de la vraisemblance empirique. Une grande partie de ces résultats énoncés se trouvent dans [Owen (1988)], [Owen (1990)] et [Owen (2001)].

Définition

Soit une suite de vecteurs aléatoires $(\mathbf{X}_i)_{1,\dots,n}$. On suppose que les $\mathbf{X}_i \in \mathbb{R}^p$ sont indépendants et identiquement distribués. On cherche à construire des régions de confiance pour l'espérance μ_0 de $\mathbf{X}$. L'idée de [Owen (1988)] dont on reprend ici la présentation originale, est de construire une vraisemblance en se donnant comme modèle l'ensemble des multinomiales ne chargeant que l'échantillon. On associe ainsi une probabilité $p_i > 0$ à chacune des observations, la somme des ces probabilités étant fixée à 1. On définit une mesure de probabilité

$$\mathcal{T} = \sum_{i=1}^{n} p_i \delta_{\mathbf{X}_i}.$$

La vraisemblance $\mathcal{L}(\mathcal{T})$ du modèle multinomial s'écrit alors

$$\mathcal{L}(\mathcal{T}) = \prod_{i=1}^{n} p_i.$$

Sous la contrainte $\sum_{i=1}^{n} p_i = 1$, $\mathcal{L}$ est maximale en

$$\mathcal{T}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\mathbf{X}_i}.$$

Alors, le rapport de vraisemblance est

$$\mathcal{R}(\mathcal{T}) = \frac{\mathcal{L}(\mathcal{T})}{\mathcal{L}(\mathcal{T}_n)} = \prod_{i=1}^{n} np_i.$$

Puisque plusieurs multinomiales peuvent avoir la même espérance, pour définir le rapport de vraisemblance en une valeur μ du paramètre, [Owen (1990)] propose de déterminer la multinomiale qui maximise le rapport de vraisemblance sous la contrainte

$$\sum_{i=1}^{n} p_i \mathbf{X}_i = \mu.$$

Ce problème peut alors s'écrire

$$\mathcal{R}(\mu) = \sup_{p_1,\dots,p_n} \Big\{ \prod_{i=1}^{n} np_i | \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \mathbf{X}_i = \mu, p_i > 0 \Big\},$$

pour $i = 1, \dots, n$. Afin de résoudre notre problème, on utilise la méthode des multiplicateurs de Lagrange. On considère le log du produit des np_i, ce qui sert à linéariser l'optimisation et les multiplicateurs de Lagrange permettent d'omettre la contrainte de positivité des p_i. On pose le processus aléatoire suivant

$$\sum_{i=1}^{n} \log(np_i) - n\boldsymbol{\lambda} \sum_{i=1}^{n} p_i(\mathbf{X}_i - \mu) - n\gamma(\sum_{i=1}^{n} p_i - 1), \tag{1.6}$$

où $\boldsymbol{\lambda} \in \mathbb{R}^p$ et $\gamma \in \mathbb{R}$ sont les multiplicateurs de Lagrange.
On annule la dérivée du processus (1.6) par rapport à p_i, on obtient

$$\frac{1}{p_i} - n\boldsymbol{\lambda}(\mathbf{X}_i - \mu) - n\gamma = 0. \tag{1.7}$$

En multipliant (1.7) par p_i, puis en faisant la sommation sur i, on obtient

$$\sum_{i=1}^{n} (1 - n\boldsymbol{\lambda} p_i(\mathbf{X}_i - \mu) - np_i\gamma) = 0.$$

En utilisant les contraintes $\sum_{i=1}^{n} p_i = 1$ et $\boldsymbol{\lambda} \sum_{i=1}^{n} p_i(\mathbf{X}_i - \mu) = 0$, on obtient que $n = n\gamma$, donc $\gamma = 1$. Alors, grâce à la relation (1.7), la probabilité p_i devient

$$p_i = \frac{1}{n(1 + \boldsymbol{\lambda}(\mathbf{X}_i - \mu))}.$$

Le rapport de vraisemblance empirique permet de construire des régions de confiance asymptotiques pour μ_0. Pour ceci, on définit une région de confiance $\mathcal{C}_{\eta,n}$, par inversion du test du rapport de vraisemblance

$$\mathcal{C}_{\eta,n} = \Big\{\mu | -2\log(\mathcal{R}(\mu)) \leq \eta\Big\},$$

où η est un paramètre qui permet de régler le niveau de confiance que l'on veut atteindre. On note la statistique de test par

$$\mathcal{Z}_n(\mu) = -2\log(\mathcal{R}(\mu)).$$

Théorème de convergence

Le théorème suivant a été établi par [Owen (1988)] et [Owen (1990)].

Théorème 1.1.1. ***(Owen)*** *Si $\mathbf{X}_1, \ldots, \mathbf{X}_n$ sont des vecteurs aléatoires de $\mathbb{R}^p$, indépendants et identiquement distribués, d'espérance μ_0 et de variance Σ, de rang $q > 0$, alors, pour tout $\eta \in \mathbb{R}$, $\mathcal{C}_{\eta,n}$ est convexe et*

$$\mathbb{P}(\mu_0 \in \mathcal{C}_{\eta,n}) = \mathbb{P}(\mathcal{Z}_n(\mu_0) \leq \eta) \underset{n\to\infty}{\longrightarrow} F_{\chi^2(q)}(\eta),$$

où $F_{\chi^2(q)}(.)$ est la fonction de répartition de la distribution de $\chi^2(q)$.

1.2 Organisation de la thèse

Le corps de cette thèse se présente en deux parties. Dans la première partie (Chapitres 2 et 3), on travaille sur des modèles de régression non-linéaires. Dans le Chapitre 2, on propose une méthode non paramétrique basée sur la vraisemblance empirique pour tester la présence de changement dans les paramètres de modèle de régression non-linéaire en testant l'hypothèse nulle de non-changement dans les paramètres de régression, contre l'hypothèse alternative d'un seul changement dans les paramètres de régression. Sous l'hypothèse nulle, on prouve la consistance et la vitesse de convergence des estimateurs de paramètres de régression. Avec des simulations de Monte-Carlo on montre la performance de la statistique de test proposée. Dans ce chapitre on étudie également le modèle épidémique avec deux points de rupture. Ce chapitre fait l'objet d'un article publié dans la revue *Metrika* en collaboration avec *Gabriela Ciuperca* ([Ciuperca and Salloum (2015a)]).

Toujours avec la vraisemblance empirique, dans le Chapitre 3 on construit les régions de confiance pour la différence entre les paramètres des deux phases d'un modèle

non-linéaire. On montre que le rapport de vraisemblance empirique a une distribution asymptotique χ^2. Le modèle non-linéaire avec des variables réponse manquantes au hasard (Missing At Random (MAR)) sera également considéré. Ainsi, on construit les régions de confiance du différence entre les paramètres des deux phases d'un modèle, en proposant trois statistiques de vraisemblance empiriques : la vraisemblance empirique basée sur les données complètes, la vraisemblance empirique pondérée et la vraisemblance empirique par des valeurs imputées. On prouve que les trois rapports de vraisemblance empiriques ont des distributions asymptotiques χ^2. Afin de démontrer l'efficacité des approches proposées on réalise des simulations de Monte-Carlo pour calculer les probabilités de recouvrement et les longueurs des intervalles de confiance obtenus. L'article correspondant est soumis ([Salloum(2015)]).

Dans la deuxième partie (composée du Chapitre 4), on considère les modèles de régression linéaire en grande dimension. Dans le Chapitre 4, on construit une statistique de test basée sur la vraisemblance empirique pour détecter le changement dans les coefficients d'un modèle linéaire en grande dimension, où le nombre des variables du modèle peut augmenter avec la taille de l'échantillon. En utilisant les comportements asymptotiques théoriques de la statistique du rapport de vraisemblance empirique, on propose une statistique de test pour l'utiliser dans la pratique. Sous l'hypothèse nulle, on prouve la normalité asymptotique de la statistique de test proposée. Ce résultat est différent de celui obtenu pour le cas où le nombre des variables est fixé, quand la loi asymptotique est une χ^2. Des simulations de Monte-Carlo seront considérées afin d'étudier les comportements de la statistique de test proposée. L'article correspondant est soumis en collaboration avec *Gabriela Ciuperca* ([Ciuperca and Salloum (2015b)]).

Remarque : Pour ne pas confondre les notations dans ce manuscrit, on a considéré que les Chapitres 2, 3 et 4 sont indépendants.

Première partie

Test par la vraisemblance empirique pour de modèles non-linéaire avec points de rupture

Chapitre 2

Test de la vraisemblance empirique pour un modèle non-linéaire avec points de rupture

Sommaire

2.1 Introduction 38
2.2 Suppositions 40
2.3 Test avec un seul point de rupture 41
2.3.1 Hypothèses 41
2.3.2 Statistique de test 42
2.3.3 Comportement asymptotique de la statistique de test 47
2.4 Test d'un modèle épidémique non-linéaire 63
2.5 Simulations 65
2.5.1 Modèle avec un seul point de rupture 66
2.5.2 Modèle épidémique 68
2.6 Annexe 69

Abstract : In this chapter, in order to test whether changes have occurred in a nonlinear parametric regression, we propose a nonparametric method based on the empirical likelihood. Firstly, we test the null hypothesis of no-change against the alternative of one change in the regression parameters. Under null hypothesis, the consistency and the convergence rate of the regression parameter estimators are proved. The asymptotic distribution of the test

statistic under the null hypothesis is obtained, which allows to find the asymptotic critical value. On the other hand, we prove that the proposed test statistic has the asymptotic power equal to 1. These theoretical results allows find a simple test statistic, very useful for applications. The epidemic model, a particular model with two change-points under the alternative hypothesis, is also studied. Numerical studies by Monte-Carlo simulations show the performance of the proposed test statistic.

Key-words : Change-point ; Nonlinear parametric model ; Empirical likelihood test ; Asymptotic behaviour.

Résumé : Dans ce chapitre, on propose une méthode basée sur la vraisemblance empirique pour tester la présence de changement dans les paramètres d'un modèle de régression non-linéaire. Tout d'abord, on teste l'hypothèse nulle de non-changement dans les paramètres de régression, contre l'hypothèse alternative d'un seul changement dans ces paramètres. Sous l'hypothèse nulle, on prouve la consistance et la vitesse de convergence des estimateurs des paramètres de régression. La loi asymptotique de la statistique de test sous l'hypothèse nulle obtenue nous permet de trouver la valeur critique asymptotique. D'autre part, on prouve que la puissance asymptotique de la statistique de test proposée est égale à 1. Les résultats théoriques obtenus nous permettront de trouver une simple statistique de test qui sera utilisé aisément dans les applications. Le cas particulier du modèle épidémique avec deux points de rupture est également étudié. Afin de montrer la performance de la statistique de test proposée, des études numériques par des simulations de Monte-Carlo ont été considérées.

Mots-clefs : Point de rupture. Modèle paramétrique non-linéaire. Test de la vraisemblance empirique. Comportement asymptotique.

2.1 Introduction

Un modèle de régression paramétrique "classique" a la forme :

$$Y_i = f(\mathbf{X}_i; \boldsymbol{\beta}) + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{2.1}$$

Pour chaque observation i, Y_i désigne la variable réponse. On considère par $\mathbf{X}_i \in \mathbb{R}^p$ des vecteurs aléatoires indépendantes et identiquement distribuées, de fonction de répartition $H(\mathbf{x})$, avec $\mathbf{x} \in \Upsilon$, $\Upsilon \subseteq \mathbb{R}^p$. ε_i est l'erreur d'espérance zéro et de variance σ^2.
La suite de vecteurs aléatoires $(\mathbf{X}_i, \varepsilon_i)_{1 \leq i \leq n}$ est indépendante et identiquement distribués

(i.i.d), avec la même distribution de $(\mathbf{X}, \varepsilon)$. Pour tout i, ε_i est indépendante de $\mathbf{X}_i$. La fonction de régression $f : \Upsilon \times \Gamma \to \mathbb{R}$, avec $\Upsilon \subseteq \mathbb{R}^p$ et $\Gamma \subseteq \mathbb{R}^d$, est de forme explicite connue. L'ensemble de paramètres Γ est supposé compact.
Dans la suite, pour $\mathbf{x} \in \Upsilon$ et $\boldsymbol{\beta} \in \Gamma$, on utilise les notations : $\dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta}) \equiv \partial f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$ et $\ddot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta}) \equiv \partial^2 f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}^2$.

Dans ce chapitre, on considère le problème de points de rupture dans un modèle non-linéaire général, par la méthode de vraisemblance empirique. On généralise le travail de [Liu et al. (2008)] où le modèle linéaire était considéré. Une des difficultés qui présente dans l'étude du modèle non-linéaire par rapport au modèle linéaire est lié au fait que les fonctions scores correspondantes pour trouver la statistique de test dépendent des paramètres de régression, et surtout, la forme analytique de ces dérivées est inconnue. D'autre part, pour les modèles linéaires, des nombreuses preuves sont basées sur la convexité de la fonction de régression par rapport au paramètre de régression. Alors, la valeur extrême de la fonction convexe est atteint sur les bornes de l'intervalle considéré. Ces deux facteurs produisent des difficultés dans l'étude théorique de la statistique de test associée dans l'étude des modèles non-linéaire.

L'intervention de points de rupture dans l'estimation de paramètre de régression entraîne une difficulté supplémentaire dans l'étude de propriétés de la statistique de test. Le papier de [Boldea and Hall (2013)] est le seul papier qui a considéré le test d'hypothèse dans un modèle non-linéaire avec de points de rupture en utilisant la méthode des moindres carrés. Cette méthode par rapport à celle du vraisemblance empirique, a l'inconvénient qu'elle est sensible aux données aberrantes. En outre, dans la Section 2.3 on va voir que les suppositions imposées par [Boldea and Hall (2013)] sont beaucoup plus fortes que celles du présent travail.
Notons que [Hall and Sen (1999)] ont testé la stabilité structurelle dans le modèle non-linéaire par la méthode des moments généralisés, et non pas le changement des paramètres de régression.

Ce chapitre est organisé comme suit. Dans la Section 2.3 on construit une statistique pour tester le changement dans les paramètres de régression dans un modèle non-linéaire. Le comportement asymptotique de la statistique de test sous l'hypothèse nulle et sous son alternative est considéré dans la Section 2.4. Dans la Section 2.5, on donne des simulations par Monte-Carlo pour montrer la performance du test proposé, concernant le niveau empirique, la puissance empirique et l'estimation du point de rupture, en particulier quand la distribution de l'erreur est non-gaussienne. Des lemmes avec leurs preuves, nécessaires

pour les démonstrations des théorèmes et des propositions, sont donnés dans la Section 2.6.

2.2 Suppositions

En ce qui concerne la variable aléatoire ε, on fait la supposition suivante :

(A1) $\mathbb{E}[\varepsilon_i] = 0$ et $\mathbb{E}[\varepsilon_i^2] < \infty$, pour tout $i = 1, \ldots, n$.

Soit $\eta > 0$ quelconque et le voisinage de $\boldsymbol{\beta^0}$: $\mathcal{V}_\eta(\boldsymbol{\beta^0}) = \{\boldsymbol{\beta} \in \Gamma; \|\boldsymbol{\beta} - \boldsymbol{\beta^0}\| \leq \eta\}$.

La fonction de régression $f : \Upsilon \times \Gamma \to \mathbb{R}$ et le vecteur aléatoire $\mathbf{X}$ satisfont les conditions suivantes :

(A2) pour tout $\mathbf{x} \in \Upsilon$ et pour $\boldsymbol{\beta} \in \Gamma$, la fonction $f(\mathbf{x}, \boldsymbol{\beta})$ est trois fois dérivable en $\boldsymbol{\beta}$ et continue en Υ.

(A3) Il existe $\eta_0 > 0$ tel que pour tout $\eta \in [0, \eta_0]$, $(|\frac{\partial^2 f(\mathbf{x},\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k}|)_{1\leq j,k\leq d}$ et $(|\frac{\partial^3 f(\mathbf{x},\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k\partial\beta_l}|)_{1\leq j,k,l\leq d}$ sont bornées, pour tout $\mathbf{x} \in \Upsilon$ et tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$.

(A4) Il existe $\eta_0 > 0$ tel que pour tout $\eta \in [0, \eta_0]$, $\mathbb{E}[\|\dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\beta})\|_1] < \infty$, $\mathbb{E}[\|\dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\beta})\dot{\mathbf{f}}^t(\mathbf{X}, \boldsymbol{\beta})\|_1] < \infty$ et $\mathbb{E}[|\frac{\partial^2 f(\mathbf{X},\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k}|] < \infty$, pour tout $1 \leq j, k \leq d$ et tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$.

Les suppositions (A3) et (A4) sont des conditions standard pour des modèles non-linéaires, voir par exemple le livre de [Seber and Wild (2003)]. On remarque que la supposition (A4) est plus faible que celle employée dans le papier de [Boldea and Hall (2013)], dont la méthode des moindres carrés a été utilisé pour tester l'hypothèse de non-changement.

Notons que, les suppositions du papier de [Boldea and Hall (2013)] sont : $\mathbb{E}[\varepsilon_i^{2s}] < \infty$, $\sup_{\boldsymbol{\beta}} \mathbb{E}[f(\mathbf{X}, \boldsymbol{\beta})]^{2s} < \infty$, $\sup_{\boldsymbol{\beta}} \mathbb{E}[\|\dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\beta})\|]^{2s} < \infty$ et pour tout $1 \leq j, k \leq d$, $\sup_{\boldsymbol{\beta}} \mathbb{E}[|\frac{\partial^2 f(\mathbf{X},\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k}|^s] < \infty$, pour $s > 2$.

2.3 Test avec un seul point de rupture

Dans cette section, on teste l'hypothèse de non-changement dans les paramètres de régression du modèle non-linéaire donné par la relation (2.1) contre l'hypothèse qu'il y a un seul changement à l'observation k inconnue, dans le paramètre, ce qui implique le modèle (1.4).

2.3.1 Hypothèses

Sous l'hypothèse H_0, on suppose qu'aucun changement n'a eu lieu dans les paramètres du modèle, donc on a la forme (2.1). Alors, l'hypothèse nulle en fonction des paramètres du modèle (1.4) peut s'écrire

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}. \tag{2.2}$$

L'hypothèse alternative suppose la présence d'un seul changement dans les paramètres de régression,

$$H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2. \tag{2.3}$$

Sous l'hypothèse nulle H_0, on désigne par $\boldsymbol{\beta^0}$ la vraie valeur (supposé inconnue) du paramètre $\boldsymbol{\beta}$. Sous l'hypothèse H_1, $\boldsymbol{\beta}_1^0$ et $\boldsymbol{\beta}_2^0$ (aussi inconnus) sont les vrais paramètres de $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$.

Considérons des notations supplémentaires. Tout d'abord on donne des notations qui seront utilisées sous l'hypothèse nulle H_0. Soit le vecteur aléatoire, de dimension d, suivant

$$\mathbf{g}(\mathbf{X}_i, \boldsymbol{\beta}) \equiv \mathbf{g}_i(\boldsymbol{\beta}) \equiv \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta})[Y_i - f(\mathbf{X}_i, \boldsymbol{\beta})].$$

On remarque que, sous l'hypothèse H_0, pour tout $i = 1, \ldots, n$, on a $\mathbf{g}_i(\boldsymbol{\beta^0}) = \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta^0})\varepsilon_i$ et $\mathbb{E}[\mathbf{g}_i(\boldsymbol{\beta^0})] = \mathbf{0}_d$.
On considère également la matrice suivante de dimension $d \times d$

$$\mathbf{V} \equiv \mathbb{E}[\dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta^0})\dot{\mathbf{f}}^t(\mathbf{X}_i, \boldsymbol{\beta^0})].$$

Alors, $\mathbb{V}\mathrm{ar}\,[\varepsilon_i \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta^0})] = \sigma^2\mathbf{V}$.

On considère maintenant les deux ensembles d'entiers naturels $I \equiv \{1, ..., k\}$ et $J \equiv \{k+1, ..., n\}$. Ces deux ensembles contiennent les indices d'observations des deux segments du modèle (1.4). Correspondant à ces deux ensembles, soient les vecteurs de probabilité : $(p_1, \ldots, p_k)$ et $(q_{k+1}, \ldots, q_n)$. Évidemment, on a que

$$\sum_{i \in I} p_i = 1 \text{ et } \sum_{j \in J} q_j = 1.$$

2.3.2 Statistique de test

Sous l'hypothèse H_0 donnée par la relation (2.2), le profil de la vraisemblance empirique pour $\boldsymbol{\beta}$ est

$$\mathcal{R}_{0,nk}(\boldsymbol{\beta}) = \sup_{(p_1,\ldots,p_k)} \sup_{(q_{k+1},\ldots,q_n)} \Big\{ \prod_{i\in I} p_i \prod_{j\in J} q_j ; \sum_{i\in I} p_i = 1; \sum_{j\in J} q_j = 1, \\ \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\},$$

Sans la contrainte $\sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{0}_d$, les maximums de $\prod_{i\in I} p_i$, $\prod_{j\in J} q_j$ sont atteints respectivement pour $p_i = k^{-1}$ et $q_j = (n-k)^{-1}$. Alors, le profil du rapport de la vraisemblance empirique pour $\boldsymbol{\beta}$ a la forme

$$\mathcal{R}'_{0,nk}(\boldsymbol{\beta}) = \sup_{(p_1,\ldots,p_k)} \sup_{(q_{k+1},\ldots,q_n)} \Big\{ \prod_{i\in I} kp_i \prod_{j\in J} (n-k) q_j ; \sum_{i\in I} p_i = 1, \sum_{j\in J} q_j = 1, \\ \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\}. \tag{2.4}$$

De même, sous l'hypothèse H_1 donnée par la relation (2.3), le profil de la vraisemblance empirique est

$$\mathcal{R}_{1,nk}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sup_{(p_1,\ldots,p_k)} \sup_{(q_{k+1},\ldots,q_n)} \Big\{ \prod_{i\in I} p_i \prod_{j\in J} q_j ; \sum_{i\in I} p_i = 1, \sum_{j\in J} q_j = 1, \\ \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}_1) = \mathbf{0}_d, \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}_2) = \mathbf{0}_d \Big\}.$$

Alors, le profil du rapport de vraisemblance empirique pour $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ a la forme

$$\mathcal{R}'_{1,nk}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sup_{(p_1,\ldots,p_k)} \sup_{(q_{k+1},\ldots,q_n)} \Big\{ \prod_{i\in I} kp_i \prod_{j\in J} (n-k) q_j , \sum_{i\in I} p_i = 1, \sum_{j\in J} q_j = 1, \\ \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}_1) = \mathbf{0}_d, \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}_2) = \mathbf{0}_d \Big\}.$$

Pour tester l'hypothèse H_0 contre H_1, on utilise une idée similaire au test du maximum de la vraisemblance. On considère alors, le profil du rapport de la vraisemblance empirique suivant

$$\frac{\mathcal{R}_{0,nk}(\boldsymbol{\beta})}{\mathcal{R}_{1,nk}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)} = \frac{\mathcal{R}'_{0,nk}(\boldsymbol{\beta})}{\mathcal{R}'_{1,nk}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}, \tag{2.5}$$

Afin de trouver une forme simple pour la statistique de test, on étudie le comportement du processus du dénominateur de la relation (2.5).

Le résultat suivant désigne une généralisation de la version non paramétrique du Théorème de Wilks. Plus spécifiquement, sous l'hypothèse H_1, puisque les observations sont indépendantes, on peut appliquer le théorème de Wilks sur chacun de deux segments. Alors, on prouve que sous l'hypothèse H_1, le profil du rapport de vraisemblance empirique pour $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ a une distribution asymptotique χ^2. Notons que, le théorème suivant est valable avec les vraies valeurs de $\boldsymbol{\beta}_1$ et $\boldsymbol{\beta}_2$ qui sont respectivement $\boldsymbol{\beta}_1^0$ et $\boldsymbol{\beta}_2^0$. Pour de raison de lisibilité, on les note par $\boldsymbol{\beta}_1$ et $\boldsymbol{\beta}_2$.

Théorème 2.3.1. *Sous les suppositions (A1)-(A3), si l'hypothèse* H_1 *est vraie, alors*

$$-2\log \mathcal{R}'_{1,nk}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2(2d).$$

Preuve. Sous l'hypothèse H_1, pour le premier segment généré par les observations $i \in I$, le profil de la vraisemblance empirique pour $\boldsymbol{\beta}_1$, pour une valeur de k fixée, est

$$\sup_{(p_1,\ldots,p_k)} \left\{ \prod_{i\in I} kp_i; 0 \le p_i \le 1, \sum_{i\in I} p_i = 1, \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}_1) = \mathbf{0}_d \right\}.$$

En utilisant la méthode des multiplicateurs de Lagrange, on pose le processus aléatoire suivant

$$\sum_{i\in I} \log p_i + \eta(\sum_{i\in I} p_i - 1) - k\boldsymbol{\lambda}_1^t(\sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}_1)), \tag{2.6}$$

où $\boldsymbol{\lambda}_1 \in \mathbb{R}^d$ et $\eta \in \mathbb{R}$ sont les multiplicateurs de Lagrange.

Annulant la dérivée par rapport à p_i du processus (2.6), on obtient

$$p_i = \frac{1}{k\boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta}_1) - \eta}. \tag{2.7}$$

Pour calculer η, par la relation (2.7) on a que

$$1 + \eta p_i - kp_i \boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta}_1) = 0,$$

en faisant la somme pour $i = 1, \ldots, k$, on obtient

$$\sum_{i\in I} kp_i \boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta}_1) - \sum_{i\in I} \eta p_i = k. \tag{2.8}$$

La relation (2.8) avec les contraintes $\sum_{i\in I} kp_i\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta}_1)=0$ et $\sum_{i\in I} p_i=1$ impliquent que

$$\eta=-k.$$

Alors, les probabilités p_i données par la relation (2.7) deviennent

$$p_i=\frac{1}{k(1+\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta}_1))}.$$

De même, le profil de la vraisemblance empirique pour le deuxième segment généré par les observations $j\in J$, est

$$\sup_{(q_{k+1},\ldots,q_n)}\left\{\prod_{j\in J}(n-k)q_j; 0\leq q_j\leq 1, \sum_{j\in J} q_j=1, \sum_{j\in J} q_j\mathbf{g}_j(\boldsymbol{\beta}_2)=\mathbf{0}_d\right\}.$$

Cette fonction est maximisé pour

$$q_j=\frac{1}{(n-k)(1-\boldsymbol{\lambda}_2^t\mathbf{g}_i(\boldsymbol{\beta}_2))},$$

où $\boldsymbol{\lambda}_2\in\mathbb{R}^p$ est le multiplicateur de Lagrange. Alors, la statistique du rapport de log-vraisemblance empirique peut s'écrire

$$-2\log\mathcal{R}'_{1,nk}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=2\sum_{i\in I}\log\left[1+\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta}_1)\right]+2\sum_{j\in J}\log\left[1-\boldsymbol{\lambda}_2^t\mathbf{g}_i(\boldsymbol{\beta}_2)\right]. \tag{2.9}$$

D'après le Théorème 2 de [Qin and Lawless (1994)] et le Théorème 4.1 de [Ciuperca (2013a)], en utilisant les suppositions (A1), (A2) et (A3), on a que chacune de deux sommes de la partie droite de la relation (2.9) converge en loi vers une distribution de $\chi^2(d)$. La preuve se termine en tenant compte du fait que les deux termes de la relation (2.9) contiennent des observations indépendantes. ■

Conséquence de ce théorème, sous l'hypothèse H_1, le dénominateur du rapport de vraisemblance empirique donné par la relation (2.5), ne dépend pas asymptotiquement de paramètres $\boldsymbol{\beta}_1$ et $\boldsymbol{\beta}_2$. Alors, on peut considérer comme statistique de test le terme $-2\log\mathcal{R}'_{0,nk}(\boldsymbol{\beta})$.

Tenant compte de l'expression $\mathcal{R}'_{0,nk}(\boldsymbol{\beta})$ donnée par la relation (2.4) et en utilisant la

méthode des multiplicateurs de Lagrange, maximiser le terme $-2\log\mathcal{R}'_{0,nk}(\boldsymbol{\beta})$ est équivalent à maximiser la statistique de test suivante par rapport à $\boldsymbol{\beta}$, η_1, η_2, $\boldsymbol{\lambda}_1$ et $\boldsymbol{\lambda}_2$:

$$\sum_{i\in I}\big[\log p_i - n\boldsymbol{\lambda}_1^t p_i\mathbf{g}_i(\boldsymbol{\beta})\big] + \sum_{j\in J}\big[\log q_j + n\boldsymbol{\lambda}_2^t q_j\mathbf{g}_j(\boldsymbol{\beta})\big] + \eta_1\big(\sum_{i\in I} p_i - 1\big) + \eta_2\big(\sum_{j\in J} q_j - 1\big), \tag{2.10}$$

où $\boldsymbol{\beta}\in\Gamma$, $\eta_1,\eta_2\in\mathbb{R}$ et $\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2\in\mathbb{R}^d$.

Puisque les dérivées de (2.10) par rapport à p_i et q_j sont nulles, en utilisant un argument similaire à celui de la preuve du Théorème 2.3.1, on obtient que

$$p_i = \frac{1}{k + n\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta})}, \qquad q_j = \frac{1}{n-k - n\boldsymbol{\lambda}_2^t\mathbf{g}_j(\boldsymbol{\beta})}. \tag{2.11}$$

Alors, la statistique $-2\log\mathcal{R}'_{nk,0}(\boldsymbol{\beta})$ devient

$$2\sum_{i\in I}\log\big[1 + \frac{n}{k}\boldsymbol{\lambda}_1^t\mathbf{g}_i(\boldsymbol{\beta})\big] + 2\sum_{j\in J}\log\big[1 - \frac{n}{n-k}\boldsymbol{\lambda}_2^t\mathbf{g}_j(\boldsymbol{\beta})\big]. \tag{2.12}$$

Tenant compte des probabilités p_i et q_j obtenues par la relation (2.11), la dérivée de (2.12) par rapport à $\boldsymbol{\beta}$ peut s'écrire

$$2n\big[\sum_{i\in I} p_i\boldsymbol{\lambda}_1^t\dot{\mathbf{g}}_i(\boldsymbol{\beta}) - \sum_{j\in J} q_j\boldsymbol{\lambda}_2^t\dot{\mathbf{g}}_j(\boldsymbol{\beta})\big] = 0,$$

où $\dot{\mathbf{g}}_i(\boldsymbol{\beta})$ est une matrices de dimension $d\times d$, qui contient les dérivées du vecteur $\mathbf{g}_i(\boldsymbol{\beta})$ par rapport à $\boldsymbol{\beta}$, pour $i = 1,\ldots,k$.

Pour avoir un seul paramètre $\boldsymbol{\lambda}$, on se limitera à un cas particulier, quand $\boldsymbol{\lambda}_1$ et $\boldsymbol{\lambda}_2$ satisfont la contrainte :

$$\mathbf{V}_{1n}(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \mathbf{V}_{2n}(\boldsymbol{\beta})\boldsymbol{\lambda}_2,$$

avec

$$\mathbf{V}_{1n}(\boldsymbol{\beta}) \equiv \frac{1}{k}\sum_{i\in I}\dot{\mathbf{g}}_i(\boldsymbol{\beta}), \qquad \mathbf{V}_{2n}(\boldsymbol{\beta}) \equiv \frac{1}{n-k}\sum_{j\in J}\dot{\mathbf{g}}_j(\boldsymbol{\beta}).$$

Pour le vrai paramètre $\boldsymbol{\beta}^{\mathbf{0}}$, ces deux dernières matrices deviennent, respectivement

$$\mathbf{V}^0_{1n} \equiv \mathbf{V}_{1n}(\boldsymbol{\beta}^{\mathbf{0}}), \qquad \mathbf{V}^0_{2n} \equiv \mathbf{V}_{2n}(\boldsymbol{\beta}^{\mathbf{0}}).$$

Tenant compte de la contrainte $\mathbf{V}_{1n}(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \mathbf{V}_{2n}(\boldsymbol{\beta})\boldsymbol{\lambda}_2$, la statistique de test de la relation (2.12) devient

$$2\sum_{i\in I}\log\left[1+\frac{n}{k}\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})\right]+2\sum_{j\in J}\log\left[1-\frac{n}{n-k}\boldsymbol{\lambda}^t\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})\right]. \tag{2.13}$$

D'autre part, afin que les paramètres appartiennent à un ensemble borné, à la place du paramètre k, on considère $\theta_{nk} \equiv k/n$ et on note la statistique donnée par la relation (2.13) par $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$.

Sous l'hypothèse H_1, si k_0 est le vrai point où le modèle subit un changement, on note ce point par $\theta_{0n} = k_0/n$.

D'une façon similaire au test de maximum de la vraisemblance classique, mais pour les modèles sans point de rupture, afin d'étudier la statistique de test de maximum de la log-vraisemblance empirique, on calcule les fonctions scores associées à la statistique de test donnée par (2.13). Alors on obtient

$$\begin{aligned}\boldsymbol{\phi}_{1n}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta}) &\equiv \frac{\partial Z_{nk}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})}{2\partial\boldsymbol{\lambda}}\\ &= \sum_{i\in I}\frac{\mathbf{g}_i(\boldsymbol{\beta})}{\theta_{nk}+\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})}-\sum_{j\in J}\frac{\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}{1-\theta_{nk}-\boldsymbol{\lambda}^t\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}.\end{aligned} \tag{2.14}$$

$$\begin{aligned}\boldsymbol{\phi}_{2n}(\theta_{nk},\boldsymbol{\lambda},\beta) &\equiv \frac{\partial Z_{nk}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})}{2\partial\boldsymbol{\beta}}\\ &= \sum_{i\in I}\frac{\dot{\mathbf{g}}_i(\boldsymbol{\beta})\boldsymbol{\lambda}^t}{\theta_{nk}+\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{g}_i(\boldsymbol{\beta})}-\sum_{j\in J}\frac{\partial(\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta}))/\partial\boldsymbol{\beta}}{1-\theta_{nk}-\boldsymbol{\lambda}^t\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}\boldsymbol{\lambda}^t.\end{aligned} \tag{2.15}$$

En résolvant le système suivant

$$\begin{cases}\boldsymbol{\phi}_{1n}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})=\mathbf{0}_d\\ \boldsymbol{\phi}_{2n}(\theta_{nk},\boldsymbol{\lambda},\boldsymbol{\beta})=\mathbf{0}_d,\end{cases} \tag{2.16}$$

on obtient le couple des solutions $(\widehat{\boldsymbol{\lambda}}(\theta_{nk}), \widehat{\boldsymbol{\beta}}(\theta_{nk}))$ qui maximisent la statistique de test donnée par (2.13). Remarquons que ces solutions dépendent du point de rupture k.

Finalement, on obtient le profil du maximum de vraisemblance empirique $Z_{nk}(\theta_{nk}, \widehat{\boldsymbol{\lambda}}(\theta_{nk}), \widehat{\boldsymbol{\beta}}(\theta_{nk}))$, qui dépend seulement du paramètre du point de rupture θ_{nk}.
Notons que, par rapport au modèle linéaire, dans notre cas les matrices $\mathbf{V}_{1n}(\boldsymbol{\beta})$, $\mathbf{V}_{2n}(\boldsymbol{\beta})$ et la dérivée $\dot{\mathbf{g}}(\boldsymbol{\beta})$ dépendent de $\boldsymbol{\beta}$. Ces dépendances, ainsi que la non-linéarité de $\mathbf{g}(\boldsymbol{\beta})$, produisent des difficultés dans l'étude de la statistique $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$ et des solutions $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$ et $\widehat{\boldsymbol{\beta}}(\theta_{nk})$.

2.3.3 Comportement asymptotique de la statistique de test

Dans cette sous-section, pour les probabilités p_i et q_j données par la relation (2.11), on prouve que sous la contrainte $\mathbf{V}_{1n}(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \mathbf{V}_{2n}(\boldsymbol{\beta})\boldsymbol{\lambda}_2$, les deux termes kp_i et $(n-k)q_j$ peuvent être encadrés par deux constantes strictement positives, ce qui implique que la statistique de test $Z_{nk}(\theta_{nk}, \widehat{\boldsymbol{\lambda}}(\theta_{nk}), \widehat{\boldsymbol{\beta}}(\theta_{nk}))$ est bien définie.

Tenant compte des propriétés établies pour les solutions $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$ et $\widehat{\boldsymbol{\beta}}(\theta_{nk})$ de la relation (2.13), on peut considérer à la place de la statistique donnée par (2.13), une statistique de test plus simple qui sera donnée plus tard dans la relation (2.29).

Dans cette sous-section, on étudie le comportement asymptotique de la statistique de test obtenue dans la relation (2.13), d'abord sous l'hypothèse nulle H_0 et ensuite sous l'alternative H_1.

Comportement asymptotique sous H_0

Afin d'étudier le comportement asymptotique de $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$ et $\widehat{\boldsymbol{\beta}}(\theta_{nk})$, on commence par étudier kp_i, pour $i \in I$, et $(n-k)q_j$, pour $j \in J$, avec les probabilités p_i, q_j données par la relation (2.11).

Plus précisément, on montre que si $\boldsymbol{\beta}$ est dans le voisinage de $\boldsymbol{\beta^0}$, kp_i et $(n-k)q_j$ peuvent être encadrés par deux constantes strictement bornées et positives, avec une probabilité proche de 1.

Proposition 2.3.1. *Sous l'hypothèse* H_0*, on suppose que (A1)-(A4) sont vérifiées. Il existe* $\eta_0 > 0$ *tel que pour tout* $\eta \in [0, \eta_0]$ *on a*

(i) Pour $i \in I$*, pour tout* $\epsilon > 0$*, il existe deux constantes positives* M_1 *et* M_2*, telle*

que, pour tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$

$$\mathbb{P}\Big[\frac{1}{M_2} \leq \frac{1}{1+\frac{\boldsymbol{\lambda}^t}{\theta_{nk}} g_i(\boldsymbol{\beta})} \leq \frac{1}{M_1}\Big] \geq 1-\epsilon. \tag{2.17}$$

(ii) Pour $j \in J$*, pour tout* $\epsilon > 0$*, il existe deux constantes positives* M_3 *et* $M_4 > 0$*, telle que, pour tout* $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$

$$\mathbb{P}\Big[\frac{1}{M_4} \leq \frac{1}{1-\frac{\boldsymbol{\lambda}^t}{1-\theta_{nk}} \mathbf{V}_{1n}(\boldsymbol{\beta})(\mathbf{V}_{2n}(\boldsymbol{\beta}))^{-1}\mathbf{g}_j(\boldsymbol{\beta})} \leq \frac{1}{M_3}\Big] \geq 1-\epsilon. \tag{2.18}$$

Preuve. *(i)* Pour le multiplicateur de Lagrange $\boldsymbol{\lambda}$, on considère la décomposition suivante :

$$\boldsymbol{\lambda} = \rho\boldsymbol{\phi},$$

tel que $\rho \geq 0$ et $\|\boldsymbol{\phi}\|_1 = 1$.

Le Lemme 2.6.2 implique que, il existe $M_2 > 0$, tel que

$$\frac{1}{1+\frac{\boldsymbol{\lambda}^t}{\theta_{nk}}\mathbf{g}_i(\boldsymbol{\beta})} \geq \frac{1}{1+\frac{\rho}{\theta_{nk}}\|\boldsymbol{\phi}^t\mathbf{g}_i(\boldsymbol{\beta})\|_1} \geq \frac{1}{1+\frac{\rho}{\theta_{nk}}\|\mathbf{g}_i(\boldsymbol{\beta})\|_1} \geq \frac{1}{M_2},$$

avec une probabilité proche de 1, c'est à dire, pour tout $\epsilon > 0$, on a

$$\mathbb{P}\Big[\frac{1}{1+\frac{\rho}{\theta_{nk}}\|\mathbf{g}_i(\boldsymbol{\beta})\|_1} \geq \frac{1}{M_2}\Big] \geq 1-\frac{\epsilon}{2}. \tag{2.19}$$

Par contre, pour la partie droite de la relation (2.17), on suppose qu'il existe une constante positive M_1 telle que

$$\sup_{i\in I, \boldsymbol{\beta}\in\Gamma} \frac{1}{1+\frac{\boldsymbol{\lambda}^t}{\theta_{nk}}\mathbf{g}_i(\boldsymbol{\beta})} \geq \frac{1}{M_1}.$$

Ce qui est équivalant au fait qu'il existe une constante positive M_5, tel que

$$\inf_{i\in I, \boldsymbol{\beta}\in\Gamma} \frac{\boldsymbol{\lambda}^t}{\theta_{nk}}\mathbf{g}_i(\boldsymbol{\beta}) \leq -M_5.$$

Puisque $\boldsymbol{\lambda} = \rho\boldsymbol{\phi}$, $\rho > 0$ et $0 < \theta_{nk} < 1$, alors il existe $M_6 > 0$ tel que

$$\inf_{i\in I, \boldsymbol{\beta}\in\Gamma} \boldsymbol{\phi}^t\mathbf{g}_i(\boldsymbol{\beta}) \leq -M_6. \tag{2.20}$$

D'autre part, on a que

$$\inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \boldsymbol{\phi}^t \mathbf{g}_i(\boldsymbol{\beta}) \geq - \inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \|\mathbf{g}_i(\boldsymbol{\beta})\|_1,$$

avec une probabilité égale à 1.

Tenant compte de la relation (2.20), il existe une constante $M_6 > 0$ tel que

$$- \inf_{i\in I,\boldsymbol{\beta}\in\Gamma} \|\mathbf{g}_i(\boldsymbol{\beta})\|_1 \leq -M_6$$

ou encore

$$\sup_{i\in I,\boldsymbol{\beta}\in\Gamma} \|\mathbf{g}_i(\boldsymbol{\beta})\|_1 \geq M_6,$$

ce qui contredit la relation (2.20) et qui implique la relation (2.17).

(ii) De la même façon que *(i)*, on peut démontrer la relation (2.18) pour $j \in J$. ■

Par le résultat suivant, on montre que $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$ et $\widehat{\boldsymbol{\beta}}(\theta_{nk})$, les solutions du système (2.16), sont des bons estimateurs. Plus précisément, on montre que sous l'hypothèse H_0, $\|\widehat{\boldsymbol{\lambda}}(\theta_{nk})\|$ converge en probabilité vers 0 quand $n \to \infty$ et que $\widehat{\boldsymbol{\beta}}(\theta_{nk})$ est un estimateur consistant de $\boldsymbol{\beta}^0$. On obtient également la vitesse de convergence de ces deux estimateurs, ce qui nous permet de proposer une statistique de test simple à la place de $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$.

Théorème 2.3.2. *Sous les suppositions (A1)-(A4), si l'hypothèse* H_0 *est vraie, alors on a*

$$\widehat{\boldsymbol{\lambda}}(\theta_{nk}) = \min\{\theta_{nk}, 1-\theta_{nk}\} O_{\mathbb{P}}((n \min\{\theta_{nk}, 1-\theta_{nk}\})^{-1/2})$$

et

$$\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^0 = O_{\mathbb{P}}((n \min\{\theta_{nk}, 1-\theta_{nk}\})^{-1/2}).$$

Preuve. La structure de cette preuve est similaire à celle du Lemme A1 de [Liu et al. (2008)] faite pour le cas du modèle linéaire. Par contre, des modifications importantes et des résultats supplémentaires, qui sont dûs au fait que le modèle considéré est non-linéaire, sont nécessaires.

Sans réduire la généralité, on suppose que $\min\{\theta_{nk}, 1-\theta_{nk}\} = \theta_{nk}$. L'autre cas est similaire.

Vu la définition du profil du rapport de vraisemblance empirique $\mathcal{R}'_0(\boldsymbol{\beta})$, on a la contrainte suivante

$$\mathbf{0}_d = \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}). \tag{2.21}$$

Sous l'hypothèse nulle H_0, pour chaque observation $i \in I$, on remplace p_i par sa valeur obtenue dans la relation (2.11). Alors on obtient

$$\sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \frac{1}{\theta_{nk}} \sum_{i\in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})}{1 + \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} = \mathbf{0}_d.$$

ou encore

$$\begin{aligned}
\frac{1}{n\theta_{nk}} \sum_{i\in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})}{1 + \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} &= \frac{1}{n\theta_{nk}} \sum_{i\in I} \frac{1}{1 + \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} \left[\mathbf{g}_i(\boldsymbol{\beta}) + \frac{\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})}{\theta_{nk}} \boldsymbol{\lambda}^t(\boldsymbol{\beta}) - \frac{\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})}{\theta_{nk}} \boldsymbol{\lambda}(\boldsymbol{\beta}) \right] \\
&= \frac{1}{n\theta_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta}) - \frac{1}{n\theta_{nk}^2} \sum_{i\in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})}{1 + \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} \boldsymbol{\lambda}^t(\boldsymbol{\beta}).
\end{aligned} \tag{2.22}$$

On obtient donc

$$\mathbf{0}_d = \frac{1}{n\theta_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta}) - \frac{1}{n\theta_{nk}^2} \sum_{i\in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})}{1 + \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} \boldsymbol{\lambda}(\boldsymbol{\beta}). \tag{2.23}$$

On a noté $\boldsymbol{\lambda}$ par $\boldsymbol{\lambda}(\boldsymbol{\beta})$ pour mettre en évidence que pour chaque valeur de $\boldsymbol{\beta}$, solution de l'équation (2.23), on aura une différente valeur pour $\boldsymbol{\lambda}$.

On considère $\boldsymbol{\beta} = \boldsymbol{\beta}^0 \pm (n\theta_{nk})^{-r} \mathbf{1}_d$, telle que r désigne une constante positive qui sera spécifiée plus tard. Donc, $\|\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathbf{0}}\| = (n\theta_{nk})^{-r} \to 0$, pour $n\theta_{nk} \to \infty$.

Pour le premier terme du membre droit de la relation (2.23), par le Lemme 2.6.3, on a

$$\frac{1}{n\theta_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta}) = O_{\mathbb{P}}((n\theta_{nk})^{-1/2}) + \mathbf{V}_{1n}^0(\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathbf{0}}) + o_{\mathbb{P}}(\boldsymbol{\beta} - \boldsymbol{\beta}^0).$$

On considère maintenant le deuxième terme du membre droit de la relation (2.23). Par la Proposition 2.3.1, on a montré que pour tout $\epsilon > 0$, il existe deux constantes positives M_1 et $M_2 > 0$ tel que

$$\mathbb{P}\Big[\frac{1}{M_1} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta}) \leq \sum_{i\in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})}{1 + \frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} \leq \frac{1}{M_2} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta}) \Big] < \epsilon.$$

Ce qui implique que, afin d'étudier le deuxième terme du membre droit de la relation (2.23), il faut étudier seulement le terme $(n\theta_{nk})^{-1}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta})$.

Par le développement de Taylor de $\mathbf{g}_i(\boldsymbol{\beta})$ dans le voisinage de $\boldsymbol{\beta}^{\mathbf{0}}$, en utilisant un argument similaire à celui utilisé pour le premier terme de la partie droite de la relation (2.23), tenant compte de la supposition (A3), on obtient

$$\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i^t(\boldsymbol{\beta}) = \frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(1+o_{\mathbb{P}}(1)). \tag{2.24}$$

Tenant compte du Lemme 2.6.3 et de la relation (2.24), alors la relation (2.23) devient

$$\mathbf{0}_d = \Big[O_{\mathbb{P}}((n\theta_{nk})^{-1/2}) + \mathbf{V}_{1n}^0(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) - \frac{1}{n\theta_{nk}^2}\sum_{i=1}^{n\theta_{nk}}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_i^t(\boldsymbol{\beta}^{\mathbf{0}})\boldsymbol{\lambda}(\boldsymbol{\beta})\Big](1+o_{\mathbb{P}}(1)). \tag{2.25}$$

On considère la constante r telle que $1/3 \leq r < 1/2$.

Si $\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathbf{0}} + (n\theta_{nk})^{-r}\mathbf{1}_d$, alors $(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{1}_d > 0$ et si $\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathbf{0}} - (n\theta_{nk})^{-r}\mathbf{1}_d$ alors $(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{1}_d < 0$. Donc, la relation (2.25) implique

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{\beta}^{\mathbf{0}} \pm (n\theta_{nk})^{-r}\mathbf{1}_d) &= \pm\Big[\theta_{nk}\Big(\frac{1}{n\theta_{nk}}\sum_{i\in I}\varepsilon_i^2\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})\Big)^{-1}\mathbf{V}_{1n}^0(n\theta_{nk})^{-r}\mathbf{1}_d \\ &\quad + O_{\mathbb{P}}((n\theta_{nk})^{-1/2})\Big](1+o_{\mathbb{P}}(1)).\end{aligned} \tag{2.26}$$

Pour chaque observation $j \in J$, on considère la fonction $\mathbf{v} : \Gamma \to \mathbb{R}^d$ définie par

$$\mathbf{v}(\boldsymbol{\beta}) = \sum_{j\in J} q_j\mathbf{g}_j(\boldsymbol{\beta}) = \frac{1}{n-n\theta_{nk}}\sum_{j\in J}\frac{\mathbf{g}_j(\boldsymbol{\beta})}{1-\frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{1-\theta_{nk}}\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}.$$

Pour $\mathbf{v}(\boldsymbol{\beta})$, on a la décomposition suivante

$$\mathbf{v}(\boldsymbol{\beta}) = \frac{1}{n(1-\theta_{nk})}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}) + \frac{\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})}{n(1-\theta_{nk})^2}\sum_{j\in J}\frac{\mathbf{g}_j(\boldsymbol{\beta})\mathbf{g}_j^t(\boldsymbol{\beta})}{1-\frac{\boldsymbol{\lambda}^t(\boldsymbol{\beta})}{1-\theta_{nk}}\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}\boldsymbol{\lambda}(\boldsymbol{\beta}).$$

Pour faciliter l'écriture, on considère les deux matrices carrés suivantes définies par

$$\mathbf{D}_{1n}^0 = \frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_i^t(\boldsymbol{\beta}^{\mathbf{0}}), \quad \mathbf{D}_{2n}^0 = \frac{1}{n-n\theta_{nk}}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_j^t(\boldsymbol{\beta}^{\mathbf{0}}). \tag{2.27}$$

De la même façon que (2.25), pour les observations $j \in J$, on obtient la relation suivante

$$\mathbf{v}(\boldsymbol{\beta}) = \Big[\mathbf{V}^0_{2n}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + (1-\theta_{nk})^{-1}\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1}\mathbf{D}^0_{2n}\boldsymbol{\lambda}(\boldsymbol{\beta}) + O_{\mathbb{P}}((n(1-\theta_{nk}))^{-1/2})\Big](1+o_{\mathbb{P}}(1)).$$

On remplace $\boldsymbol{\lambda}(\boldsymbol{\beta})$ par sa valeur obtenue en (2.26), alors on obtient

$$\begin{aligned}\mathbf{v}(\boldsymbol{\beta}) \;=\; & \Big[\mathbf{V}^0_{2n}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + (\theta_{nk})(1-\theta_{nk})^{-1}\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1}\mathbf{D}^0_{2n}(\mathbf{D}^0_{1n})^{-1}\mathbf{V}^0_{1n}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) \\ & + O_{\mathbb{P}}((n(1-\theta_{nk}))^{-1/2}) + O_{\mathbb{P}}((n\theta_{nk})^{-1/2})\Big](1+o_{\mathbb{P}}(1)).\end{aligned}$$

Puisque $\boldsymbol{\beta} = \boldsymbol{\beta^0} \pm (n\theta_{nk})^{-r}\mathbf{1}_d$, $1/3 \le r < 1/2$ et $\min\{\theta_{nk}, 1-\theta_{nk}\} = \theta_{nk}$, alors la dernière relation de $v(\boldsymbol{\beta})$ devient

$$\Big[\Big(\mathbf{V}^0_{2n} + \frac{\theta_{nk}}{1-\theta_{nk}}\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1}\mathbf{D}^0_{2n}(\mathbf{D}^0_{1n})^{-1}\mathbf{V}^0_{1n}\Big)(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + O_{\mathbb{P}}((n\theta_{nk})^{-1/2}))\Big]\big(1+o_{\mathbb{P}}(1)\big). \quad (2.28)$$

Ceci implique que $\mathbf{v}(\boldsymbol{\beta^0} + (n\theta_{nk})^{-r}\mathbf{1}_d)$ et $\mathbf{v}(\boldsymbol{\beta^0} - (n\theta_{nk})^{-r}\mathbf{1}_d)$ ont des signes différents, composante par composante. De plus, puisque $\mathbf{v}$ contient des fonctions continues en $\boldsymbol{\beta}$ dans le voisinage de $\boldsymbol{\beta^0}$, alors, d'après le théorème des valeurs intermédiaires, il existe un $\boldsymbol{\beta}$ tel que $\mathbf{v}(\boldsymbol{\beta}) = \mathbf{0}_d$.

D'autre part, puisque $\mathbf{v}(\widehat{\boldsymbol{\beta}}(\theta_{nk})) = \mathbf{0}_d$, on a que

$$\widehat{\boldsymbol{\beta}}(\theta_{nk}) \in [\boldsymbol{\beta^0} - (n\theta_{nk})^{-r}\mathbf{1}_d, \boldsymbol{\beta^0} + (n\theta_{nk})^{-r}\mathbf{1}_d],$$

ce qui implique que

$$\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta^0} = O_{\mathbb{P}}((n\theta_{nk})^{-r}) \ge O_{\mathbb{P}}((n\theta_{nk})^{-1/2}),$$

car $r < 1/2$.

Puisque $\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta^0}$ est le coefficient d'une matrice strictement positive, alors la dernière relation, avec la relation (2.28) impliquent que pour avoir $\mathbf{v}(\widehat{\boldsymbol{\beta}}(\theta_{nk})) = \mathbf{0}_d$, il faut nécessairement que

$$\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta^0} = O_{\mathbb{P}}(n\theta_{nk})^{-1/2}.$$

Tenant compte du dernier résultat, pour la relation (2.26), on obtient que

$$\boldsymbol{\lambda}(\widehat{\boldsymbol{\beta}}(\theta_{nk})) = \theta_{nk}O_{\mathbb{P}}(n\theta_{nk})^{-1/2}.$$

Le théorème est prouvé en totalité. ■

Remarque 2.3.1. *Par la preuve du Lemme 2.6.5, d'après le loi des grands nombres, on a que* $\mathbf{V}_{1n}^0(\mathbf{V}_{2n}^0)^{-1} \xrightarrow[n\to\infty]{p.s.} I_d$*, alors, sous l'hypothèse* H_0*, on peut considérer à la place de la statistique* $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$*, donnée par la relation (2.13), la statistique suivante*

$$T_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = 2\sum_{i\in I} \log(1 + \frac{1}{\theta_{nk}}\boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})) + 2\sum_{j\in J} \log(1 - \frac{1}{1-\theta_{nk}}\boldsymbol{\lambda}^t \mathbf{g}_j(\boldsymbol{\beta})). \tag{2.29}$$

Puisque la fonction de régression est non-linéaire, afin que le maximum de la vraisemblance empirique existe toujours, on considère le paramètre $\theta_{nk} \in [\Theta_{1n}, \Theta_{2n}] \subset (0,1)$, tel que $n\Theta_{1n} \to \infty$, $n(1-\Theta_{2n}) \to \infty$, quand $n \to \infty$. [1].

Finalement, la statistique de test pour tester l'hypothèse H_0 contre son alternative H_1 est

$$\tilde{T_n} \equiv \max_{\theta_{nk}\in[\Theta_{1n},\Theta_{2n}]} T_{nk}(\theta_{nk}, \tilde{\boldsymbol{\lambda}}(\theta_{nk}), \tilde{\boldsymbol{\beta}}(\theta_{nk})), \tag{2.30}$$

où $\tilde{\boldsymbol{\lambda}}(\theta_{nk}), \tilde{\boldsymbol{\beta}}(\theta_{nk})$ sont les solutions du système

$$\begin{cases} \dfrac{\partial T_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})}{2\partial\boldsymbol{\lambda}} = \displaystyle\sum_{i\in I} \frac{\mathbf{g}_i(\boldsymbol{\beta})}{\theta_{nk} + \boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})} - \sum_{j\in J} \frac{\mathbf{g}_j(\boldsymbol{\beta})}{1-\theta_{nk} - \boldsymbol{\lambda}^t \mathbf{g}_j(\boldsymbol{\beta})} = \mathbf{0}_d, \\ \dfrac{\partial T_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})}{2\partial\boldsymbol{\beta}} = \displaystyle\sum_{i\in I} \frac{\dot{\mathbf{g}}_i(\boldsymbol{\beta})\boldsymbol{\lambda}}{\theta_{nk} + \boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})} - \sum_{j\in J} \frac{\dot{\mathbf{g}}_j(\boldsymbol{\beta})\boldsymbol{\lambda}}{1-\theta_{nk} - \boldsymbol{\lambda}^t \mathbf{g}_j(\boldsymbol{\beta})} = \mathbf{0}_d. \end{cases} \tag{2.31}$$

Alors, on peut considérer comme un estimateur du point de rupture k^0, l'estimateur du maximum de la vraisemblance empirique

$$\tilde{k}_n \equiv n\tilde{\theta}_n \equiv n\min\{\tilde{\theta}_{nk}; \tilde{\theta}_{nk} = \underset{\theta_{nk}\in[\Theta_{1n},\Theta_{2n}]}{\arg\max}\ T_{nk}(\theta_{nk}, \tilde{\boldsymbol{\lambda}}(\theta_{nk}), \tilde{\boldsymbol{\beta}}(\theta_{nk}))\}.$$

Le résultat suivant donne la distribution asymptotique de la statistique de test $\tilde{T_n}$ donnée par la relation (2.30), sous l'hypothèse nulle de non-changement H_0.

Pour ceci, on considère les fonctions suivantes :

$$A(x) \equiv (2\log x)^{1/2},$$

$$D(x) = 2\log x + \log\log x$$

1. Le lecteur peut trouver une discussion concernant les valeurs possibles de Θ_{1n}, Θ_{2n} dans les papiers de [Zou et al. (2007)] et de [Liu et al. (2008)]

et

$$u(n) = \frac{1 - \Theta_{1n}\Theta_{2n}}{\Theta_{1n}(1 - \Theta_{2n})} \to \infty, \text{ quand } \quad n \to \infty.$$

Théorème 2.3.3. *Sous les suppositions (A1)-(A4), si l'hypothèse* H_0 *est vraie, alors pour tout* $t \in \mathbb{R}$*, on a*

$$\lim_{n\to\infty} \mathbb{P}\{A(\log u(n))(\tilde{T}_n)^{\frac{1}{2}} \leq t + D(\log u(n))\} = exp(-e^{-t}). \tag{2.32}$$

Preuve. La structure de cette preuve est similaire à celle du Théorème 1.3.1 de [Csörgö and Horváth (1997)]. Le Lemme 2.6.5 est utilisé à la place du Théorème 1.1.1 de [Csörgö and Horváth (1997)] et le Théorème A.3.4 de [Csörgö and Horváth (1997)] est utilisé à la place du Corollaire A.3.1 de [Csörgö and Horváth (1997)]. ■

Corollaire 2.3.1. *Conséquence du Théorème 2.3.3 , pour une valeur fixée de* $\alpha \in (0, 1)$*, on peut déduire la région critique suivante*

$$(\tilde{T}_n)^{1/2} \geq \frac{-\log(-\log\alpha) + D(\log u(n))}{A(\log u(n))}.$$

La non-linéarité en $\boldsymbol{\beta}$ du modèle, produisent des difficultés numériques pour résoudre le système donné par la relation (2.31). Ce qui rend l'utilisation du Théorème 2.3.3 dans les applications assez compliqués. Alors, on propose une forme plus simple pour la statistique de test, pour l'utiliser aisément en pratique. Cette forme doit préserver tous les propriétés théoriques de la statistique de test donnée par la relation (2.30).

Remarque 2.3.2. *Tenant compte de la dernière relation du Lemme 2.6.5, le Théorème 2.3.3 implique que, dans la pratique, afin de tester l'hypothèse* H_0 *contre* H_1*, on utilise la forme approximée de la statistique de test :*

$$T(\theta_{nk}) = n\sigma^{-2}\theta_{nk}(1 - \theta_{nk})(\boldsymbol{W}^0_{1n} - \boldsymbol{W}^0_{2n})^t\mathbf{V}^{-1}(\boldsymbol{W}^0_{1n} - \boldsymbol{W}^0_{2n}), \tag{2.33}$$

où

$$\boldsymbol{W}^0_{1n} = \frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}), \qquad \boldsymbol{W}^0_{2n} = \frac{1}{n(1-\theta_{nk})}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}}). \tag{2.34}$$

Dans les applications, puisque le paramètre $\boldsymbol{\beta}^{\mathbf{0}}$ *est inconnu, on le remplace par un estimateur consistent, par exemple, l'estimateur du moindre carré ordinaire noté par* $\widehat{\boldsymbol{\beta}}_{MC}$*.*
Sous H_0*, la variance du l'erreur* σ^2 *est estimé par* $n^{-1}\sum_{i=1}^{n}[\boldsymbol{Y}_i - f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_{MC})]^2$ *et la matrice*

$\mathbf{V}$ *par* $n^{-1}\sum_{i=1}^{n}\dot{\mathbf{f}}(\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_{MC})\dot{\mathbf{f}}^{t}(\mathbf{X}_i,\widehat{\boldsymbol{\beta}}_{MC})$.

L'estimateur du maximum de la vraisemblance empirique modifié pour le point de rupture k^0 *est*

$$\widehat{k}_n = n\widehat{\theta}_n = n\min\{\widehat{\theta}_{nk};\widehat{\theta}_{nk} = \underset{\Theta_{1n}\leq\theta_{nk}\leq\Theta_{2n}}{\arg\max}\ T(\theta_{nk})\}. \tag{2.35}$$

Comportement asymptotique de T_{nk} et $\tilde{T}_{nk}$ sous H_1

Considérons maintenant que l'hypothèse H_1 est vraie. Tout d'abord, on prouve que le maximum de la statistique T_{nk} converge vers le maximum de sa distribution limite. Alors, on montre que la statistique de test $\tilde{T}_n$ est consistante (a la puissance asymptotique égale à 1).

Si k^0 est le vrai temps du changement, on note par $\theta_{n0} = k^0/n$ et on suppose que $\theta_0 \equiv \lim_{n\to\infty}\theta_{n0}$, où $\theta_0 \in (0,1)$.

Pour $\mathbf{x} \in \Upsilon$ et $e \in \mathbb{R}$, soient $F_{\mathbf{x}}(e)$ la distribution conditionnelle de $\mathbf{g}(\mathbf{X}_i,\boldsymbol{\beta})$ quand $\mathbf{X}_i = \mathbf{x}$, pour les observations $i \in I$ et $G_{\mathbf{x}}(e)$ la distribution conditionnelle de $\mathbf{g}(\mathbf{X}_j,\boldsymbol{\beta})$ quand $\mathbf{X}_j = \mathbf{x}$, pour les observations $j \in J$.

Nous rappelons que la fonction de répartition de $\mathbf{X}$ est $H(\mathbf{x})$. Pour $\mathbf{x} \in \Upsilon$ et une valeur fixée de θ, on définit

$$dP_{\mathbf{x}}(e) \equiv (\theta\mathbb{1}_{\{\theta\leq\theta_0\}} + \theta_0\mathbb{1}_{\{\theta>\theta_0\}})dF_{\mathbf{x}}(e) + (\theta-\theta_0)\mathbb{1}_{\{\theta>\theta_0\}}dG_{\mathbf{x}}(e),$$

$$dQ_{\mathbf{x}}(e) \equiv ((1-\theta)\mathbb{1}_{\{\theta\geq\theta_0\}} + (1-\theta_0)\mathbb{1}_{\{\theta<\theta_0\}})dG_{\mathbf{x}}(e) + (\theta_0-\theta)\mathbb{1}_{\{\theta<\theta_0\}}dF_{\mathbf{x}}(e),$$

$$dR_{\mathbf{x}}(e) \equiv \mathbb{1}_{\{\theta<\theta_0\}}dF_{\mathbf{x}}(e) + \mathbb{1}_{\{\theta>\theta_0\}}dG_{\mathbf{x}}(e).$$

Comme on a prouvé que sous l'hypothèse H_0 on peut considérer à la place de la statistique de vraisemblance empirique (2.13) la statistique (2.29), alors, on définit la statistique suivante

$$\Lambda_{nk}(\theta_{nk}) = \frac{T_{nk}(\theta_{nk},\tilde{\boldsymbol{\lambda}}(\theta_{nk}),\tilde{\boldsymbol{\beta}}(\theta_{nk})}{2n}, \Lambda_n(0) = \Lambda_n(1) = 0, \tag{2.36}$$

avec T_{nk} donné par la relation (2.29).

Par une preuve similaire à celle du Théorème 2.3.2, on a que

$$\begin{aligned}\tilde{\boldsymbol{\lambda}}(\theta_{nk}) &= min\{\theta_{nk}, 1-\theta_{nk}\}O_{\mathbb{P}}((n\, min\{\theta_{nk}, 1-\theta_{nk}\})^{-1/2}),\\ \tilde{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^0 &= O_{\mathbb{P}}((n\, min\{\theta_{nk}, 1-\theta_{nk}\})^{-1/2}).\end{aligned} \tag{2.37}$$

Pour un $\boldsymbol{\lambda}$ et $\boldsymbol{\beta}$ quelconque, soit la fonction $K : \Upsilon \times \mathbb{R} \times (0,1)$ définie par

$$K(\mathbf{x}, e, \theta) = \theta + \boldsymbol{\lambda}^t \dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta})[e - f(\mathbf{x}, \boldsymbol{\beta}) + f(\mathbf{x}, \boldsymbol{\beta}^{\mathbf{0}})].$$

Soit également

$$\begin{aligned}\psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) &= \int_{\Upsilon} \Big(\int_{\mathbb{R}} \log K(\mathbf{x}, e, \theta) dP_{\mathbf{x}}(e) + \int_{\mathbb{R}} \log(1 - K(\mathbf{x}, e, \theta)) dQ_{\mathbf{x}}(e) \Big) dH(\mathbf{x})\\ &- \theta \log \theta - (1-\theta)\log(1-\theta).\end{aligned} \tag{2.38}$$

Par le Théorème 2.3.4, on prouve que sous l'hypothèse H_1, ψ est le processus limite de Λ_{nk}.

Alors, pour une valeur fixée de θ, telle que $\theta \in (0,1)$, on considère par $\tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta)$ la solution du système suivant

$$\begin{cases} \mathbf{z}_1(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \displaystyle\int_{\Upsilon} \Big(\int_{\mathbb{R}} \frac{\mathbf{g}(\mathbf{x}, \boldsymbol{\beta})}{K(\mathbf{x}, e, \theta)} dP_{\mathbf{x}}(e) - \int_{\mathbb{R}} \frac{\mathbf{g}(\mathbf{x}, \boldsymbol{\beta})}{1 - K(\mathbf{x}, e, \theta)} dQ_{\mathbf{x}}(e) \Big) dH(\mathbf{x}) = \mathbf{0}_d, \\ \mathbf{z}_2(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \displaystyle\int_{\Upsilon} \Big(\int_{\mathbb{R}} \frac{\dot{\mathbf{g}}(\mathbf{x}, \boldsymbol{\beta})\boldsymbol{\lambda}}{K(\mathbf{x}, e, \theta)} dP_{\mathbf{x}}(e) - \int_{\mathbb{R}} \frac{\dot{\mathbf{g}}(\mathbf{x}, \boldsymbol{\beta})\boldsymbol{\lambda}}{1 - K(\mathbf{x}, e, \theta)} dQ_{\mathbf{x}}(e) \Big) dH(\mathbf{x}) = \mathbf{0}_d, \end{cases} \tag{2.39}$$

où,

$$\mathbf{z}_1(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \frac{\partial \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\lambda}}, \qquad \mathbf{z}_2(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \frac{\partial \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}.$$

Pour la suite, on a besoin des suppositions suivantes :

(A5) La matrice $\begin{pmatrix} -\dfrac{\partial \mathbf{z}_1(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\lambda}} & -\dfrac{\partial \mathbf{z}_1(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ -\dfrac{\partial \mathbf{z}_2(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\lambda}} & -\dfrac{\partial \mathbf{z}_2(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \beta} \end{pmatrix}$ est inversible pour tout $\theta \in (0,1)$.

(A6) On a :

$$\int_{\Upsilon} \int_{\mathbb{R}} \Big(\frac{\mathbf{g}(\mathbf{x}, \boldsymbol{\beta}) \mathbf{g}^t(\mathbf{x}, \boldsymbol{\beta})}{K^2(\mathbf{x}, e, \theta)} + \frac{\mathbf{g}(\mathbf{x}, \boldsymbol{\beta}) \mathbf{g}^t(\mathbf{x}, \boldsymbol{\beta})}{(1 - K(\mathbf{x}, e, \theta))^2} \Big) d(F_{\mathbf{x}}(e) + G_{\mathbf{x}}(e)) < \infty,$$

$$\int_{\Upsilon}\int_{\mathbb{R}}(\frac{\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})}{K^2(\mathbf{x},e,\theta)}+\frac{\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})\dot{\mathbf{g}}(\mathbf{x},\boldsymbol{\beta})}{(1-K(\mathbf{x},e,\theta))^2})d(F_{\mathbf{x}}(e)+G_{\mathbf{x}}(e))<\infty,$$

avec la convention que ces deux intégrales sont appliquées composante par composante à la matrice correspondante.

Soit $\mathbf{g}_l(\mathbf{x},\boldsymbol{\beta})$ la l-ème composante du vecteur $\mathbf{g}(\mathbf{x},\boldsymbol{\beta})$. Pour tout $1\leq j,k\leq d$, on suppose que

$$\int_{\Upsilon}\int_{\mathbb{R}}(\frac{1}{K^2(\mathbf{x},e,\theta)}\frac{\partial^2\mathbf{g}_l(\mathbf{x},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k}+\frac{1}{(1-K(\mathbf{x},e,\theta))^2}\frac{\partial^2\mathbf{g}_l(\mathbf{x},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k})d(F_{\mathbf{x}}(e)+G_{\mathbf{x}}(e))<\infty.$$

(A7) Les fonctions $f(\mathbf{x},\boldsymbol{\beta})$ et $\dot{\mathbf{f}}(\mathbf{x},\boldsymbol{\beta})$ sont equicontinues en $\boldsymbol{\beta}$ dans Γ.

Remarque 2.3.3. *Une condition suffisante pour que les fonctions $f(\mathbf{x},\boldsymbol{\beta})$ et $\dot{\mathbf{f}}(\mathbf{x},\boldsymbol{\beta})$ soient équicontinues est qu'elles soient Lipschitziennes par rapport à $\boldsymbol{\beta}$ dans Γ.*

Par le théorème suivant, on montre que si θ_{nk} converge vers la vraie valeur θ_0, alors, le maximum de la statistique de test modifiée converge vers le maximum de sa distribution limite.

Théorème 2.3.4. *Sous l'hypothèse alternative* H_1*, si les suppositions (A5)-(A7) sont vérifiées, $\theta_0\in(0,1)$ et $\lim_{n\to\infty}\theta_{nk}=\theta\in[0,1]$, alors*

$$\Lambda_{nk}(\theta_{nk})\xrightarrow[n\to\infty]{p.s.}\psi(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta)).$$

De plus, $\psi(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta))$ est une fonction strictement croissante en $(0,\theta_0)$ et décroissante en $(\theta_0,1)$ et

$$\max_{0\leq\theta\leq1}\psi(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta))=\psi(\theta_0,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta_0),\tilde{\tilde{\boldsymbol{\beta}}}(\theta_0)).$$

Preuve. Ce théorème sera prouvé en deux étapes.

Étape 1. Tout d'abord, on montre que, pour tout θ fixé, telle que $\theta\in(0,1)$, on a que

$$\arg\max_{(\boldsymbol{\lambda},\boldsymbol{\beta})}T_{nk}(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})\xrightarrow[n\to\infty]{p.s.}\arg\max_{(\boldsymbol{\lambda},\boldsymbol{\beta})}\psi(\theta,\boldsymbol{\lambda},\boldsymbol{\beta}).\tag{2.40}$$

Évidement, par la loi des grands nombres, pour tout $(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})\in(0,1)\times\mathbb{R}\times\Gamma$, on a que

$$\frac{1}{2n}T_{nk}(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})\xrightarrow[n\to\infty]{p.s.}\psi(\theta,\boldsymbol{\lambda},\boldsymbol{\beta}).$$

D'autre part, par la supposition (A5), $\arg\max_{(\boldsymbol{\lambda},\boldsymbol{\beta})} \psi(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})$ est la solution unique du système (2.39). Vu les suppositions (A6) et (A7), la fonction $(2n)^{-1}T_{nk}(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})$ est équicontinue et bornée en $\boldsymbol{\lambda}$ et $\boldsymbol{\beta}$. Alors, en utilisant le Théorème 1.12.1 de [Van der Vaart and Wellner (1996)] , la convergence de $(2n)^{-1}T_{nk}(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})$ vers $\psi(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})$ est uniforme en $(\boldsymbol{\lambda},\boldsymbol{\beta})$.
Tenant compte du fait que la solution du système (2.31) est unique, alors on obtient la relation (2.40).

Étape 2. On montre que

$$\max_{\theta_{nk}} \Lambda_{nk}(\theta_{nk}) \xrightarrow[n\to\infty]{p.s.} \max_{\theta} \psi(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta)), \tag{2.41}$$

avec $\tilde{\tilde{\boldsymbol{\lambda}}}(\theta)$ et $\tilde{\tilde{\boldsymbol{\beta}}}(\theta)$ les solutions du système (2.39). Pour $\theta = \lim_{n\to\infty}\theta_{nk}$, on a

$$\begin{aligned}
\Lambda_{nk}(\theta_{nk}) &= \frac{1}{2n}\Big[2\sum_{i\in I}\log\big[1+\frac{1}{\theta}\tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta)\mathbf{g}_i(\tilde{\tilde{\boldsymbol{\beta}}}(\theta))\big]+2\sum_{j\in J}\log\big[1-\frac{1}{1-\theta}\tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta)\mathbf{g}_j(\boldsymbol{\beta})\tilde{\tilde{\boldsymbol{\beta}}}(\theta)\big]\Big] \\
&= \frac{1}{n}\sum_{i\in I}\log\big[\theta+\tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta)\mathbf{g}_i(\tilde{\tilde{\boldsymbol{\beta}}}(\theta))\big]+\frac{1}{n}\sum_{j\in J}\log\big[(1-\theta)-\tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta)\mathbf{g}_j(\tilde{\tilde{\boldsymbol{\beta}}}(\theta))\big] \\
&\quad -\theta\log\theta-(1-\theta)\log(1-\theta)+o_{\mathbb{P}}(1).
\end{aligned}$$

La relation ci-dessus avec la loi des grands nombres, impliquent que

$$\Lambda_{nk}(\theta_{nk}) \xrightarrow[n\to\infty]{p.s.} \psi(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta)), \tag{2.42}$$

avec $\theta = \lim\limits_{n\to\infty}\theta_{nk}$.

Pour $\theta \notin \{0,1,\theta_0\}$, la dérivée partielle devient

$$\begin{aligned}
\frac{\partial\psi(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial\theta} &= \int_{\Upsilon}\int_{\mathbb{R}}\Big[[\log K(\mathbf{x},e,\theta)\mathbb{1}_{\{\theta<\theta_0\}}dF_{\mathbf{x}}(e)+\log K(\mathbf{x},e,\theta)\mathbb{1}_{\{\theta>\theta_0\}}dG_{\mathbf{x}}(e)] \\
&\quad -[\log(1-K(\mathbf{x},e,\theta))\mathbb{1}_{\{\theta<\theta_0\}}dF_{\mathbf{x}}(e)+\log(1-K(\mathbf{x},e,\theta))\mathbb{1}_{\{\theta>\theta_0\}}dG_{\mathbf{x}}(e)]\Big]dH(\mathbf{x}) \\
&\quad +\log(1-\theta)-\log\theta.
\end{aligned}$$

D'autre part, on a que

$$dR_{\mathbf{x}}(e) = \mathbb{1}_{\{\theta<\theta_0\}}dF_{\mathbf{x}}(e)+\mathbb{1}_{\{\theta>\theta_0\}}dG_{\mathbf{x}}(e).$$

D'où,

$$\frac{\partial\psi(\theta,\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial\theta} = \int_{\Upsilon}\int_{\mathbb{R}}[\log K(\mathbf{x},e,\theta)-\log(1-K(\mathbf{x},e,\theta))]dR_{\mathbf{x}}(e)dH(\mathbf{x}) - \log\theta+\log(1-\theta).$$

Comme $\tilde{\tilde{\boldsymbol{\lambda}}}^t(\theta)\mathbf{g}(\mathbf{x},\tilde{\tilde{\boldsymbol{\beta}}}(\theta)) = K(\mathbf{x},e,\theta)-\theta$ et $\mathbf{z}_1(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta)) = \mathbf{0}_d$, alors, on obtient

$$\int_{\Upsilon}\Big[\int_{\mathbb{R}}(1-\frac{\theta}{K(\mathbf{x},e,\theta)})dP_{\mathbf{x}}(e)+\int_{\mathbb{R}}(1-\frac{1-\theta}{1-K(\mathbf{x},e,\theta)})dQ_{\mathbf{x}}(e)\Big]dH(\mathbf{x})=0. \tag{2.43}$$

D'autre part, on a

$$\mathbf{z}_2(\theta,\tilde{\tilde{\boldsymbol{\lambda}}}(\theta),\tilde{\tilde{\boldsymbol{\beta}}}(\theta)) = \mathbf{0}_d.$$

Alors

$$\int_{\Upsilon}\Big[\dot{\mathbf{g}}(\mathbf{x},\tilde{\tilde{\boldsymbol{\beta}}}(\theta))\int_{\mathbb{R}}\frac{dP_{\mathbf{x}}(e)}{K(\mathbf{x},e,\theta)}\Big]dH(\mathbf{x}) = \int_{\Upsilon}\Big[\dot{\mathbf{g}}(\mathbf{x},\tilde{\tilde{\boldsymbol{\beta}}}(\theta))\int_{\mathbb{R}}\frac{dQ_{\mathbf{x}}(e)}{1-K(\mathbf{x},e,\theta)}\Big]dH(\mathbf{x})=\mathbf{0}_d.$$

Puisque

$$\int_{\Upsilon}\Big[\int_{\mathbb{R}}dP_{\mathbf{x}}(e)+\int_{\mathbb{R}}dQ_{\mathbf{x}}(e)\Big]dH(\mathbf{x})=1,$$

alors, la relation (2.43) devient

$$1-\theta\int_{\Upsilon}\int_{\mathbb{R}}\Big[\frac{dP_{\mathbf{x}}(e)}{K(\mathbf{x},e,\theta)}-\frac{dQ_{\mathbf{x}}(e)}{1-K(\mathbf{x},e,\theta)}\Big]dH(\mathbf{x})-\int_{\Upsilon}\int_{\mathbb{R}}\frac{dQ_{\mathbf{x}}(e)}{1-K(\mathbf{x},e,\theta)}dH(\mathbf{x})=0. \tag{2.44}$$

La relation (2.44) est valable pour tout $\theta\in(0,1)$. Si $\theta=0$, la relation (2.44) implique

$$\int_{\Upsilon}\int_{\mathbb{R}}\frac{dQ_{\mathbf{x}}(e)}{1-K(\mathbf{x},e,\theta)}dH(\mathbf{x})=1. \tag{2.45}$$

Si $\theta=1$, la relation (2.44) implique

$$\int_{\Upsilon}\int_{\mathbb{R}}\frac{dP_{\mathbf{x}}(e)}{K(\mathbf{x},e,\theta)}dH(\mathbf{x})=1. \tag{2.46}$$

Pour $\theta<\theta_0$, la relation (2.46) sera

$$\int_{\Upsilon}\int_{\mathbb{R}}\frac{1}{K(\mathbf{x},e,\theta)}\Big[\theta\mathbb{1}_{\{\theta<\theta_0\}}dF_{\mathbf{x}}(e)\Big]dH(\mathbf{x})=1.$$

Puisque $\mathbb{1}_{\{\theta<\theta_0\}}$ désigne la fonction indicatrice de l'ensemble $\theta < \theta_0$, alors

$$\int_\Upsilon \int_{\mathbb{R}} \frac{\theta}{K(\mathbf{x}, e, \theta)} dF_{\mathbf{x}}(e) dH(\mathbf{x}) = 1,$$

ce qui donne

$$\int_\Upsilon \int_{\mathbb{R}} \frac{1}{K(\mathbf{x}, e, \theta)} dF_{\mathbf{x}}(e) dH(\mathbf{x}) = \frac{1}{\theta}. \tag{2.47}$$

La dérivée partielle devient :

$$\begin{aligned}
\frac{\partial \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \theta} &= \int_\Upsilon \int_{\mathbb{R}} \Big[\log K(\mathbf{x}, e, \theta) \mathbb{1}_{\{\theta<\theta_0\}} - \log(1 - K(\mathbf{x}, e, \theta)) \mathbb{1}_{\{\theta<\theta_0\}} \Big] dF_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&\quad + \log(1-\theta) - \log\theta \\
&= \int_\Upsilon \int_{\mathbb{R}} \log \frac{K(\mathbf{x}, e, \theta)}{1 - K(\mathbf{x}, e, \theta)} dF_{\mathbf{x}}(e) dH(\mathbf{x}) - \int_\Upsilon \int_{\mathbb{R}} \log \frac{\theta}{1-\theta} dF_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&= \int_\Upsilon \int_{\mathbb{R}} \log \frac{(1-\theta) K(\mathbf{x}, e, \theta)}{\theta(1 - K(\mathbf{x}, e, \theta))} dF_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&= \int_\Upsilon \int_{\mathbb{R}} - \log \frac{\theta(1 - K(\mathbf{x}, e, \theta))}{(1-\theta) K(\mathbf{x}, e, \theta)} dF_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&> - \log \int_\Upsilon \int_{\mathbb{R}} \frac{\theta(1 - K(\mathbf{x}, e, \theta))}{(1-\theta) K(\mathbf{x}, e, \theta)} dF_{\mathbf{x}}(e) dH(\mathbf{x}).
\end{aligned}$$

En utilisant la relation (2.47), on obtient que

$$- \log \int_\Upsilon \int_{\mathbb{R}} \frac{\theta(1 - K(\mathbf{x}, e, \theta))}{-K(\mathbf{x}, e, \theta(1-\theta))} dF_{\mathbf{x}}(e) dH(\mathbf{x}) = 0.$$

Alors

$$\frac{\partial \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \theta} > 0.$$

On étudie maintenant le cas où $\theta > \theta_0$. En utilisant la relation (2.45), on obtient que

$$\int_\Upsilon \int_{\mathbb{R}} \frac{1}{1 - K(\mathbf{x}, e, \theta)} dG_{\mathbf{x}}(e) dH(\mathbf{x}) = \frac{1}{1-\theta}. \tag{2.48}$$

De la même façon que pour $\theta < \theta_0$, en utilisant la relation (2.48), pour $\theta > \theta_0$, on obtient que

$$\frac{\partial \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \theta} < 0.$$

Pour $\theta = \theta_0$, les relations (2.47) et (2.48) deviennent respectivement

$$\int_{\Upsilon}\int_{\mathbb{R}} \frac{1}{K(\mathbf{x}, e, \theta_0)} dF_{\mathbf{x}}(e) dH(\mathbf{x}) = \frac{1}{\theta_0} \tag{2.49}$$

et

$$\int_{\Upsilon}\int_{\mathbb{R}} \frac{1-\theta}{1-K(\mathbf{x}, e, \theta_0)} dG_{\mathbf{x}}(e) dH(\mathbf{x}) = \frac{1}{1-\theta_0}. \tag{2.50}$$

Alors, on obtient

$$\begin{aligned}
\psi(\theta_0, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta_0), \tilde{\tilde{\boldsymbol{\beta}}}(\theta_0)) &= \int_{\Upsilon}\int_{\mathbb{R}} \theta_0 \log K(\mathbf{x}, e, \theta_0) \mathbb{1}_{\{\theta<\theta_0\}} dF_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&\quad + \int_{\Upsilon}\int_{\mathbb{R}} (1-\theta_0) \log(1-K(\mathbf{x}, e, \theta_0)) \mathbb{1}_{\{\theta<\theta_0\}} dG_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&\quad -\theta_0 \log \theta_0 + (1-\theta_0)\log(1-\theta_0) \\
&> -\theta_0 \log \int_{\Upsilon}\int_{\mathbb{R}} \frac{1}{K(\mathbf{x}, e, \theta_0)\mathbb{1}_{\{\theta<\theta_0\}}} dF_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&\quad -(1-\theta_0)\log \int_{\Upsilon}\int_{\mathbb{R}} \frac{1}{(1-K(\mathbf{x}, e, \theta_0))} \mathbb{1}_{\{\theta<\theta_0\}} dG_{\mathbf{x}}(e) dH(\mathbf{x}) \\
&\quad -\theta_0 \log \theta_0 + (1-\theta_0)\log(1-\theta_0) \\
&= 0.
\end{aligned}$$

Donc, on peut constater que la fonction $\psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta})$ est strictement croissante sur $(0, \theta_0)$ et décroissante sur $(\theta_0, 1)$ et

$$\max_{0\le\theta\le 1} \psi(\theta, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \psi(\theta_0, \boldsymbol{\lambda}, \boldsymbol{\beta}).$$

Sous l'hypothèse H_1, en utilisant la relation (2.42), on obtient que

$$\Lambda_{nk}(\theta_{nk}) \xrightarrow[n\to\infty]{p.s.} \psi(\theta_0, \boldsymbol{\lambda}, \boldsymbol{\beta})$$

ainsi que, pour tout $\gamma \in (0, \min(\theta_0, 1-\theta_0))$, on a

$$\max_{|k-n\theta_0|\ge n\gamma} \Lambda_{nk}(\theta_{nk}) \xrightarrow[n\to\infty]{p.s.} \max_{|\theta-\theta_0|\ge\gamma} \psi(\theta, \tilde{\tilde{\boldsymbol{\lambda}}}(\theta), \tilde{\tilde{\boldsymbol{\beta}}}(\theta)). \tag{2.51}$$

Ce qui implique que

$$\lim_{n\to\infty} \mathbb{P}[|\arg\max_k \Lambda_{nk}(\theta_{nk}) - \theta_0| \ge \gamma] = 0.$$

■

Corollaire 2.3.2. *La preuve du Théorème 2.3.4 implique que l'estimateur de θ_0 défini par*

$$\tilde{\theta}_n \equiv \min\{\tilde{\theta}_{nk}; \tilde{\theta}_{nk} = \underset{\theta_{nk}\in[\Theta_{1n},\Theta_{2n}]}{\arg\max}\ T_{nk}(\theta_{nk}, \tilde{\boldsymbol{\lambda}}(\theta_{nk}), \tilde{\boldsymbol{\beta}}(\theta_{nk}))\}$$

vérifie la propriété que

$$\tilde{\theta}_n - \theta_{n0} \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Tenant compte de la Remarque 2.3.2, on a également que $\widehat{\theta}_n - \theta_{n0}$ converge vers 0 en probabilité, avec $\widehat{\theta}_n$ l'estimateur de θ_0 défini dans la Remarque 2.3.2.

Par le théorème suivant, on prouve que la statistique de test $\tilde{T}_n$ donnée par la relation (2.30) a la puissance asymptotique égale à 1.

Théorème 2.3.5. *Sous les suppositions (A1)-(A7), la puissance du test du rapport du vraisemblance empirique $\tilde{T}_n$ converge vers 1.*

Preuve. Par la preuve du Théorème 2.3.4, sous l'hypothèse H_1, on a

$$\max_{\theta_{nk}\in[\Theta_{1n},\Theta_{2n}]} \frac{T_{nk}(\theta_{nk}, \tilde{\boldsymbol{\lambda}}(\theta_{nk}), \tilde{\boldsymbol{\beta}}(\theta_{nk}))}{2n} \xrightarrow[n\to\infty]{p.s.} 0.$$

Tenant compte de la relation (2.37), on a

$$\frac{\tilde{T}_n}{2n} \xrightarrow[n\to\infty]{p.s.} 0. \tag{2.52}$$

D'autre part, si on suppose que l'hypothèse H_0 est vraie, pour tout $t \in \mathbb{R}$, par le Théorème 2.3.3, on a

$$\mathbb{P}\Big[A(\log u(n))\tilde{T}_n^{\frac{1}{2}} \leq t + D(\log u(n))\Big] = exp(-e^{-t}).$$

En considérant dans la dernière relation $t = -\log\log u(n)$, on obtient

$$\lim_{n\to\infty} \mathbb{P}\Big[A(\log u(n))(\frac{\tilde{T}_n}{n})^{\frac{1}{2}} \leq -\frac{\log\log u(n)}{\sqrt{n}} + \frac{D(\log u(n))}{\sqrt{n}}\Big] = 0. \tag{2.53}$$

Ce qui termine la démonstration. ■

Notons que, des résultats similaires aux Théorèmes 2.3.3, 2.3.4 et 2.3.5 sont obtenus par d'autres auteurs pour des modèles plus simples. Le lecteur peut trouver les résultats correspondants pour le test de détection d'un changement dans la suite de variables aléatoires dans le papier de [Zou et al. (2007)] et pour détecter le changement dans les paramètres d'un modèle linéaire dans le papier de [Liu et al. (2008)].

2.4 Test d'un modèle épidémique non-linéaire

Dans cette section, on considère le modèle épidémique. [Ning and al. (2012)] ont testé un modèle épidémique linéaire par la vraisemblance empirique. Pour la moyenne d'une suite de variables aléatoires normales indépendantes, [Yao (1993)] a testé un modèle épidémique par des différents tests statistiques tels que : rapport de vraisemblance, rapport de sémi-vraisemblance. [Ramanayake and Gupta (2003)] et [Ramanayake (2004)] ont étudié le changement épidémique dans la moyenne d'une suite des variables aléatoires exponentielles et dans la forme des paramètres d'une suite des variables aléatoires qui suivent une loi gamma, en utilisant la méthode du rapport de vraisemblance.

Sous l'hypothèse alternative, notée par H_2, on suppose que le modèle a deux points de rupture k_1 et k_2 $(1 < k_1 < k_2 < n)$, tel que le modèle est le même sur le premier et le troisième segment. Alors, le modèle peut s'écrire

$$\mathrm{H}_2 : \mathbf{Y}_i = \begin{cases} f(\mathbf{X}_i, \boldsymbol{\beta}_1) + \varepsilon_i & i = 1, \ldots, k_1 \\ f(\mathbf{X}_i, \boldsymbol{\beta}_2) + \varepsilon_i & i = k_1 + 1, \ldots, k_2 \\ f(\mathbf{X}_i, \boldsymbol{\beta}_1) + \varepsilon_i & i = k_2 + 1, \ldots, n. \end{cases} \tag{2.54}$$

On teste l'hypothèse nulle H_0 de non-changement donnée par la relation (2.1), contre l'hypothèse alternative H_2 donnée par (2.54).

Sous l'hypothèse H_2, on considère les deux ensembles de nombres naturels, $I' = \{1, \ldots, k_1, k_2 + 1, \ldots, n\}$ et $J' = \{k_1 + 1, \ldots, k_2\}$. On définit les vecteurs de probabilités correspondants $(u_1, \ldots, u_{k_1}, u_{k_2+1}, \ldots, u_n)$ et $(v_{k_1+1}, \ldots, v_{k_2})$. Évidement, ces probabilités satisfont les relations

$$\sum_{i \in I'} u_i = 1, \qquad \sum_{j \in J'} v_j = 1.$$

Sous l'hypothèse H_0, le profil du rapport de vraisemblance empirique pour $\boldsymbol{\beta}$ est

$$U'_{0,n,k_1,k_2}(\boldsymbol{\beta}) = \sup_{(u_1,\ldots,u_{k_1},u_{k_2+1},\ldots,u_n)} \sup_{(v_{k_1+1},\ldots,v_{k_2})} \Big\{ \prod_{i \in I'} (n - k_2 + k_1) u_i \prod_{j \in J'} (k_2 - k_1) v_j ; \sum_{i \in I'} u_i = \sum_{j \in J'} v_j = 1, \sum_{i \in I'} u_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j \in J'} v_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\}.$$

Sous l'hypothèse H_2, le profil du rapport de vraisemblance empirique pour $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ a la forme

$$U'_{1,n,k_1,k_2}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = \sup_{(u_1,\ldots,u_{k_1},u_{k_2+1},\ldots,u_n)} \sup_{(v_{k_1+1},\ldots,v_{k_2})} \Big\{ \prod_{i\in I'}(n-k_2+k_1)u_i \prod_{j\in J'}(k_2-k_1)v_j; \sum_{i\in I'} u_i = \sum_{j\in J'} v_j = 1, \sum_{i\in I'} u_i \mathbf{g}_i(\boldsymbol{\beta}_1) = \sum_{j\in J'} v_j \mathbf{g}_j(\boldsymbol{\beta}_2) = \mathbf{0}_d \Big\}.$$

Le profil du rapport de vraisemblance empirique pour tester H_0 contre H_2 est

$$U'_{0,n,k_1,k_2}(\boldsymbol{\beta}) / U'_{1,n,k_1,k_2}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2).$$

De la même manière que dans la Section 2.3, quand on a testé l'hypothèse de non-changement H_0 contre son alternative H_1, en utilisant la méthode des multiplicateurs de Lagrange, on obtient que sous l'hypothèse nulle H_0, les probabilités u_i et v_j sont

$$u_i = \frac{1}{(n-k_2+k_1) + n\boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta})} \qquad v_j = \frac{1}{(k_2-k_1) - n\boldsymbol{\lambda}_2^t \mathbf{g}_j(\boldsymbol{\beta})}. \tag{2.55}$$

En utilisant les mêmes arguments de la preuve du Théorème 2.3.1, on déduit que la loi limite de $-2\log U'_{1,n,k_1,k_2}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ est $\chi^2(2d)$, alors on peut considérer comme statistique de test $-2\log U'_{0,n,k_1,k_2}(\boldsymbol{\beta})$.

On se limite au cas où $\boldsymbol{\lambda}_1$ et $\boldsymbol{\lambda}_2$ vérifient la contrainte

$$\mathbf{V}_{1n}^{(1)}(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \mathbf{V}_{2n}^{(1)}(\boldsymbol{\beta})\boldsymbol{\lambda}_2,$$

avec

$$\mathbf{V}_{1n}^{(1)}(\boldsymbol{\beta}) = \frac{1}{n+k_1-k_2}\sum_{i\in I'} \dot{\mathbf{g}}(\boldsymbol{\beta}), \qquad \mathbf{V}_{2n}^{(1)}(\boldsymbol{\beta}) = \frac{1}{k_2-k_1}\sum_{j\in J'} \dot{\mathbf{g}}(\boldsymbol{\beta}).$$

Dans ce cas, on considère le paramètre $\theta_{n,k_1,k_2} = n^{-1}(n-k_2+k_1)$, qui dépend de deux points de rupture k_1, k_2. Alors la statistique de test peut s'écrire

$$2\sum_{i\in I'} \log\Big[1 + \frac{1}{\theta_{n,k_1,k_2}}\boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})\Big] + 2\sum_{j\in J'} \log\Big[1 - \frac{1}{1-\theta_{n,k_1,k_2}}\boldsymbol{\lambda}^t \mathbf{V}_{1n}^{(1)}(\boldsymbol{\beta})(\mathbf{V}_{2n}^{(1)})^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})\Big].$$

On note par $\widehat{\boldsymbol{\lambda}}(\theta_{n,k_1,k_2})$, $\widehat{\boldsymbol{\beta}}(\theta_{n,k_1,k_2})$ les solutions d'equations scores égales à zéro. De la même façon que dans la Section 2.3, on peut montrer que sous l'hypothèse H_0, cette statistique est équivalente asymptotiquement à la statistique

$$U_{n,k_1,k_2}(\theta_{n,k_1,k_2},\boldsymbol{\lambda},\boldsymbol{\beta}) \equiv 2\sum_{i\in I'} \log\Big[1 + \frac{1}{\theta_{n,k_1,k_2}}\boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})\Big] + 2\sum_{j\in J'} \log\Big[1 - \frac{1}{1-\theta_{n,k_1,k_2}}\boldsymbol{\lambda}^t \mathbf{g}_j(\boldsymbol{\beta})\Big].$$

Alors, pour tester l'hypothèse nulle H_0 contre l'hypothèse alternative H_2, on considère la statistique de test suivante

$$\max_{1<k_1<k_2<n} \left\{U_{n,k_1,k_2}(\theta_{n,k_1,k_2}, \widehat{\boldsymbol{\lambda}}(\theta_{n,k_1,k_2}), \widehat{\boldsymbol{\beta}}(\theta_{n,k_1,k_2}))\right\}.$$

Dans le cas où k_1 ou $n-k_2$ sont trop proches de 0, le maximum de la vraisemblance empirique peut ne pas exister. Dans ce cas, le test proposé ne peut pas détecter la présence des points de rupture dans le modèle. Pour que le maximum de vraisemblance empirique existe toujours, on considère les deux entières naturels Θ_{n1} et Θ_{n2}, tel que

$$\Theta_{n1} < k_1 < k_2 < n - \Theta_{n2}.$$

Finalement, pour tester l'hypothèse H_0 contre H_2, la statistique de test devient

$$\tilde{U}_n \equiv \max_{\Theta_{n1}<k_1<k_2<n-\Theta_{n2}} \left\{U_{n,k_1,k_2}(\theta_{n,k_1,k_2}, \widehat{\boldsymbol{\lambda}}(\theta_{n,k_1,k_2}), \widehat{\boldsymbol{\beta}}(\theta_{n,k_1,k_2}))\right\}. \tag{2.56}$$

Remarque 2.4.1. *Sous l'hypothèse nulle* H_0*, la statistique* $\tilde{U}_n$ *donnée par la relation (2.56) a la même distribution asymptotique que celle de la statistique* $\tilde{T}_n$ *donnée par la relation (2.30). En conséquence, pour une valeur fixée de* $\alpha \in (0,1)$*, on peut déduire la région critique suivante*

$$(\tilde{U}_n)^{1/2} \geq \frac{-\log(-\log\alpha) + D(\log u(n))}{A(\log u(n))},$$

où les fonctions $A(x)$*,* $D(x)$ *et* $u(n)$ *sont données dans la section 2.3.3.*

Dans la pratique, afin de faciliter l'utilisation de la statistique de test, on peut tout simplement obtenir la statistique correspondante donnée par la relation (2.33) de la Remarque 2.3.2.

2.5 Simulations

Dans cette section, afin d'évaluer la performance de la statistique de test proposée, on réalise des simulations en utilisant la méthode du Monte-Carlo. Pour un niveau théorique fixé, on calcule tout d'abord la valeur critique de la statistique de test, pour différentes valeurs de n. Ensuite, on utilise le technique du Monte-Carlo pour calculer le niveau empirique et la puissance du test empirique pour un modèle non-linéaire avec un point de rupture. Finalement, la statistique de test est étudiée pour un modèle épidémique. Toutes les simulations ont été effectuées en utilisant le logiciel R.

On considère la fonction non-linéaire suivante

$$f(x, \boldsymbol{\beta}) = a\frac{1-x^b}{b}, \tag{2.57}$$

avec $\boldsymbol{\beta} = (a, b) \in [-100, 100] \times [0.1, 20]$.

La même fonction a été considérée par [Ciuperca (2011b)], où le modèle est estimé par la méthode du minimum de la déviation absolue pénalisée.

2.5.1 Modèle avec un seul point de rupture

Pour la fonction non-linéaire donnée par la relation (2.57), sous l'hypothèse H_1, le modèle non-linéaire avec deux phases (un seul point de rupture) est

$$Y_i = a_1\frac{1-X_i^{b_1}}{b_1}\mathbb{1}_{i\leq k_0} + a_2\frac{1-X_i^{b_2}}{b_2}\mathbb{1}_{i>k_0} + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.58}$$

où $\mathbf{X} \sim \chi^2(2)$ et les vraies valeurs des paramètres $a_1^0 = 10$, $b_1^0 = 2$, $a_2^0 = 7$, $b_2^0 = 1.75$.

Sous l'hypothèse H_0, le modèle est

$$Y_i = a\frac{1-X_i^b}{b} + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.59}$$

avec les vrais valeurs des paramètres a, b : $a^0 = 10$, $b^0 = 2$.

L'absence du changement contre la présence d'un seul changement dans le modèle est testé en utilisant la statistique de test $\max_{\theta_{nk}} T^{1/2}(\theta_{nk})$, avec $T(\theta_{nk})$ donnée par la relation (2.33), les vecteurs $\mathbf{W}_{1n}^0$, $\mathbf{W}_{2n}^0$ étant calculés par la relation (2.34) et la matrice $\mathbf{V}$ par $n^{-1}\sum_{i=1}^n \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta}^0)\dot{\mathbf{f}}^t(\mathbf{X}_i, \boldsymbol{\beta}^0)$.

Pour le modèle sans point de rupture, spécifié par (2.59), afin de calculer le niveau empirique du test, on compte, le nombre des fois, parmi toutes les répliques Monte-Carlo qu'on a obtenu si $\max_{\theta_{nk}} T^{1/2}(\theta_{nk}) \geq c_\alpha$, où c_α désigne la valeur critique générique. Pour un niveau fixé $\alpha \in (0, 1)$, on considère la valeur critique conformément au Corollaire 2.3.1 :

$$c_\alpha = \frac{-\log(-\log\alpha) + D(\log u(n))}{A(\log u(n))},$$

tel que

$$u(n) = (n^2 + (2[\log n])^2 - 2n[\log n])(2[\log n])^{-2}.$$ [2]

Pour un niveau théorique $\alpha = 0.05$, les valeurs critique c_α sont calculées en prenant des différentes valeurs pour la taille de l'échantillon n (voir le Tableau 2.5.1).

Pour les modèles (2.58) et (2.59), on réalise 5000 répliques Monte-Carlo pour chacune de quatre distributions des erreurs suivantes : $\mathcal{N}(0,1)$, $2\mathcal{E}xp(2)-1$, $1/\sqrt{6}(\chi^2(3)-3)$ et $2/\sqrt{6}t(6)$, où $\mathcal{E}xp(2)$, $\chi^2(3)$ et $t(6)$ sont respectivement la distribution normale standard, la distribution exponentielle de moyenne 1/2, la distribution de chi-deux avec degré de liberté 3 et la distribution student de degré de liberté 6.

On note le niveau empirique et la puissance correspondante à la valeur critique c_α respectivement par $\widehat{\alpha}$ et $\widehat{\pi}$. Dans le Tableau 2.5.1, pour différentes valeurs de n et différentes positions du point de rupture, on donne les niveaux empiriques et les puissances obtenus pour les quatre distributions de l'erreur. Pour chaque localisation du point de rupture, on trouve que la puissance de la statistique de test empirique est 1. Comme il est mentionné

Tableau 2.5.1 – Valeur critique c_α, niveau empirique $\widehat{\alpha}$, puissance $\widehat{\pi}$, α=0.05

Erreur	n	k_0	c_α	$\widehat{\alpha}$	$\widehat{\pi}$
normale	200	75	1.133	0.112	1
	200	100	1.133	0.089	1
	400	250	1.340	0.068	1
	600	400	1.434	0.066	1
	1000	200	1.544	0.051	1
	1000	600	1.544	0.062	1
exponentielle	200	75	1.133	0.097	1
	1000	600	1.544	0.073	1
χ^2	200	75	1.133	0.108	1
	1000	600	1.544	0.067	1
student	200	75	1.133	0.124	1
	1000	600	1.544	0.081	1

2. voir aussi Ning et al. (2012).

Tableau 2.5.2 – Statistiques descriptive pour les estimations du point de rupture. Modèle avec deux phases, $n = 1000$, M= 5000.

distribution de l'erreur	k^0	$\widehat{k}_n$			
		$\min(\widehat{k}_n)$	$\max(\widehat{k}_n)$	moyenne$(\widehat{k}_n)$	médiane$(\widehat{k}_n)$
$\varepsilon_i \sim \mathcal{N}(0,1)$	200	171	229	201	198
	400	370	427	402	400
	600	573	630	597	600
$\varepsilon_i \sim 2/\sqrt{6}t(6)$	200	172	230	202	200
	400	370	427	400	397
	600	572	630	599	598
$\varepsilon_i \sim 2\mathcal{E}xp(2) - 1$	200	170	230	203	199
	400	372	430	400	403
	600	570	627	601	599
$\varepsilon_i \sim 1/\sqrt{6}(\chi^2(3) - 3)$	200	178	227	202	200
	400	373	430	405	401
	600	596	629	598	602

dans la Remarque 2.3.2, la localisation du point de rupture peut être estimée par la méthode de la vraisemblance empirique. Dans le Tableau 2.5.2 on donne quelques éléments de statistique descriptive (minimum, maximum, moyenne, médiane) pour les estimations $\widehat{k}_n$, calculées avec (2.35), pour 5000 répliques Monte-Carlo.

Vu les résultats présentés dans le Tableau 2.5.2, on déduit que la méthode d'estimation proposée donne une bonne approche pour la vraie valeur k^0, quelque soit la distribution des erreurs et la position du point de rupture dans l'intervalle $[1 : n]$.

Notons que, dans toutes les situations, le médiane et la moyenne des estimations du point de rupture soit ils coïncident, soit il sont très proches de la vraie valeur k^0.

2.5.2 Modèle épidémique

Pour la fonction non-linéaire donnée par la relation (2.57), sous l'hypothèse H_2 donnée par (2.54), on considère le modèle à trois phases (deux points de rupture) suivant

$$Y_i = a_1 \frac{1 - X_i^{b_1}}{b_1} \mathbb{1}_{i \leq k_1} + a_2 \frac{1 - X_i^{b_2}}{b_2} \mathbb{1}_{k_1 < i \leq k_2} + a_1 \frac{1 - X_i^{b_1}}{b_1} \mathbb{1}_{k_2 < i \leq n} + \varepsilon_i, \tag{2.60}$$

où $\mathbf{X} \sim \chi^2(2)$ et les vraies valeurs des paramètres $a_1^0 = 10$, $b_1^0 = 2$, $a_2^0 = 7$, $b_2^0 = 1.75$.

Sous l'hypothèse nulle H_0 représentée par le modèle (2.59), les paramètres sont $a^0 = 10$, $b^0 = 2$.

Dans le Tableau 2.5.3 on donne les résultats obtenus avec 200 répliques Monte-Carlo pour le modèle (2.60) sous l'hypothèse H_2 et pour le modèle (2.59) sous l'hypothèse H_0, pour $n = 400$ et des erreurs gaussiennes. Conformément à la Remarque 2.4.1, pour $n = 400$ la valeur critique calculée est $c_\alpha = 1.34$. On observe que le niveau empirique est inférieur à 0.05 ainsi que la puissance empirique est 1, pour toutes les positions de deux points de rupture considérées.

Tableau 2.5.3 – Puissance empirique et niveau empirique pour le modèle épidémique, $\alpha = 0.05$, $n = 400$, $\varepsilon \sim \mathcal{N}(0, 1)$.

k_1	k_2	puissance
no-change		0.033
100	300	1
100	200	1
150	300	1

2.6 Annexe

Le lemme suivant sera utilisé dans la démonstration des preuves des propositions, des théorèmes et des lemmes qui suivent.

Lemme 2.6.1. *Soit* $\mathbf{X} = (X_1, \ldots, X_p)$ *un vecteur aléatoire (colonne), tel que* $X_1, \ldots, X_p$ *sont des variables aléatoires qui ne sont pas nécessairement indépendantes, et* $\boldsymbol{M}$ *une matrice d'élément général* $(m_{ij})_{1 \le i,j \le p}$, *tel que* $\boldsymbol{M} = \mathbf{X}\mathbf{X}^t$. *Si pour j=1, ..., p, on a*

$$\text{pour tout } \eta_j > 0, \text{il existe } \delta_j > 0 \quad \text{tel que} \quad \mathbb{P}[|X_j| \ge \delta_j] \le \eta_j, \tag{2.61}$$

alors

(i) $\mathbb{P}\big[\|\mathbf{X}\|_1 \geq p \max_{1\leq j\leq p} \delta_j\big] \leq \max_{1\leq j\leq p} \eta_j$,

(ii) $\mathbb{P}\big[\|\mathbf{X}\| \geq \sqrt{p} \max_{1\leq j\leq p} \delta_j\big] \leq \max_{1\leq j\leq p} \eta_j$,

(iii) $\mathbb{P}\big[\|\boldsymbol{M}\|_1 \geq p \max_{1\leq i,j\leq p}\{\delta_i^2, \delta_j^2\}\big] \leq \max_{1\leq i,j\leq p}\{\eta_i^2, \eta_j^2\}$,

où $\|\boldsymbol{M}\|_1 = \max_{1\leq j\leq p}\{\sum_{i=1}^p |m_{ij}|\}$ *est la norme subordonné à la norme* $\|.\|_1$.

Preuve. (i) En utilisant la relation (2.61), on peut écrire

$$\begin{aligned}\mathbb{P}\big[\|\mathbf{X}\|_1 \geq p \max_{1\leq j\leq p} \delta_j\big] &\leq \mathbb{P}\big[p \max_{1\leq j\leq p} |X_j| \geq p \max_{1\leq j\leq p} \delta_j\big] \\ &= \mathbb{P}\big[\max_{1\leq j\leq p} |X_j| \geq \max_{1\leq j\leq p} \delta_j\big] \\ &\leq \max_{1\leq j\leq p} \eta_j.\end{aligned}$$

(ii) La relation de l'équation (2.61) est équivalente à

$$\mathbb{P}\big[X_j^2 \geq \delta_j^2\big] \leq \eta_j,$$

ce qui implique que

$$\begin{aligned}\mathbb{P}\big[\|\mathbf{X}\|^2 \geq p \max_{1\leq j\leq p} \delta_j^2\big] &= \mathbb{P}\big[\max_{1\leq j\leq p} X_j^2 \geq \max_{1\leq j\leq p} \delta_j^2\big] \\ &\leq \max_{1\leq j\leq p} \eta_j.\end{aligned}$$

L'affirmation *(ii)* est prouvée.

(iii) Pour $1 \leq i, j \leq p$, on a

$$\begin{aligned}\mathbb{P}\big[|X_iX_j| \geq \max\{\delta_i^2, \delta_j^2\}\big] &\leq \mathbb{P}\big[\max\{X_i^2, X_j^2\} \geq \max\{\delta_i^2, \delta_j^2\}\big] \\ &\leq \max\{\eta_i^2, \eta_j^2\}.\end{aligned}$$

Alors,

$$\mathbb{P}\big[|m_{ij}| \geq \max\{\delta_i^2, \delta_j^2\}\big] \leq \max\{\eta_i^2, \eta_j^2\}.$$

Donc, pour chaque $1 \leq j \leq p$, on peut écrire

$$\begin{aligned}\mathbb{P}\Big[\sum_{i=1}^p |m_{ij}| \geq p \max\{\delta_i^2, \delta_j^2\}\Big] &\leq \mathbb{P}\big[p \max_{1\leq i\leq p} |m_{ij}| \geq p \max\{\delta_i^2, \delta_j^2\}\big] \\ &\leq \max\{\eta_i^2, \eta_j^2\}.\end{aligned}$$

Alors,

$$\mathbb{P}\Big[\sum_{i=1}^{p}|m_{ij}| \geq p\max\{\delta_i^2,\delta_j^2\}\Big] \;\leq\; \max\{\eta_i^2,\eta_j^2\}.$$

■

Lemme 2.6.2. *Sous les suppositions (A1)-(A4), pour tout $\epsilon > 0$, il existe une constante positive M, tel que, pour $\eta_0 > 0$, pour tout $\eta \in [0,\eta_0]$ et tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$, on a*

$$\mathbb{P}\big[\|\mathbf{g}_i(\boldsymbol{\beta})\|_1 \geq M\big] \leq \epsilon.$$

Preuve. Dans la suite, pour alléger les notations, on note la fonction $\dot{\mathbf{f}}(\mathbf{X}_i,\boldsymbol{\beta})$ par $\dot{\mathbf{f}}_i(\boldsymbol{\beta})$ et la fonction $\ddot{\mathbf{f}}(\mathbf{X}_i,\boldsymbol{\beta})$ par $\ddot{\mathbf{f}}_i(\boldsymbol{\beta})$.

Le développement du Taylor de $\mathbf{g}_i(\boldsymbol{\beta})$, au voisinage de $\boldsymbol{\beta} = \boldsymbol{\beta}^0$, jusqu'à l'ordre 2, est

$$\begin{aligned}\mathbf{g}_i(\boldsymbol{\beta}) \;=\;& \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i + \frac{1}{2}\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\varepsilon_i - \frac{1}{2}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\\ &-\frac{1}{6}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) - \frac{1}{4}\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\\ &-\frac{1}{12}\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0}),\end{aligned} \tag{2.62}$$

où $\mathbf{M}_{1i} = \left(\frac{\partial^2 f_i(\boldsymbol{\beta}_{i,jk}^{(1)})}{\partial\beta_j\partial\beta_k}\right)_{1\leq j,k\leq d}$, $\mathbf{M}_{2i} = \left(\frac{\partial^2 f_i(\boldsymbol{\beta}_{i,jk}^{(2)})}{\partial\beta_j\partial\beta_k}\right)_{1\leq j,k\leq d}$ et $\boldsymbol{\beta}_{i,jk}^{(1)} = \boldsymbol{\beta^0} + u_{i,jk}(\boldsymbol{\beta}-\boldsymbol{\beta^0})$, $\boldsymbol{\beta}_{i,jk}^{(2)} = \boldsymbol{\beta^0} + v_{i,jk}(\boldsymbol{\beta}-\boldsymbol{\beta^0})$, avec $u_{i,jk}, v_{i,jk} \in [0,1]$. Faisons la remarque que $\boldsymbol{\beta}_{i,jk}^{(1)}$ et $\boldsymbol{\beta}_{i,jk}^{(2)}$ sont des vecteurs aléatoires qui dépendent de $\mathbf{X}_i$.

Pour le terme $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i$, on sait que $\mathbf{X}_i$ et ε_i sont indépendants ainsi que $\mathbb{E}(\varepsilon_i) = 0$, ce qui implique

$$\mathbb{E}[\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i] = 0$$

et

$$\mathbb{V}\mathrm{ar}\,[\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i] = \sigma^2\mathbf{V}.$$

Pour la j-ème composante de $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})$, par l'inégalité de Bienaymé-Tchebychev, pour $1 \leq j \leq d$, pour tout $\epsilon_1 > 0$, on a

$$\mathbb{P}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\varepsilon_i\Big| \geq \epsilon_1\Big] \leq \frac{\sigma^2}{\epsilon_1^2}V_{jj}, \tag{2.63}$$

avec V_{jj} est le j-ème élément diagonal de la matrice $\mathbf{V}$.

Pour tout $\epsilon > 0$, on prend $\epsilon_1 = \sigma\sqrt{6V_{jj}/\epsilon}$. Alors la relation (2.63) devient

$$\mathbb{P}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\varepsilon_i\Big| \geq \sigma\sqrt{6V_{jj}/\epsilon}\Big] \leq \frac{\epsilon}{6}.$$

En appliquant le Lemme 2.6.1 *(i)*, pour tout $\epsilon > 0$, on obtient

$$\mathbb{P}\Big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i\|_1 \geq \frac{\sigma d}{\sqrt{\epsilon}}\max_{1\leq j\leq d}\sqrt{6V_{jj}}\Big] \leq \frac{\epsilon}{6}. \tag{2.64}$$

Pour le deuxième terme du membre droit de la relation (2.62), d'après la supposition (A3), pour $1 \leq j, k \leq d$, pour tout $\epsilon > 0$ il existe $\epsilon_2 > 0$, tel que

$$\mathbb{P}\Big[\Big|\frac{\partial^2 f_i(\boldsymbol{\beta}^{(1)}_{i,jk})}{\partial\beta_j\partial\beta_k}\Big| \geq \epsilon_2\Big] \leq \frac{\epsilon}{6}.$$

Par le Lemme 2.6.1 (iii), pour tout $\epsilon > 0$, on a que

$$\mathbb{P}\big[\|\mathbf{M}_{1i}\|_1 \geq \epsilon_2\big] \leq \frac{\epsilon}{6}. \tag{2.65}$$

En utilisant l'inégalité de Bienaymé-Tchebychev et la supposition (A1), pour tout $C_1 > 0$, on obtient que

$$\mathbb{P}\big[|\varepsilon_i| > C_1\big] \leq \frac{\sigma^2}{C_1}. \tag{2.66}$$

Rappelons que $\|\boldsymbol{\beta} - \boldsymbol{\beta^0}\| < \eta$. Or les normes vectorielles $||.||_1$ et $||.||$ sont équivalentes. Alors, il existe $C > 0$ tel que

$$\|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|_1 \leq C\|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|.$$

Par les relations (2.65) et (2.66), on peut déduire que pour tout $\epsilon > 0$, il existe $\epsilon_2 > 0$ tel que,

$$\begin{aligned}
\mathbb{P}\big[\|\mathbf{M}_{1i}(\boldsymbol{\beta} - \boldsymbol{\beta^0})\varepsilon_i\|_1 \geq \epsilon_2\big] &\leq \mathbb{P}\big[\|\mathbf{M}_{1i}\|_1|\varepsilon_i|\ \|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|_1 \geq \epsilon_2\big] \\
&\leq \mathbb{P}\big[\|\mathbf{M}_{1i}\|_1 \geq \frac{\epsilon_2}{C_1\eta}\big] \\
&\leq \mathbb{P}\big[\|\mathbf{M}_{1i}\|_1 \geq \epsilon_2\big] \\
&\leq \frac{\epsilon}{6}.
\end{aligned}$$

Donc, pour tout $\epsilon > 0$, il existe $\epsilon_2 > 0$ tel que

$$\mathbb{P}\left[\|\mathbf{M}_{1i}(\boldsymbol{\beta} - \boldsymbol{\beta^0})\varepsilon_i\|_1 \geq \epsilon_2\right] \leq \frac{\epsilon}{6}. \tag{2.67}$$

Considérons maintenant le terme $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})$ de la relation (2.62).

Par l'inégalité de Markov et en tenant compte de la supposition (A4), pour $1 \leq j, l \leq d$, pour tout $\epsilon_3 > 0$, on obtient que

$$\mathbb{P}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right| \geq \epsilon_3\right] \leq \frac{1}{\epsilon_3}\mathbb{E}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right|\right]. \tag{2.68}$$

Pour tout $\epsilon > 0$, on choisit

$$\epsilon_3 = \frac{6}{\epsilon}\mathbb{E}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right|\right].$$

Alors, la relation (2.68) devient

$$\mathbb{P}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right| \geq \frac{6}{\epsilon}\mathbb{E}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right|\right]\right] \leq \frac{\epsilon}{6}.$$

En utilisant le Lemme 2.6.1 (iii), on obtient

$$\mathbb{P}\left[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j,l\leq d}\mathbb{E}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right|\right]\right] \leq \frac{\epsilon}{6}.$$

Puisque pour tout $C_2 > 0$, on a $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 \leq C_2\eta$, alors la dernière relation implique que

$$\begin{aligned}
&\mathbb{P}\left[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j,l\leq d}\mathbb{E}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right|\right]\right] \\
&\leq \mathbb{P}\left[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j,l\leq d}\mathbb{E}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right|\right]\right] \\
&\leq \frac{\epsilon}{6}.
\end{aligned}$$

Alors, pour tout $\epsilon > 0$

$$\mathbb{P}\left[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j,l\leq d}\mathbb{E}\left[\left|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_l}\right|\right]\right] \leq \frac{\epsilon}{6}. \tag{2.69}$$

Pour le terme $\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})$ de la relation (2.62), on utilise la supposition (A3) ainsi l'inégalité de Markov. Alors pour chaque terme $\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}$ qui représente la j-ème composante du vecteur $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})$, avec $1 \leq j \leq p$, pour tout $\epsilon_4 > 0$, on obtient

$$\mathbb{P}\Big[|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}| \geq \epsilon_4\Big] \leq \frac{1}{\epsilon_4}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big]. \tag{2.70}$$

Pour tout $\epsilon > 0$, on choisit

$$\epsilon_4 = \frac{6}{\epsilon}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big],$$

alors la relation (2.70) devient

$$\mathbb{P}\left[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big| \geq \frac{6}{\epsilon}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big]\right] \leq \frac{\epsilon}{6}.$$

Par le Lemme 2.6.1 *(i)*, pour tout $\epsilon > 0$, on obtient

$$\mathbb{P}\left[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big]\right] \leq \frac{\epsilon}{6}. \tag{2.71}$$

En utilisant la supposition (A3) et les relations (2.65) et (2.71), on peut écrire

$$\begin{aligned}
&\mathbb{P}\left[\|\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big]\right] \\
&\leq \mathbb{P}\left[\|\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big]\right] \\
&\leq \frac{\epsilon}{6}.
\end{aligned}$$

Donc, pour tout $\epsilon > 0$,

$$\mathbb{P}\left[\|\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big]\right] \leq \frac{\epsilon}{6}. \tag{2.72}$$

De même, en tenant compte des relations (2.65) et (2.71) et de suppositions (A3) et (A4), pour tout $\epsilon > 0$, on a que

$$\mathbb{P}\left[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \frac{6d}{\epsilon}\max_{1\leq j\leq d}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial \beta_j}\Big|\Big]\right] \leq \frac{\epsilon}{6}. \tag{2.73}$$

Pour le dernier terme du membre droit de la relation (2.62), en utilisant la supposition (A3), on a pour tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$, pour tout $\epsilon > 0$, qu'il existe $\epsilon_5 > 0$, tel que

$$\mathbb{P}\Big[\|\mathbf{M}_{1i}\|_1 \|\mathbf{M}_{2i}\|_1 \geq \epsilon_5\Big] \leq \frac{\epsilon}{6}.$$

En utilisant cette dernière relation, on obtient

$$\begin{aligned}
&\mathbb{P}\Big[\|\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \epsilon_5\Big] \\
&\leq \mathbb{P}\Big[\|\mathbf{M}_{1i}\|_1\|\mathbf{M}_{2i}\|_1 \geq \epsilon_5\Big] \\
&\leq \frac{\epsilon}{6}.
\end{aligned}$$

Alors, pour tout $\epsilon > 0$, il existe $\epsilon_5 > 0$, tel que,

$$\mathbb{P}\Big[\|\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq \epsilon_5\Big] \leq \frac{\epsilon}{6}. \tag{2.74}$$

En utilisant le développement limité de $\mathbf{g}_i(\boldsymbol{\beta})$, donné par la relation (2.62) au voisinage de $\boldsymbol{\beta} = \boldsymbol{\beta^0}$, on obtient pour tout $M > 0$:

$$\begin{aligned}
\mathbb{P}[\|\mathbf{g}_i(\boldsymbol{\beta})\|_1 \geq M] \;\leq\; & \mathbb{P}[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i\|_1 \geq M] + \mathbb{P}[\|\frac{1}{2}\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\varepsilon_i\|_1 \geq M] \\
&+\mathbb{P}[\|\frac{1}{2}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq M] \\
&+\mathbb{P}[\|\frac{1}{6}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq M] \\
&+\mathbb{P}[\|\frac{1}{4}\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq M] \\
&+\mathbb{P}[\|\frac{1}{12}\mathbf{M}_{1i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1 \geq M].
\end{aligned}$$

On choisit

$$M = \sup\left\{\epsilon_2, \epsilon_5, \frac{6d}{\epsilon}\max_{1\leq j,l\leq d}\left\{\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial\beta_j}\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial\beta_l}\Big|\Big], \mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta^0})}{\partial\beta_j}\Big|\Big]\right\}, \frac{\sigma d}{\sqrt{\epsilon}}\max_{1\leq j\leq d}\sqrt{6V_{jj}}\right\}.$$

En combinant les relations (2.64), (2.67), (2.69), (2.72), (2.73), (2.74), le lemme est prouvé. ■

Lemme 2.6.3. *Sous les mêmes suppositions du Théorème 2.3.2, on a*

$$\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}) = O_{\mathbb{P}}((n\theta_{nk})^{-1/2}) + \mathbf{V}^0_{1n}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^0).$$

Preuve. Le développement de Taylor de $\mathbf{g}_i(\boldsymbol{\beta})$ au voisinage de $\boldsymbol{\beta}=\boldsymbol{\beta^0}$ jusqu'à l'ordre 3 est

$$\begin{aligned}
\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}) \;=\; & \frac{1}{n\theta_{nk}}\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i + \frac{1}{2n\theta_{nk}}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\varepsilon_i \\
& -\frac{1}{2n\theta_{nk}}\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^{t}(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0}) \\
& -\frac{1}{6n\theta_{nk}}\sum_{i\in I}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) \\
& -\frac{1}{4n\theta_{nk}}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^{t}(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0}) \\
& -\frac{1}{12n\theta_{nk}}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) \\
& +\frac{1}{6n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i\varepsilon_i - \frac{1}{12n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0}),
\end{aligned} \tag{2.75}$$

où

$$\mathbf{M}_i = \left(\sum_{l=1}^{d}\sum_{k=1}^{d}\frac{\partial^2\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(3)}_{i,kl})}{\partial\beta_k\partial\beta_l}(\beta_k-\beta_k^0)(\beta_l-\beta_l^0)\right)_{1\le k,l\le d}$$

est le vecteur de dimension $(d\times 1)$, $\boldsymbol{\beta}^{(3)}_{i,kl} = \boldsymbol{\beta^0} + w_{i,kl}(\boldsymbol{\beta}-\boldsymbol{\beta^0})$, avec $w_{i,kl}\in[0,1]$, et $\mathbf{M}_{2i}$ défini dans le Lemme 2.6.2.
Pour le premier terme du membre droit de la relation (2.75), par le Théorème Central Limite et par le fait que $\mathbb{E}[\mathbf{g}_i(\boldsymbol{\beta^0})]=0$, on a

$$\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta^0}) = O_{\mathbb{P}}((n\theta_{nk})^{-1/2}). \tag{2.76}$$

Pour le deuxième terme du membre droit de la relation (2.75), par la loi forte des grands nombres, le terme $(n\theta_{nk})^{-1}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\varepsilon_i$ converge presque sûrement vers l'espérance de $\ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\varepsilon_i$, quand $n\to\infty$.

De plus, puisque ε_i est indépendant de $\mathbf{X}_i$ et $\mathbb{E}[\varepsilon_i] = 0$, on a

$$\frac{1}{n\theta_{nk}} \sum_{i\in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})\varepsilon_i = o_{\mathbb{P}}(\boldsymbol{\beta} - \boldsymbol{\beta^0}). \tag{2.77}$$

Pour le troisième terme du membre droit de la relation (2.75), en utilisant la supposition (A4) ainsi que la loi des grands nombres, on obtient que le terme $(n\theta_{nk})^{-1} \sum_{i\in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})$ converge presque sûrement vers l'espérance de $\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})$, quand $n \to \infty$.

D'autre part, puisque

$$\frac{1}{n\theta_{nk}} \sum_{i\in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i \xrightarrow{p.s.} 0,$$

on a

$$\frac{1}{n\theta_{nk}} \sum_{i\in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0}) = -\mathbf{V}_{1n}^0(\boldsymbol{\beta} - \boldsymbol{\beta^0})(1 + o_{\mathbb{P}}(1)). \tag{2.78}$$

Pour le quatrième terme du membre droit de la relation (2.75), en utilisant la loi des grands nombres, la supposition (A3) et la relation (2.71), on peut écrire

$$\begin{aligned}
&\|\frac{1}{6n\theta_{nk}} \sum_{i\in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t \mathbf{M}_{2i}(\boldsymbol{\beta} - \boldsymbol{\beta^0})\|_1 \\
&\leq \frac{1}{6n\theta_{nk}} \| \sum_{i\in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t \mathbf{M}_{2i}(\boldsymbol{\beta} - \boldsymbol{\beta^0})\|_1 \\
&= O_{\mathbb{P}}(\|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|_2^2),
\end{aligned}$$

ce qui implique que

$$\frac{1}{6n\theta_{nk}} \sum_{i\in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t \mathbf{M}_{2i}(\boldsymbol{\beta} - \boldsymbol{\beta^0}) = o_{\mathbb{P}}(\boldsymbol{\beta} - \boldsymbol{\beta^0}). \tag{2.79}$$

De même, pour le cinquième terme du membre droit de la relation (2.75), en utilisant la supposition (A3) et la relation (2.71), on obtient

$$\frac{1}{4n\theta_{nk}} \sum_{i\in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0}) = o_{\mathbb{P}}(\boldsymbol{\beta} - \boldsymbol{\beta^0}). \tag{2.80}$$

Pour le sixième terme du membre droit de la relation (2.75), en utilisant la supposition (A3), on obtient

$$\frac{1}{12n\theta_{nk}}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{2.81}$$

Pour $1 \leq j \leq d$, pour chaque i fixé, tel que $1 \leq i \leq n\theta_{nk}$, on note par M_{ij} la variable aléatoire suivante qui désigne la j-ème composante du vecteur $\mathbf{M}_i$, telle que

$$M_{ij} = \sum_{l=1}^{d}\sum_{k=1}^{d}\frac{\partial^3 f_i(\boldsymbol{\beta}^{(3)}_{i,kl})}{\partial\beta_k\partial\beta_l\partial\beta_j}(\beta_k-\beta_k^0)(\beta_l-\beta_l^0).$$

D'après la supposition (A3), il existe $C_3 > 0$ tel que

$$\mathbb{P}\Big[|M_{ij}| \leq C_3\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|^2\Big] = 1.$$

En utilisant le Lemme 2.6.1 *(i)*, on obtient

$$\|\mathbf{M}_i\|_1 \leq C_3\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|^2. \tag{2.82}$$

Pour le terme $(6n\theta_{nk})^{-1}\sum_{i\in I}\mathbf{M}_i\varepsilon_i$, en utilisant les relations (2.66) et (2.82), on a

$$\begin{aligned}
\|\frac{1}{6n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i\varepsilon_i\|_1 &\leq \frac{1}{6n\theta_{nk}}\|\sum_{i\in I}\mathbf{M}_i\varepsilon_i\|_1\\
&\leq \frac{1}{6n\theta_{nk}}\sum_{i\in I}\|\mathbf{M}_i\|_1|\varepsilon_i|\\
&\leq C_4\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|^2.
\end{aligned}$$

D'où,

$$\frac{1}{6n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i\varepsilon_i = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{2.83}$$

Finalement, pour le dernier terme du membre droit de la relation (2.75), en utilisant la supposition (A3) et la relation (2.82), on obtient avec une probabilité égale à 1, que

$$\begin{aligned}
&\|\frac{1}{12n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\|_1\\
&\leq \frac{1}{12n\theta_{nk}}\sum_{i\in I}\|\mathbf{M}_i\|_1\|(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\|_1\|\mathbf{M}_{2i}\|_1\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|_1\\
&\leq C_5\|\boldsymbol{\beta}-\boldsymbol{\beta^0}\|^2.
\end{aligned}$$

Ce qui donne,

$$\frac{1}{12n\theta_{nk}}\sum_{i\in I}\mathbf{M}_i(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t\mathbf{M}_{2i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})=o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{2.84}$$

Alors, en combinant les relations (2.76)-(2.81), (2.83) et (2.84), on obtient le lemme. ■

Lemme 2.6.4. *Sous les suppositions du Théorème 2.3.3, pour tout $\varrho>0$, il existe deux constantes positives $B=B(\varrho)$, $T=T(\varrho)$ telles que*

$\mathbb{P}[\max_{\frac{T}{n}\le\theta_{nk}\le1-\frac{T}{n}}(n\theta_{nk}/\log\log n\theta_{nk})^{1/2}\|\frac{\widehat{\boldsymbol{\lambda}}(\theta_{nk})}{\min\{\theta_{nk},1-\theta_{nk}\}}\|>B]\le\varrho,$

$\mathbb{P}[\max_{\frac{T}{n}\le\theta_{nk}\le1-\frac{T}{n}}(n\theta_{nk}/\log\log n\theta_{nk})^{1/2}\|\widehat{\boldsymbol{\beta}}(\theta_{nk})-\boldsymbol{\beta^0}\|>B]\le\varrho,$

$\mathbb{P}[n^{-1/2}\max_{\frac{T}{n}\le\theta_{nk}\le1-\frac{T}{n}}n\theta_{nk}\|\frac{\widehat{\boldsymbol{\lambda}}(\theta_{nk})}{\min\{\theta_{nk},1-\theta_{nk}\}}\|>B]\le\varrho,$

$\mathbb{P}[n^{-1/2}\max_{\frac{T}{n}\le\theta_{nk}\le1-\frac{T}{n}}n\theta_{nk}\|\widehat{\boldsymbol{\beta}}(\theta_{nk})-\boldsymbol{\beta^0}\|>B]\le\varrho.$

Preuve. La preuve de ce lemme est similaire à celle du Lemme 1.2.2 de [Csörgö and Horváth (1997)] . ■

Afin de prouver le Lemme 2.6.5, on considère

$$R_k=n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\mathbf{W}^0_{1n}-\mathbf{W}^0_{2n})^t\mathbf{V}^{-1}(\mathbf{W}^0_{1n}-\mathbf{W}^0_{2n}).$$

Rappelons que, pour $i=1,...,n$

$$\mathbf{V}\equiv\mathbb{E}[\dot{\mathbf{f}}(\mathbf{X}_i,\boldsymbol{\beta^0})\dot{\mathbf{f}}^t(\mathbf{X}_i,\boldsymbol{\beta^0})].$$

Les résultats du Lemme 2.6.5 sont similaires à ceux du Théorème 1.1.1 de [Csörgö and Horváth (1997)] .

Lemme 2.6.5. *Sous les suppositions (A1)-(A4), si l'hypothèse nulle* H_0 *est vraie, pour tout* $0\le\alpha<1/2$ *on a*

(i) $n^\alpha\max_{\theta_{nk}\in\Theta_{nk}}[\theta_{nk}(1-\theta_{nk})]^\alpha|Z_{nk}(\theta_{nk},\widehat{\boldsymbol{\lambda}}(\theta_{nk}),\widehat{\boldsymbol{\beta}}(\theta_{nk}))-R_k|=O_{\mathbb{P}}(1).$

(ii) $\max_{\theta_{nk}\in\Theta_{nk}}[\theta_{nk}(1-\theta_{nk})]|Z_{nk}(\theta_{nk},\widehat{\boldsymbol{\lambda}}(\theta_{nk}),\widehat{\boldsymbol{\beta}}(\theta_{nk}))-R_k|=O_{\mathbb{P}}(n^{-1/2}(\log\log n)^{3/2}).$

Preuve. Pour la fonction score $\boldsymbol{\phi}_{1n}$ de l'équation (2.14), on remplace les deux termes du membre droit par leurs décompositions obtenues respectivement par les relations (2.25) et (2.28).

D'autre part, on a

$$\boldsymbol{\phi}_{1n}(\theta_{nk}, \widehat{\boldsymbol{\lambda}}(\theta_{nk}), \widehat{\boldsymbol{\beta}}(\theta_{nk})) = \mathbf{0}_d.$$

Alors, on peut écrire

$$\begin{aligned}
&\Big[\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) + \mathbf{V}^0_{1n}(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}) - \frac{1}{n\theta^2_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}^t_i(\boldsymbol{\beta}^{\mathbf{0}})\widehat{\boldsymbol{\lambda}}(\theta_{nk})\Big](1+o_{\mathbb{P}}(1))\\
&-\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1}\Big[\frac{1}{n(1-\theta_{nk})^2}\mathbf{V}_{1n}(\mathbf{V}^0_{2n})^{-1}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}^t_j(\boldsymbol{\beta}^{\mathbf{0}})\widehat{\boldsymbol{\lambda}}(\theta_{nk})\\
&+\frac{1}{n(1-\theta_{nk})}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}}) + \mathbf{V}^0_{2n}(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}})\Big](1+o_{\mathbb{P}}(1)) = \mathbf{0}_d.
\end{aligned}$$

Ce qui donne,

$$\begin{aligned}
\widehat{\boldsymbol{\lambda}}(\theta_{nk}) \;=\; &\Big(\frac{1}{\theta_{nk}}\mathbf{D}^0_{1n} + \frac{1}{1-\theta_{nk}}(\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1})(\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1})^t\mathbf{D}^0_{2n}\Big)^{-1}\\
&\cdot\Big(\frac{1}{n\theta_{nk}}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) - \frac{\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1}}{n(1-\theta_{nk})}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\Big)(1+o_{\mathbb{P}}(1))\\
&+o_{\mathbb{P}}(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}),
\end{aligned}$$

où les matrices $\mathbf{D}^0_{1n}$ et $\mathbf{D}^0_{2n}$ sont données par la relation (2.27).

D'autre part, par la loi forte des grands nombres, on a $-\mathbf{V}^0_{1n} \xrightarrow[n\to\infty]{p.s.} \mathbf{V}$ et $-\mathbf{V}^0_{2n} \xrightarrow[n\to\infty]{p.s.} \mathbf{V}$.

Alors,

$$\mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1} \xrightarrow[n\to\infty]{p.s.} I_d.$$

Toujours, avec la loi des grands nombres, les deux matrices $\mathbf{D}^0_{1n}$ et $\mathbf{D}^0_{2n}$ convergent presque sûrement vers $\sigma^2\mathbf{V}$, quand $n \to \infty$.

Par le Théorème 2.3.2, on a démontré que $\widehat{\boldsymbol{\lambda}}(\theta_{nk}) = \theta_{nk}O_{\mathbb{P}}((n\theta_{nk})^{-1/2})$. Alors, on obtient

$$\widehat{\boldsymbol{\lambda}}(\theta_{nk}) = \sigma^{-2}\theta_{nk}(1-\theta_{nk})\mathbf{V}^{-1}(\mathbf{W}^0_{1n} - \mathbf{W}^0_{2n})(1+o_{\mathbb{P}}(1)) + o_{\mathbb{P}}(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}). \tag{2.85}$$

Le développement limité de la statistique $Z_{nk}(\theta_{nk}, \widehat{\boldsymbol{\lambda}}(\theta_{nk}), \widehat{\boldsymbol{\beta}}(\theta_{nk}))$, spécifiée par la relation (2.13), au voisinage de $(\boldsymbol{\lambda}, \boldsymbol{\beta}) = (\mathbf{0}_d, \boldsymbol{\beta}^{\mathbf{0}})$ jusqu'à l'ordre 2, est

$$
\begin{aligned}
& \Big[\frac{2\widehat{\boldsymbol{\lambda}}^t(\theta_{nk})}{\theta_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) - \frac{2\widehat{\boldsymbol{\lambda}}^t(\theta_{nk})}{1-\theta_{nk}} \mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1} \sum_{j\in J} \mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\Big] \\
- \; & \Big[\frac{\widehat{\boldsymbol{\lambda}}^t(\theta_{nk})}{(1-\theta_{nk})^2} \mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1} \sum_{j\in J} \mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}}) \mathbf{g}_j^t(\boldsymbol{\beta}^{\mathbf{0}}) \mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1} \widehat{\boldsymbol{\lambda}}(\theta_{nk}) \\
+ \; & \frac{\widehat{\boldsymbol{\lambda}}^t(\theta_{nk})}{\theta^2_{nk}} \sum_{i\in I} \mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}}) \mathbf{g}_i^t(\boldsymbol{\beta}^{\mathbf{0}}) \widehat{\boldsymbol{\lambda}}(\theta_{nk})\Big] + \Big[2\widehat{\boldsymbol{\lambda}}^t(\theta_{nk})\Big(\frac{1}{\theta_{nk}} \sum_{i\in I} \dot{\mathbf{g}}_i(\boldsymbol{\beta}^{\mathbf{0}}) \\
- \; & \frac{1}{1-\theta_{nk}} \mathbf{V}^0_{1n}(\mathbf{V}^0_{2n})^{-1} \sum_{j\in J} \dot{\mathbf{g}}_j(\boldsymbol{\beta}^{\mathbf{0}})\Big)(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}})\Big] \\
- \; & \Big[2\widehat{\boldsymbol{\lambda}}^t(\theta_{nk})\Big(\frac{1}{1-\theta_{nk}} \sum_{j\in J} \mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}}) \frac{\partial(\mathbf{V}_{1n}(\boldsymbol{\beta})(\mathbf{V}_{2n}(\boldsymbol{\beta}))^{-1})}{\partial \boldsymbol{\beta}}\Big)(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}})\Big] \\
+ \; & \frac{1}{3!}\Big[S_1 + 3S_2 + 3S_3 + S_4\Big], \\
& \equiv A_1 - A_2 + A_3 - A_4 + A_5. \qquad (2.86)
\end{aligned}
$$

avec

$S_1 = \sum_{j=1}^d \sum_{l=1}^d \sum_{k=1}^d \frac{\partial^3 Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}^{(1)}_{jkl}, \boldsymbol{\beta}^{(1)}_{jkl})}{\partial\beta_j \partial\beta_k \partial\beta_l} (\widehat{\beta}_j - \beta^0_j)(\widehat{\beta}_k - \beta^0_k)(\widehat{\beta}_l - \beta^0_l)$,

$S_2 = \sum_{j=1}^d \sum_{l=1}^d \sum_{k=1}^d \frac{\partial^3 Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}^{(2)}_{jkl}, \boldsymbol{\beta}^{(2)}_{jkl})}{\partial\lambda_j \partial\lambda_k \partial\beta_l} (\widehat{\lambda}_j)(\widehat{\lambda}_k)(\widehat{\beta}_l - \beta^0_l)$,

$S_3 = \sum_{j=1}^d \sum_{l=1}^d \sum_{k=1}^d \frac{\partial^3 Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}^{(3)}_{jkl}, \boldsymbol{\beta}^{(3)}_{jkl})}{\partial\lambda_j \partial\beta_k \partial\beta_l} (\widehat{\lambda}_j)(\widehat{\beta}_k - \beta^0_k)(\widehat{\beta}_l - \beta^0_l)$,

$S_4 = \sum_{j=1}^d \sum_{l=1}^d \sum_{k=1}^d \frac{\partial^3 Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}^{(4)}_{jkl}, \boldsymbol{\beta}^{(4)}_{jkl})}{\partial\lambda_j \partial\lambda_k \partial\lambda_l} (\widehat{\lambda}_j)(\widehat{\lambda}_k)(\widehat{\lambda}_l)$,

où, pour $1 \leq j \leq d$, $\widehat{\beta}_j$ est la j-ème composante de $\widehat{\boldsymbol{\beta}}(\theta_{nk})$, et $\widehat{\lambda}_j$ est la j-ème composante de $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$.

Dans les expressions de S_1, S_2,S_3, S_4, pour tout $1 \leq j, k, l \leq d$, on a $\boldsymbol{\lambda}^{(a)}_{jkl} = u^{(a)}_{jkl}(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}})$, et $\boldsymbol{\beta}^{(a)}_{jkl} = \boldsymbol{\beta}^{\mathbf{0}} + v^{(a)}_{jkl}(\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}})$, avec $u^{(a)}_{jkl}, v^{(a)}_{jkl} \in [0, 1]$ et $a \in \{1, 2, 3, 4\}$.

On note que, la dérivée de $\mathbf{V}_{1n}(\boldsymbol{\beta})(\mathbf{V}_{2n}(\boldsymbol{\beta}))^{-1}$ par rapport à $\boldsymbol{\beta}$ est considérée terme par terme.

Maintenant, on remplace $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$ dans la relation (2.86) par sa valeur obtenue en (2.85). Pour le terme A_1 de la relation (2.86), en utilisant les notations données par la relation (2.34), et le fait que $\mathbf{V}_{1n}^0(\mathbf{V}_{2n}^0)^{-1} \xrightarrow[n\to\infty]{p.s.} I_d$, quand $n \to \infty$, on obtient que

$$A_1 = 2n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0)^t\mathbf{V}^{-1}(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0) + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}\|). \tag{2.87}$$

De même, pour le terme A_2 de la relation (2.86), en utilisant les notations données par (2.27) et le fait que $\mathbf{D}_{1n}^0$ et $\mathbf{D}_{2n}^0$ convergent vers $\sigma^2\mathbf{V}$, quand $n \to \infty$, on obtient que

$$A_2 = n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0)^t\mathbf{V}^{-1}(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0) + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}\|). \tag{2.88}$$

Pour le terme A_3 de la relation (2.86), on sait que

$$\mathbf{V}_{1n}^0 = \frac{1}{n\theta_{nk}}\sum_{i\in I}\dot{\mathbf{g}}_i(\boldsymbol{\beta}^{\mathbf{0}}), \qquad \mathbf{V}_{2n}^0 = \frac{1}{n(1-\theta_{nk})}\sum_{j\in J}\dot{\mathbf{g}}_j(\boldsymbol{\beta}^{\mathbf{0}}).$$

D'autre part, par la loi des grands nombres, les deux matrices $\mathbf{V}_{1n}^0$ et $\mathbf{V}_{2n}^0$ convergent presque sûrement vers $-\mathbf{V}$, quand $n \to \infty$, ainsi que $\mathbf{V}_{1n}^0(\mathbf{V}_{2n}^0)^{-1} \xrightarrow[n\to\infty]{p.s.} I_d$. Ce qui impliquent que

$$A_3 \xrightarrow{p.s.} 0, \text{ quand } n \to \infty \tag{2.89}$$

Le Théorème Central Limite implique que

$$\frac{1}{n(1-\theta_{nk})}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}}) = O_{\mathbb{P}}((n(1-\theta_{nk}))^{-1/2}).$$

Alors, pour le terme A_4, de la relation (2.86), on obtient que

$$A_4 = o_{\mathbb{P}}(n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0)^t\mathbf{V}^{-1}(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0). \tag{2.90}$$

Pour le dernier terme de la relation (2.86), en utilisant les suppositions (A2)-(A4) et par des calculs élémentaires, on obtient que

$$A_5 = o_{\mathbb{P}}(\|\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}\|) + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\lambda}}(\theta_{nk})\|) + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\lambda}}(\theta_{nk})\|\|\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^{\mathbf{0}}\|). \tag{2.91}$$

En combinant les résultats obtenus par les relations (2.87)-(2.91), on obtient

$$\begin{aligned} Z(\theta_{nk}, \widehat{\boldsymbol{\lambda}}(\theta_{nk}), \widehat{\boldsymbol{\beta}}(\theta_{nk})) &= n\sigma^{-2}\theta_{nk}(1-\theta_{nk})(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0)^t\mathbf{V}^{-1}(\mathbf{W}_{1n}^0 - \mathbf{W}_{2n}^0)(1+o_{\mathbb{P}}(1)) \\ &\quad + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta^0}\|) + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\lambda}}(\theta_{nk})\|) \\ &\quad + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\lambda}}(\theta_{nk})\|\|\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta^0}\|). \end{aligned}$$

La dernière relation, avec le Lemme 2.6.4 impliquent le Lemme 2.6.5. ■

Chapitre 3

Intervalle de confiance pour la différence entre les paramètres de deux phases d'un modèle non-linéaire

Sommaire

3.1 Introduction 86
3.2 Suppositions, hypothèses, notations 88
3.3 Modèle avec des données complètes 90
3.3.1 Statistique de test 90
3.3.2 Comportement asymptotique de la statistique $\mathcal{Z}_{nk}$ 94
3.4 Modèle avec des données manquantes 101
3.4.1 Statistique de test 103
3.4.2 Comportement asymptotique de $\mathcal{Z}_{nk,C}, \mathcal{Z}_{nk,W}$ et $\mathcal{Z}_{nk,R}$ 109
3.5 Simulations 110
3.5.1 Modèle avec des données complètes 112
3.5.2 Modèle avec des données manquantes 112
3.6 Annexe 115

Abstract In this chapter, we use the empirical likelihood method to construct the confidence regions for the difference between the parameters of a two-phases nonlinear model with random design. We show that the empirical likelihood ratio has an asymptotic chi-squared distribution. The result is a nonparametric version of Wilk's theorem. Empirical

likelihood method is also used to construct the confidence regions for the difference between the parameters of a two-phases nonlinear model with response variables missing at randoms (MAR). In order to construct the confidence regions of the parameter in question, we propose three empirical likelihood statistics : Empirical likelihood based on complete-case data, weighted empirical likelihood and empirical likelihood with imputed values. We prove that all three empirical likelihood ratios have asymptotically chi-squared distributions. The effectiveness of the proposed approaches in aspects of coverage probability and interval length is demonstrated by Monte-Carlo simulations.

Key-words : Confidence region ; Empirical likelihood ; Two-phases problem ; Nonlinear regression model ; Missing response.

Résumé Dans ce chapitre, nous utilisons la méthode de vraisemblance empirique pour construire les régions de confiance pour la différence entre les paramètres de deux phases d'un modèle non-linéaire avec des régresseurs aléatoires. Nous montrons que le rapport de vraisemblance empirique a une distribution asymptotique chi-deux. Le résultat obtenu désigne une version non paramétrique du théorème de Wilks. La méthode de vraisemblance empirique est également utilisée pour construire les régions de confiance pour le différence entre les paramètres des deux phases d'un modèle non-linéaire avec des variables de réponse manquantes au hasard (Missing At Random (MAR)). Afin de construire les régions de confiance du paramètre en question, nous proposons trois statistiques de vraisemblance empiriques : la vraisemblance empirique basée sur les données cas-complète, la vraisemblance empirique pondérée et la vraisemblance empirique par des valeurs imputées. Nous prouvons que les trois rapports de vraisemblance empiriques ont des distributions asymptotiques chi-deux. L'efficacité des approches proposées, concernant la probabilité de recouvrement et la longueur de l'intervalle, est démontrée par des simulations de Monte-Carlo.

Mots-clefs : Région de confiance ; Vraisemblance empirique ; Problème de deux phases ; Modèle de régression non-linéaire ; Réponse manquante.

3.1 Introduction

On considère le modèle non-linéaire suivant

$$Y_i = \begin{cases} f(\mathbf{X}_i;\boldsymbol{\beta}) + \varepsilon_i & i = 1,\ldots,k, \\ f(\mathbf{X}_i;\boldsymbol{\beta}_1) + \varepsilon_i & i = k+1,\ldots,n, \end{cases} \tag{3.1}$$

où $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_1$ sont des vecteurs de d-paramètres (inconnus), $\mathbf{X}_i$ est un vecteur aléatoire de p regresseurs, avec la fonction de distribution $\mathrm{H}(\mathbf{x})$, pour $\mathbf{x} \in \Upsilon$ et l'ensemble compact $\Upsilon \subseteq \mathbb{R}^p$. On considère le vecteur $\mathbf{Y} = (Y_1, \ldots, Y_n)$, où pour chaque observation i, Y_i désigne la variable réponse (qui peut avoir des valeurs manquantes). Pour $1 \leq i \leq n$, on suppose que les erreurs ε_i sont des variables aléatoires indépendantes avec une moyenne zero. De plus, pour les observations $1 \leq i \leq k$ les erreurs sont identiquement distribués de variance σ_1^2, ainsi que celles du observations $k+1 \leq i \leq n$ sont identiquement distribués de variance σ_2^2. La suite de vecteurs aléatoires continus $(\mathbf{X}_i)_{1\leq i\leq n}$ sont indépendants et identiquement distribués (i.i.d), avec la même distribution jointe que $(\mathbf{X})$. Pour tout i, ε_i est indépendante de $\mathbf{X}_i$. La fonction de régression $f : \Upsilon \times \Gamma \to \mathbb{R}$, avec $\Upsilon \subseteq \mathbb{R}^p$ et $\Gamma \subseteq \mathbb{R}^d$, est connue jusqu'au paramètre $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)$. L'ensemble de paramètres Γ est supposé compact.

Pour le modèle (3.1), on considère le paramètre $\boldsymbol{\gamma}$ qui désigne la différence entre les paramètres de deux phases du modèle,

$$\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}_1.$$

On suppose d'abord que pour chaque observation i, la variable réponse Y_i est observée. On construit la région de confiance asymptotique pour $\boldsymbol{\gamma}$, ou on teste l'hypothèse nulle

$$\mathrm{H}_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0,$$

avec $\boldsymbol{\gamma}_0$ un vecteur connu.

Ensuite, on construit les régions de confiance pour $\boldsymbol{\gamma}$, quand certaines valeurs de $\mathbf{Y}$ sont manquantes alors que toutes les valeurs de la variable $\mathbf{X}$ sont observées. Alors, on obtient un échantillon incomplète $\{(\mathbf{X}_i, Y_i, \delta_i)_{1\leq i\leq n}\}$ du modèle (3.1), dont toutes les $\mathbf{X}_i$ sont observées, et $(\delta_i)_{1\leq i\leq n}$ désigne une suite des variables aléatoires, telle que $\delta_i = 0$ si Y_i est manquante et $\delta_i = 1$ sinon.

On suppose que Y_i est manquante au hasard (Missing at random : MAR). La supposition MAR implique que δ_i et Y_i sont indépendants conditionnellement à $\mathbf{X}_i$. C'est-à-dire, pour tout $1 \leq i \leq n$, on a

$$\mathbb{P}[\delta_i = 1 | \mathbf{X}_i, Y_i] = \mathbb{P}[\delta_i = 1 | \mathbf{X}_i].$$

La supposition MAR est une condition commune dans les analyses statistiques qui étudient les données manquantes et raisonnable dans plusieurs situations pratiques. Voir par exemple [Little and Rubin (1987)], [Qin and al. (2009)] et [Ciuperca (2013a)].

Pour le modèle (3.1), on utilise la méthode de vraisemblance empirique pour construire la région de confiance asymptotique pour $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}_1$, tout d'abord, si pour chaque $i = 1, \ldots, n$ la variable réponse Y_i est observée, ensuite, quand la variable réponse Y_i peut être manquante. Pour le modèle avec des données complètes, on propose une statistique de vraisemblance empirique et on prouve qu'elle a une distribution asymptotique chi-deux, ce qui implique la région de confiance asymptotique de $\boldsymbol{\gamma}$. Pour un modèle avec des donnés manquantes, on propose trois statistiques de test : Vraisemblance empirique basée sur les données complètes, vraisemblance empirique pondérée et vraisemblance empirique avec des valeurs imputées. On montre que ces trois rapports de vraisemblance empirique ont une distribution asymptotique chi-deux. On généralise ainsi les papiers de [Zi et al. (2010)] et de [Yu et al. (2014)], dans le cas où le modèle est non-linéaire.

Ce chapitre est organisé comme suit. Dans la Section 3.2, on introduit les suppositions, les notations et les hypothèses. Dans la Section 3.3, on formule le rapport de la vraisemblance empirique pour le modèle avec des données complètes. Ensuite, on prouve que la statistique de vraisemblance empirique a une distribution asymptotique chi-deux. En se basant sur cette statistique, on construit la région de confiance pour le paramètre $\boldsymbol{\gamma}$. Pour le modèle avec des données manquantes, les régions de confiance asymptotiques pour $\boldsymbol{\gamma}$, correspondant à trois statistiques de vraisemblance empirique, sont données dans la Section 3.4. Des simulations sont présentées dans la Section 3.5. Des lemmes nécessaires pour les démonstrations des théorèmes et des propositions sont énoncées et prouvés dans la Section 3.6.

3.2 Suppositions, hypothèses, notations

Puisque l'étude dans ce chapitre sera consacré au point de rupture dans les paramètres de régression d'un modèle non-linéaire, alors, on considère les mêmes suppositions (A1)-(A4) du Chapitre 2. Nous rappelons que ces suppositions concernent les erreurs, la fonction de régression et le vecteur aléatoire $\mathbf{X}$.

Nous sommes intéressés à construire la région de confiance asymptotique pour $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}_1$ et dans le même temps de tester l'hypothèse nulle

$$\mathrm{H}_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0, \tag{3.2}$$

avec $\boldsymbol{\gamma}_0$ un $d-$vecteur connu. L'hypothèse alternative de (3.2), est

$$\mathrm{H}_1 : \boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0. \tag{3.3}$$

Sous l'hypothèse H_0, on note la vraie valeur (inconnue) de $\boldsymbol{\beta}$ par $\boldsymbol{\beta^0}$, où $\boldsymbol{\beta}$ est la valeur générique du paramètre de régression de la première phase et $\boldsymbol{\beta}_1^0$ la vraie valeur de $\boldsymbol{\beta}_1$ pour la deuxième phase. Si $\boldsymbol{\gamma}_0^0 = \mathbf{0}_d$, alors $\boldsymbol{\beta}_1^0 = \boldsymbol{\beta^0}$.

La position du point de rupture k dépend de la taille de l'échantillon n. Donc, $k = k_n$. Pour simplifier les notations on utilise k partout. De plus, on suppose que $\lim_{n\to\infty} \frac{k}{n} \in (0, 1)$ qui désigne une condition classique. D'autre part, pour une valeur n donnée, on suppose que la position du point de rupture k est fixé. Pour simplifier les notations, on considère les ensembles des entiers naturels suivants : $I \equiv \{1, ..., k\}$ et $J \equiv \{k + 1, ..., n\}$. Ces deux ensembles contiennent les observations de deux segments du modèle (3.1). Correspondant aux deux ensembles I et J, soient les vecteurs de probabilités $(p_1, \ldots, p_k)$ et $(q_{k+1}, \ldots, q_n)$, où

$$\sum_{i\in I} p_i = 1 \text{ et } \sum_{j\in J} q_j = 1.$$

Pour $i \in I$, on considère le $d-$vecteur aléatoire suivant

$$\mathbf{g}_i(\boldsymbol{\beta}) \equiv \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta})[Y_i - f(\mathbf{X}_i, \boldsymbol{\beta})].$$

Pour la deuxième phase, à la place de Y_j, on considère la nouvelle variable Y_j^* calculée sur la base de Y_j et en tenant compte de la différence entre les paramètres $\boldsymbol{\beta} - \boldsymbol{\gamma}_0$ pour la fonction de régression. Pour les observations $j \in J$, soient

$$Y_j^* \equiv Y_j - f(\mathbf{X}_j, \boldsymbol{\beta} - \boldsymbol{\gamma}^0) + f(\mathbf{X}_j, \boldsymbol{\beta})$$

et

$$\mathbf{g}_j(\boldsymbol{\beta}) \equiv \dot{\mathbf{f}}(\mathbf{X}_j, \boldsymbol{\beta})[Y_j^* - f(\mathbf{X}_j, \boldsymbol{\beta})].$$

Sous l'hypothèse H_0, pour $i \in I$, on a

$$\mathbf{g}_i(\boldsymbol{\beta^0}) = \dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta^0})\varepsilon_i$$

et pour $j \in J$ on a

$$\mathbf{g}_j(\boldsymbol{\beta^0}) = \dot{\mathbf{f}}(\mathbf{X}_j, \boldsymbol{\beta^0})\varepsilon_j.$$

Pour tout $i = 1, \ldots, n$, sous H_0, on a

$$\mathbb{E}[\mathbf{g}_i(\boldsymbol{\beta^0})] = \mathbf{0}_d.$$

On considère la matrice suivante de dimension $d \times d$

$$\mathbf{V} \equiv \mathbb{E}[\dot{\mathbf{f}}(\mathbf{X}_i, \boldsymbol{\beta^0})\dot{\mathbf{f}}^t(\mathbf{X}_i, \boldsymbol{\beta^0})].$$

3.3 Modèle avec des données complètes

Dans cette section, pour le modèle non-linéaire donné par la relation (3.1), on suppose que pour chaque observation $i = 1, \ldots, n$, la variable réponse Y_i est observée. On construit la statistique du rapport de vraisemblance empirique et on montre que cette statistique a une distribution asymptotique χ^2, ce qui nous conduit à construire la région de confiance asymptotique pour $\boldsymbol{\gamma}$.

3.3.1 Statistique de test

Dans cette section, on formule la statistique du rapport de vraisemblance empirique qu'on va l'utiliser pour construire la région de confiance asymptotique pour $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}_1$.

Sous l'hypothèse H_0, on a $\boldsymbol{\gamma}_0 = \boldsymbol{\beta^0} - \boldsymbol{\beta}_1^0$. Remarquons que si les paramètres $\boldsymbol{\gamma}_0$ et $\boldsymbol{\beta^0}$ sont donnés, alors $\boldsymbol{\beta}_1^0$ (la vraie valeur de $\boldsymbol{\beta}_1$ sous H_0) est connu.

Afin d'étudier la statistique de vraisemblance empirique, puisque le paramètre $\boldsymbol{\beta^0}$ est inconnu, on va utiliser la notation $\boldsymbol{\beta}$. Alors, le profil de la vraisemblance empirique pour $\boldsymbol{\gamma}$, égal à $\boldsymbol{\gamma}_0$ sous H_0 est défini par

$$\mathcal{R}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\beta}) \equiv \sup_{(p_1,\ldots,p_k)} \sup_{(q_{k+1},\ldots,q_n)} \Big\{ \prod_{i\in I} p_i \prod_{j\in J} q_j ; \sum_{i\in I} p_i = 1, \sum_{j\in J} q_j = 1, \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\}.$$

Sans la contrainte $\sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{0}_d$, les maximums de $\prod_{i\in I} p_i$ et $\prod_{j\in J} q_j$ sont atteints respectivement pour $p_i = k^{-1}$ et $q_j = (n-k)^{-1}$. Alors, sous H_0, le profil du rapport de vraisemblance empirique pour $\boldsymbol{\gamma}$ évalué en $\boldsymbol{\gamma}_0$, a la forme

$$\mathcal{R}'_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\beta}) \equiv \sup_{(p_1,\ldots,p_k)} \sup_{(q_{k+1},\ldots,q_n)} \Big\{ \prod_{i\in I} kp_i \prod_{j\in J} (n-k) q_j ; \sum_{i\in I} p_i = 1, \sum_{j\in J} q_j = 1, \sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d \Big\}.$$

Sous l'hypothèse H_1, on peut estimer les paramètres $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_1$. On considère $\widehat{\boldsymbol{\beta}}$ et $\widehat{\boldsymbol{\beta}}_1$ les estimateurs de moindres carrés respectivement pour $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_1$

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i\in I} (Y_i - f(\mathbf{X}_i, \boldsymbol{\beta}))^2, \quad \widehat{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{\beta}_1} \sum_{j\in J} (Y_j - f(\mathbf{X}_j, \boldsymbol{\beta}_1))^2. \tag{3.4}$$

Alors, sous l'hypothèse H_1, le profil de la vraisemblance empirique pour $\boldsymbol{\gamma}$ est

$$\mathcal{R}_{nk}(\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}_1) \equiv \sup_{(p_1,\ldots,p_k)}\sup_{(q_{k+1},\ldots,q_n)}\Big\{\prod_{i\in I}p_i\prod_{j\in J}q_j;\sum_{i\in I}p_i=1,\sum_{j\in J}q_j=1,\\ \sum_{i\in I}p_i\mathbf{g}_i(\widehat{\boldsymbol{\beta}})=\sum_{j\in J}q_j\mathbf{g}_j(\widehat{\boldsymbol{\beta}}_1)=\mathbf{0}_d\Big\},$$

où $\widehat{\boldsymbol{\gamma}}=\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_1$ et $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}_1$ sont donnés par la relation (3.4).

Sous l'hypothèse H_1, le profil du rapport de vraisemblance empirique pour $\boldsymbol{\gamma}$ a la forme

$$\mathcal{R}'_{nk}(\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}_1) \equiv \sup_{(p_1,\ldots,p_k)}\sup_{(q_{k+1},\ldots,q_n)}\Big\{\prod_{i\in I}kp_i\prod_{j\in J}(n-k)q_j;\sum_{i\in I}p_i=1,\sum_{j\in J}q_j=1,\\ \sum_{i\in I}p_i\mathbf{g}_i(\widehat{\boldsymbol{\beta}})=\mathbf{0}_d,\sum_{j\in J}q_j\mathbf{g}_j(\widehat{\boldsymbol{\beta}}_1)=\mathbf{0}_d\Big\}.$$

En utilisant une idée similaire au test du maximum de vraisemblance empirique pour tester H_0 contre H_1, on considère le profil du rapport de vraisemblance empirique

$$\frac{\mathcal{R}_{nk}(\boldsymbol{\gamma}_0,\boldsymbol{\beta})}{\mathcal{R}_{nk}(\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}_1)}=\frac{\mathcal{R}'_{nk}(\boldsymbol{\gamma}_0,\boldsymbol{\beta})}{\mathcal{R}'_{nk}(\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}_1)}. \tag{3.5}$$

Par le Théorème 2.3.1 du Chapitre 2, sous l'hypothèse H_1, le profil du rapport de vraisemblance empirique pour $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}_1$ a une distribution asymptotique χ^2 avec $2d$ degrés de liberté. Alors, $\mathcal{R}'_{nk}(\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}_1)$ ne dépend pas asymptotiquement de paramètres $\widehat{\boldsymbol{\beta}}$ et $\widehat{\boldsymbol{\beta}}_1$. Donc, la fonction du log-vraisemblance empirique notée par $\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0,\boldsymbol{\beta})$ peut s'écrire

$$\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0,\boldsymbol{\beta}) = -2\sup_{(p_1,\ldots,p_k)}\sup_{(q_{k+1},\ldots,q_n)}\Big\{\sum_{i\in I}\log(kp_i)+\sum_{j\in J}\log((n-k)q_j);\sum_{i\in I}p_i=1,\\ \sum_{j\in J}q_j=1,\sum_{i\in I}p_i\mathbf{g}_i(\boldsymbol{\beta})=\mathbf{0}_d,\sum_{j\in J}q_j\mathbf{g}_j(\boldsymbol{\beta})=\mathbf{0}_d\Big\}. \tag{3.6}$$

Afin d'étudier $\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0,\boldsymbol{\beta})$ donnée par la relation (3.6), on considère tout d'abord les observations de la première phase ($i\in I$). En utilisant la méthode des multiplicateurs de Lagrange, le processus aléatoire suivant

$$\sum_{i\in I}\log p_i+\eta(\sum_{i\in I}p_i-1)-n\boldsymbol{\lambda}_1^t\sum_{i\in I}p_i\mathbf{g}_i(\boldsymbol{\beta}), \tag{3.7}$$

doit être maximisé par rapport à $p_1, \ldots, p_k$, η et $\boldsymbol{\lambda}_1$, où $\boldsymbol{\lambda}_1$ et η sont les multiplicateurs de Lagrange tels que $\boldsymbol{\lambda}_1 \in \mathbb{R}^d$ et $\eta \in \mathbb{R}$.

En utilisant le fait que la dérivée du processus donné par (3.7) par rapport à p_i est égale à zero, alors par des calculs similaires à celles du Théorème 2.3.1, on obtient

$$p_i = \frac{1}{n\boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta}) - \eta}. \tag{3.8}$$

Par la relation (3.8), on obtient que $\eta = k$. Donc, la probabilité p_i devient

$$p_i = \frac{1}{k + n\boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta}))}. \tag{3.9}$$

De même, pour les observations de la deuxième phase de (3.6), pour $j \in J$, on obtient

$$q_j = \frac{1}{n - k - n\boldsymbol{\lambda}_2^t \mathbf{g}_j(\boldsymbol{\beta})}, \tag{3.10}$$

où $\boldsymbol{\lambda}_2 \in \mathbb{R}^d$ est le multiplicateur de Lagrange.

En utilisant les relations (3.9) et (3.10), étudier la statistique de (3.6) revient à maximiser le processus aléatoire suivant par rapport à $\boldsymbol{\lambda}_1$ et $\boldsymbol{\lambda}_2$

$$\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\beta}) \equiv 2\Big[\sum_{i\in I} \log(1 + \frac{n}{k}\boldsymbol{\lambda}_1^t \mathbf{g}_i(\boldsymbol{\beta})) + \sum_{j\in J} \log(1 - \frac{n}{n-k}\boldsymbol{\lambda}_2^t \mathbf{g}_j(\boldsymbol{\beta}))\Big]. \tag{3.11}$$

Afin d'avoir un seul multiplicateur de Lagrange, noté par $\boldsymbol{\lambda}$, on limite l'étude sur un cas particulier, où $\boldsymbol{\lambda}_1$ et $\boldsymbol{\lambda}_2$ vérifient la contrainte

$$\mathbf{V}_{1n}(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \mathbf{V}_{2n}(\boldsymbol{\beta})\boldsymbol{\lambda}_2,$$

avec

$$\mathbf{V}_{1n}(\boldsymbol{\beta}) \equiv \frac{1}{k}\sum_{i\in I} \dot{\mathbf{g}}_i(\boldsymbol{\beta}), \quad \mathbf{V}_{2n}(\boldsymbol{\beta}) \equiv \frac{1}{n-k}\sum_{j\in J} \dot{\mathbf{g}}_j(\boldsymbol{\beta}).$$

Dans le cas du vrai paramètre $\boldsymbol{\beta}^{\mathbf{0}}$, ces deux matrices $\mathbf{V}_{1n}(\boldsymbol{\beta})$ et $\mathbf{V}_{2n}(\boldsymbol{\beta})$ seront notées respectivement par

$$\mathbf{V}_{1n}^0 \equiv \mathbf{V}_{1n}(\boldsymbol{\beta}^{\mathbf{0}}), \qquad \mathbf{V}_{2n}^0 \equiv \mathbf{V}_{2n}(\boldsymbol{\beta}^{\mathbf{0}}).$$

En considérant la contrainte $\mathbf{V}_{1n}(\boldsymbol{\beta})\boldsymbol{\lambda}_1 = \mathbf{V}_{2n}(\boldsymbol{\beta})\boldsymbol{\lambda}_2$, la statistique (3.11) devient

$$\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) \equiv 2\Big[\sum_{i\in I}\log(1+\frac{n}{k}\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})) + \sum_{j\in J}\log(1-\frac{n}{n-k}\boldsymbol{\lambda}^t\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta}))\Big]. \tag{3.12}$$

Afin d'étudier le maximum de la statistique de test de log-vraisemblance empirique $\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta})$ par rapport à $\boldsymbol{\beta}$ et $\boldsymbol{\lambda}$, on calcule les fonctions scores de la statistique de test (3.12), telles que

$$\begin{aligned}\tilde{\boldsymbol{\phi}}_{1n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) &\equiv \frac{\partial\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\lambda}, \boldsymbol{\gamma}_0, \boldsymbol{\beta})}{2\partial\boldsymbol{\lambda}} \\ &= \sum_{i\in I}\frac{\mathbf{g}_i(\boldsymbol{\beta})}{\frac{k}{n}+\boldsymbol{\lambda}^t\mathbf{g}_i(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}{\frac{n-k}{n}-\boldsymbol{\lambda}^t\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}.\end{aligned} \tag{3.13}$$

$$\begin{aligned}\tilde{\boldsymbol{\phi}}_{2n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) &\equiv \frac{\partial\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\lambda}, \boldsymbol{\gamma}_0, \boldsymbol{\beta})}{2\partial\boldsymbol{\beta}} \\ &= \sum_{i\in I}\frac{\dot{\mathbf{g}}_i(\boldsymbol{\beta})\boldsymbol{\lambda}^t}{\frac{k}{n}+\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{g}_i(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\partial(\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta}))/\partial\boldsymbol{\beta}}{\frac{n-k}{n}-\boldsymbol{\lambda}^t\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta})}\boldsymbol{\lambda}^t.\end{aligned} \tag{3.14}$$

En résolvant le système $\tilde{\boldsymbol{\phi}}_{1n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$ et $\tilde{\boldsymbol{\phi}}_{2n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$, les solutions obtenues $\tilde{\boldsymbol{\lambda}}_n(k)$ et $\tilde{\boldsymbol{\beta}}_n(k)$ maximisent la statistique (3.12).

Remarque 3.3.1. *Pour obtenir la forme symétrique de la statistique $\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta})$ donnée par la relation (3.12), on considère les notations suivantes, respectivement pour $i \in I$ et $j \in J$*

$$\mathbf{z}_i(\boldsymbol{\beta}) \equiv \mathbf{M}_n^{\frac{1}{2}}(\boldsymbol{\beta})\mathbf{V}_{1n}^{-1}(\boldsymbol{\beta})\mathbf{g}_i(\boldsymbol{\beta}), \qquad \mathbf{z}_j(\boldsymbol{\beta}) \equiv \mathbf{M}_n^{\frac{1}{2}}(\boldsymbol{\beta})\mathbf{V}_{2n}^{-1}(\boldsymbol{\beta})\mathbf{g}_j(\boldsymbol{\beta}),$$

où

$$\mathbf{M}_n(\boldsymbol{\beta}) \equiv \frac{k(n-k)}{n^2}\mathbf{V}_{1n}(\boldsymbol{\beta})\mathbf{H}_n(\boldsymbol{\beta})\mathbf{V}_{2n}(\boldsymbol{\beta})$$

et

$$\mathbf{H}_n(\boldsymbol{\beta}) \equiv \Big[\frac{k}{n}\sigma_2^2\mathbf{V}_{1n}(\boldsymbol{\beta}) + \frac{n-k}{n}\sigma_1^2\mathbf{V}_{2n}(\boldsymbol{\beta})\Big]^{-1}.$$

En tenant compte des notations données ci-dessus, on considère à la place de $\tilde{\mathcal{Z}}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta})$ *donnée par (3.12), la statistique de test suivante*

$$\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) \equiv 2\Big[\sum_{i\in I}\log(1 + \frac{n}{k}\boldsymbol{\lambda}^t\mathbf{z}_i(\boldsymbol{\beta})) + \sum_{j\in J}\log(1 - \frac{n}{n-k}\boldsymbol{\lambda}^t\mathbf{z}_j(\boldsymbol{\beta})\Big]. \tag{3.15}$$

En basant sur la relation (3.15), on peut dériver les equations scores suivantes pour obtenir les estimateurs $(\widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k))$ *de* $(\boldsymbol{\lambda}, \boldsymbol{\beta})$

$$\begin{cases} \boldsymbol{\phi}_{1n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) \equiv \dfrac{\partial\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta})}{2\partial\boldsymbol{\lambda}} = \displaystyle\sum_{i\in I}\frac{\mathbf{z}_i(\boldsymbol{\beta})}{\frac{k}{n} + \boldsymbol{\lambda}^t\mathbf{z}_i(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\mathbf{z}_j(\boldsymbol{\beta})}{\frac{n-k}{n} - \boldsymbol{\lambda}^t\mathbf{z}_j(\boldsymbol{\beta})}, \\ \boldsymbol{\phi}_{2n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) \equiv \dfrac{\partial\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta})}{2\partial\boldsymbol{\beta}} = \displaystyle\sum_{i\in I}\frac{\dot{\mathbf{z}}_i(\boldsymbol{\beta})}{\frac{k}{n} + \boldsymbol{\lambda}^t\mathbf{z}_i(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\dot{\mathbf{z}}_j(\boldsymbol{\beta})}{\frac{n-k}{n} - \boldsymbol{\lambda}^t\mathbf{z}_j(\boldsymbol{\beta})}, \end{cases} \tag{3.16}$$

où $\dot{\mathbf{z}}_i(\boldsymbol{\beta})$ *et* $\dot{\mathbf{z}}_j(\boldsymbol{\beta})$ *sont respectivement les dérivées de* $\mathbf{z}_i(\boldsymbol{\beta})$ *et* $\mathbf{z}_j(\boldsymbol{\beta})$ *par rapport à* $\boldsymbol{\beta}$*.*

Pour $i \in I$, la variance de ε_i est σ_1^2 et pour $j \in J$, la variance ε_j est σ_2^2. Dans les applications, la variance de σ_1^2 est estimée par $k^{-1}\sum_{i\in I}[Y_i - f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}})]^2$ et σ_2^2 est estimée par $(n-k)^{-1}\sum_{j\in J}[Y_j - f(\mathbf{X}_j, \widehat{\boldsymbol{\beta}}_1)]^2$, où $\widehat{\boldsymbol{\beta}}$ et $\widehat{\boldsymbol{\beta}}_1$ sont donnés par la relation (3.4).

3.3.2 Comportement asymptotique de la statistique $\mathcal{Z}_{nk}$

Dans cette section, on étudie le comportement asymptotique de la statistique $\mathcal{Z}_{nk}$ donnée par la relation (3.15) sous l'hypothèse nulle H_0. On montre également que $\widehat{\boldsymbol{\lambda}}_n(k)$ et $\widehat{\boldsymbol{\beta}}_n(k)$, les solutions de l'equations scores $\boldsymbol{\phi}_{1n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$ et $\boldsymbol{\phi}_{2n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$ données par la relation (3.16) ont des bonnes propriétés.

En plus des suppositions (A1)-(A4), on a besoin les suppositions suivantes :

(A5) Pour chaque $\mathbf{X}_i \in \Upsilon$ et pour tout $\boldsymbol{\beta}$ dans le voisinage de $\boldsymbol{\beta}^0$, les matrices $\mathbf{V}_{1n}(\boldsymbol{\beta})$ et $\mathbf{V}_{2n}(\boldsymbol{\beta})$ sont inversibles et leurs déterminants sont bornés, quand n est assez grand.

(A6) $\sup_{\boldsymbol{\beta}\in\Gamma}\mathbb{E}[\|\dot{\mathbf{f}}(\mathbf{X},\boldsymbol{\beta})]\|]^{2s} < \infty$ et pour tout $1 \leq u, v \leq d$, $\sup_{\boldsymbol{\beta}\in\Gamma}\mathbb{E}[|\frac{\partial^2 f(\mathbf{X},\boldsymbol{\beta})}{\partial\beta_u\partial\beta_v}|^s] < \infty$, avec $s > 2$.

La supposition (A5) assure que les matrices $\mathbf{V}_{1n}(\boldsymbol{\beta})$ et $\mathbf{V}_{2n}(\boldsymbol{\beta})$ sont uniformément inversibles et leurs déterminants sont bornées quand n est assez grand. La supposition (A6) est une condition nécessaire dans l'inférence statistique et elle est également utilisée dans le papier de [Boldea and Hall (2013)].

Par la proposition suivante, on montre que $\widehat{\boldsymbol{\lambda}}_n(k)$ et $\widehat{\boldsymbol{\beta}}_n(k)$, les solutions des equations scores $\boldsymbol{\phi}_{1n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$ et $\boldsymbol{\phi}_{2n}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \mathbf{0}_d$ données par la relation (3.16), ont des bonnes propriétés. Plus précisément, on montre que sous l'hypothèse H_0, $\|\widehat{\boldsymbol{\lambda}}_n(k)\|$ converge en probabilité vers zero, quand n tend vers l'infini, ainsi que $\widehat{\boldsymbol{\beta}}_n(k)$ est un estimateur consistent de $\boldsymbol{\beta}^0$.

Proposition 3.3.1. *Sous l'hypothèse nulle* H_0*, si les suppositions (A1)-(A6) sont vérifiées, alors on a* $\widehat{\boldsymbol{\lambda}}_n(k) = O_{\mathbb{P}}(n^{-1/2})$ *et* $\widehat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta}_0 = o_{\mathbb{P}}(1)$*, où* $(\widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k))$ *est la solution du système (3.16).*

Preuve. Par la définition du rapport de vraisemblance empirique, on a la contrainte suivante

$$\sum_{i\in I} p_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{g}_j(\boldsymbol{\beta}) = \mathbf{0}_d,$$

ce qui donne

$$\sum_{i\in I} p_i \mathbf{z}_i(\boldsymbol{\beta}) = \sum_{j\in J} q_j \mathbf{z}_j(\boldsymbol{\beta}) = \mathbf{0}_d.$$

En utilisant les valeurs de p_i et q_j données respectivement par les relations (3.9) et (3.10) et par des calculs élémentaires similaires à ceux fait pour la relation (2.22), on obtient

$$\sum_{i\in I} p_i \mathbf{z}_i(\boldsymbol{\beta}) = \frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}) - \frac{n}{k^2}\sum_{i\in I}\frac{\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_i^t(\boldsymbol{\beta})}{1+\frac{n}{k}\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{z}_i(\boldsymbol{\beta})}\boldsymbol{\lambda}(\boldsymbol{\beta}) = \mathbf{0}_d \tag{3.17}$$

et

$$\sum_{j\in J} q_j \mathbf{z}_j(\boldsymbol{\beta}) = \frac{1}{n-k}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta}) + \frac{n}{(n-k)^2}\sum_{j\in J}\frac{\mathbf{z}_j(\boldsymbol{\beta})\mathbf{z}_j^t(\boldsymbol{\beta})}{1+\frac{n}{n-k}\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{z}_j(\boldsymbol{\beta})}\boldsymbol{\lambda}(\boldsymbol{\beta}) = \mathbf{0}_d. \tag{3.18}$$

Pour le terme $k^{-1}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})$ de la relation (3.17), par le Lemme 3.6.2, on a que

$$\frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})=\frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}})-\frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})+o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{3.19}$$

Pour le terme $(n/k^2)\sum_{i\in I}\dfrac{\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_i^t(\boldsymbol{\beta})}{1+\frac{n}{k}\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{z}_i(\boldsymbol{\beta})}\boldsymbol{\lambda}(\boldsymbol{\beta})$, en utilisant la Proposition 2.3.1, pour tout $\epsilon>0$, il existe $N_1,N_2>0$, tel que

$$\mathbb{P}\Big[\frac{1}{N_1}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_i^t(\boldsymbol{\beta})\leq\sum_{i\in I}\frac{\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_i^t(\boldsymbol{\beta})}{1+\frac{n}{k}\boldsymbol{\lambda}^t(\boldsymbol{\beta})\mathbf{z}_i(\boldsymbol{\beta})}\leq\frac{1}{N_2}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_i^t(\boldsymbol{\beta})\Big]<\epsilon.$$

Ce qui implique que, afin d'étudier le deuxième terme du membre droit de la relation (3.17), il suffit d'étudier seulement le terme $k^{-1}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_i^t(\boldsymbol{\beta})$.

Par le Lemme 3.6.2, on a que

$$\mathbf{V}_{1n}(\boldsymbol{\beta})=\mathbf{V}_{1n}^0+o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}).$$

Sous la supposition (A5), le développement de Taylor de $\mathbf{z}_i(\boldsymbol{\beta})$ en $\boldsymbol{\beta}=\boldsymbol{\beta}^0$ jusqu'à l'ordre 2 est

$$\begin{aligned}\mathbf{z}_i(\boldsymbol{\beta}) &= (\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i+\frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{M}_{6i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i\\ &\quad-\frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\ &\quad-\frac{1}{6}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{7i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\ &\quad-\frac{1}{4}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{M}_{6i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\ &\quad-\frac{1}{12}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{M}_{6i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{7i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}),\end{aligned} \tag{3.20}$$

où, $\mathbf{M}_{6i}$ et $\mathbf{M}_{7i}$ sont des matrices de dimension $d\times d$, définies respectivement par $\mathbf{M}_{6i}=\big(\frac{\partial^2 f(\boldsymbol{\beta}_{i,rl}^{(7)})}{\partial\beta_r\partial\beta_l}\big)_{1\leq r,l\leq d}$, $\mathbf{M}_{7i}=\big(\frac{\partial^2 f(\boldsymbol{\beta}_{i,rl}^{(8)})}{\partial\beta_r\partial\beta_l}\big)_{1\leq r,l\leq d}$ et $\boldsymbol{\beta}_{i,rl}^{(a)}=\boldsymbol{\beta}^{\mathbf{0}}+v_{i,rl}^{(a)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})$, avec $v_{i,rl}\in[0,1]$ et $a\in\{7,8\}$.

Pour le deuxième terme de (3.17), en utilisant le développement de Taylor de $\mathbf{z}_i(\boldsymbol{\beta})$ dans le voisinage de $\boldsymbol{\beta}^{\mathbf{0}}$ donné par la relation (3.20) et avec un argument similaire à celui

utilisé dans le Lemme 3.6.2 pour prouver le premier terme de la relation (3.17), ainsi les suppositions (A3), (A4), on obtient

$$\frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_i^t(\boldsymbol{\beta}) = \frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta^0})\mathbf{z}_i^t(\boldsymbol{\beta^0}) + \boldsymbol{\Delta}_{1n}^0 + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}), \tag{3.21}$$

où

$$\boldsymbol{\Delta}_{1n}^0 = \frac{n}{k^2}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\big[(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\big]^t.$$

De même, pour les observations $j \in J$, on obtient

$$\frac{1}{n-k}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta}) = \frac{1}{n-k}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta^0}) - \frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) \tag{3.22}$$

et

$$\frac{1}{n-k}\sum_{j\in J}\mathbf{z}_i(\boldsymbol{\beta})\mathbf{z}_j^t(\boldsymbol{\beta}) = \frac{1}{n-k}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta^0})\mathbf{z}_j^t(\boldsymbol{\beta^0}) + \boldsymbol{\Delta}_{2n}^0 + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}),$$

où

$$\boldsymbol{\Delta}_{2n}^0 = \frac{n}{(n-k)^2}\sum_{j\in J}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{2n}^0)^{-1}\dot{\mathbf{f}}_j(\boldsymbol{\beta^0})\dot{\mathbf{f}}_j^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\big[(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{2n}^0)^{-1}\dot{\mathbf{f}}_j(\boldsymbol{\beta^0})\dot{\mathbf{f}}_j^t(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})\big]^t.$$

Pour faciliter l'écriture, on considère les deux matrices suivantes, chacune de dimension $d \times d$

$$\mathbf{S}_n^0 = \frac{n}{k^2}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta^0})\mathbf{z}_i^t(\boldsymbol{\beta^0}) + \frac{n}{(n-k)^2}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta^0})\mathbf{z}_j^t(\boldsymbol{\beta^0}), \tag{3.23}$$

$$\boldsymbol{\Delta}_n^0 = \boldsymbol{\Delta}_{1n}^0 + \boldsymbol{\Delta}_{2n}^0, \tag{3.24}$$

et le vecteur

$$\boldsymbol{\psi}_n^0 = \frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta^0}) - \frac{1}{n-k}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta^0}). \tag{3.25}$$

D'autre part, on a $\boldsymbol{\phi}_{1n}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k)) = \mathbf{0}_d$. En utilisant les relations (3.61) et (3.21)-(3.23), on obtient

$$\begin{aligned}
&\Big[\frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}}) - \frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_n(k)-\boldsymbol{\beta}^{\mathbf{0}}) - \boldsymbol{\Delta}_{1n}^0\widehat{\boldsymbol{\lambda}}_n(k) - \frac{n}{k^2}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{z}_i^t(\boldsymbol{\beta}^{\mathbf{0}})\widehat{\boldsymbol{\lambda}}_n(k)\Big]\\
&-\Big[\frac{1}{k}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta}^{\mathbf{0}}) - \frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_n(k)-\boldsymbol{\beta}^{\mathbf{0}}) + \boldsymbol{\Delta}_{2n}^0\widehat{\boldsymbol{\lambda}}_n(k) + \frac{n}{(n-k)^2}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{z}_j^t(\boldsymbol{\beta}^{\mathbf{0}})\widehat{\boldsymbol{\lambda}}_n(k)\Big]\\
&=\mathbf{0}_d.
\end{aligned}$$

En utilisant les notations données par les relations (3.23), (3.24) et (3.25), alors, on obtient

$$\widehat{\boldsymbol{\lambda}}_n(k) = (\mathbf{S}_n^0 + \boldsymbol{\Delta}_n^0)^{-1}\boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(\widehat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta}^{\mathbf{0}}). \tag{3.26}$$

Le développement limité de la statistique $\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k))$, spécifiée par la relation (3.15), dans le voisinage de $(\boldsymbol{\lambda}, \boldsymbol{\beta}) = (\mathbf{0}_d, \boldsymbol{\beta}^{\mathbf{0}})$ jusqu'à l'ordre 2, peut s'écrire

$$\begin{aligned}
&\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k)) = \Big[2\widehat{\boldsymbol{\lambda}}_n^t(k)\Big(\frac{n}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}}) - \frac{n}{n-k}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta}^{\mathbf{0}})\Big)\Big]\\
&-\Big[\widehat{\boldsymbol{\lambda}}_n^t(k)\Big(\frac{n^2}{k^2}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_i^t(\boldsymbol{\beta}^{\mathbf{0}}) + \frac{n^2}{(n-k)^2}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{g}_j^t(\boldsymbol{\beta}^{\mathbf{0}})\Big)\widehat{\boldsymbol{\lambda}}_n(k)\Big]\\
&+\Big[2\widehat{\boldsymbol{\lambda}}_n^t(k)\Big((\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\frac{n}{k}\sum_{i\in I}\dot{\mathbf{g}}_i(\boldsymbol{\beta}^{\mathbf{0}}) - (\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{2n}^0)^{-1}\frac{n}{n-k}\sum_{j\in J}\dot{\mathbf{g}}_j(\boldsymbol{\beta}^{\mathbf{0}})\Big)(\widehat{\boldsymbol{\beta}}_n(k)-\boldsymbol{\beta}^{\mathbf{0}})\Big]\\
&-\Big[2\widehat{\boldsymbol{\lambda}}_n^t(k)\Big(\frac{n}{k}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta}^{\mathbf{0}})\frac{\partial(\mathbf{M}_n^{\frac{1}{2}}(\boldsymbol{\beta})(\mathbf{V}_{1n}(\boldsymbol{\beta}))^{-1})}{\partial\boldsymbol{\beta}} + \frac{n}{n-k}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta}^{\mathbf{0}})\frac{\partial(\mathbf{M}_n^{\frac{1}{2}}(\boldsymbol{\beta})(\mathbf{V}_{2n}(\boldsymbol{\beta}))^{-1})}{\partial\boldsymbol{\beta}}\Big)\\
&\cdot(\widehat{\boldsymbol{\beta}}_n(k)-\boldsymbol{\beta}^{\mathbf{0}})\Big] + \frac{1}{3!}\Big[T_1 + 3T_2 + 3T_3 + T_4\Big]\\
&\equiv B_1 - B_2 + B_3 - B_4 + B_5,
\end{aligned} \tag{3.27}$$

où

$T_1 = \sum_{r=1}^d\sum_{l=1}^d\sum_{s=1}^d \frac{\partial^3\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_{rsl}^{(1)}, \boldsymbol{\beta}_{rsl}^{(1)})}{\partial\beta_r\partial\beta_s\partial\beta_l}(\widehat{\beta}_{n,r}-\beta_r^0)(\widehat{\beta}_{n,s}-\beta_s^0)(\widehat{\beta}_{n,l}-\beta_l^0)$,

$T_2 = \sum_{r=1}^d\sum_{l=1}^d\sum_{s=1}^d \frac{\partial^3\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_{rsl}^{(2)}, \boldsymbol{\beta}_{rsl}^{(2)})}{\partial\lambda_r\partial\lambda_s\partial\beta_l}(\widehat{\lambda}_{n,r})(\widehat{\lambda}_{n,s})(\widehat{\beta}_{n,l}-\beta_l^0)$,

$T_3 = \sum_{r=1}^{d}\sum_{l=1}^{d}\sum_{s=1}^{d} \frac{\partial^3 \mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_{rsl}^{(3)}, \boldsymbol{\beta}_{rsl}^{(3)})}{\partial\lambda_r \partial\beta_s \partial\beta_l}(\widehat{\lambda}_{n,r})(\widehat{\beta}_{n,s} - \beta_s^0)(\widehat{\beta}_{n,l} - \beta_l^0)$,

$T_4 = \sum_{r=1}^{d}\sum_{l=1}^{d}\sum_{s=1}^{d} \frac{\partial^3 \mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_{rsl}^{(4)}, \boldsymbol{\beta}_{rsl}^{(4)})}{\partial\lambda_r \partial\lambda_s \partial\lambda_l}(\widehat{\lambda}_{n,r})(\widehat{\lambda}_{n,s})(\widehat{\lambda}_{n,l})$.

Pour $1 \leq r \leq d$, $\widehat{\beta}_{n,r}$ désigne la r-ème composante du vecteur $\widehat{\boldsymbol{\beta}}_n(k)$ et $\widehat{\lambda}_{n,r}$ désigne la r-ème composante du $\widehat{\boldsymbol{\lambda}}_n(k)$.
Pour tout $1 \leq r, s, l \leq d$, on a $\boldsymbol{\lambda}_{rsl}^{(b)} = u_{rsl}^{(b)}(\widehat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta^0})$ et $\boldsymbol{\beta}_{rsl}^{(b)} = \boldsymbol{\beta^0} + v_{rsl}^{(b)}(\widehat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta^0})$, avec $u_{rsl}^{(b)}, v_{rsl}^{(b)} \in [0,1]$ et $b \in \{1,2,3,4\}$.

Notons que, les dérivées de $\mathbf{M}_n^{\frac{1}{2}}(\boldsymbol{\beta})(\mathbf{V}_{1n}(\boldsymbol{\beta}))^{-1}$ et $\mathbf{M}_n^{\frac{1}{2}}(\boldsymbol{\beta})(\mathbf{V}_{2n}(\boldsymbol{\beta}))^{-1}$ par rapport à $\boldsymbol{\beta}$ sont considérées terme par terme.

Maintenant, on remplace $\widehat{\boldsymbol{\lambda}}_n(k)$ dans la relation (3.27) par sa valeur obtenue dans (3.26). Pour le terme B_1 de la relation (3.27), en utilisant les notations données par les relations (3.23), (3.24) et (3.25), on obtient

$$B_1 = 2(n\boldsymbol{\psi}_n^0)^t(\mathbf{S}_n^0 + \boldsymbol{\Delta}_n^0)^{-1}\boldsymbol{\psi}_n^0.$$

De même, pour le terme B_2 de la relation (3.27), on obtient

$$B_2 = n(\boldsymbol{\psi}_n^0)^t(\mathbf{S}_n^0 + \boldsymbol{\Delta}_n^0)^{-1}\boldsymbol{\psi}_n^0.$$

On sait que, $\mathbf{V}_{1n}^0 = k^{-1}\sum_{i\in I}\dot{\mathbf{g}}_i(\boldsymbol{\beta^0})$ et $\mathbf{V}_{2n}^0 = (n-k)^{-1}\sum_{j\in J}\dot{\mathbf{g}}_j(\boldsymbol{\beta^0})$. Alors, pour le terme B_3 de la relation (3.27), on obtient

$$B_3 \xrightarrow{p.s.} 0.$$

Par le Théorème Central Limite, on a que

$$\frac{1}{k}\sum_{i\in I}\mathbf{g}_i(\boldsymbol{\beta^0}) = O_{\mathbb{P}}(k^{-1/2})$$

et

$$\frac{1}{n-k}\sum_{j\in J}\mathbf{g}_j(\boldsymbol{\beta^0}) = O_{\mathbb{P}}((n-k)^{-1/2}).$$

Alors, pour le terme B_4 de la relation (3.27) on obtient

$$B_4 = o_{\mathbb{P}}(n(\boldsymbol{\psi}_n^0)^t(\mathbf{S}_n^0 + \boldsymbol{\Delta}_n^0)^{-1}\boldsymbol{\psi}_n^0).$$

Pour le terme B_5 de la relation (3.27), en utilisant les suppositions (A2)-(A4) et par des calculs élémentaires, on obtient que ce terme est égal à $\xi_n(k)$, où

$$\xi_n(k) \equiv o_{\mathbb{P}}(\|\widehat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta}^{\mathbf{0}}\|) + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\lambda}}_n(k)\|) + o_{\mathbb{P}}(\|\widehat{\boldsymbol{\lambda}}_n(k)\|\|\widehat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta}^{\mathbf{0}}\|). \tag{3.28}$$

En combinant les résultats obtenus,

$$\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k)) = n(\boldsymbol{\psi}_n^0)^t(\mathbf{S}_n^0 + \boldsymbol{\Delta}_n^0)^{-1}\boldsymbol{\psi}_n^0(1 + o_{\mathbb{P}}(1)) + \xi_n(k). \tag{3.29}$$

On peut voir que, quand $\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathbf{0}} = \mathbf{0}_d$, le terme $n(\boldsymbol{\psi}_n^0)^t(\mathbf{S}_n^0 + \boldsymbol{\Delta}_n^0)^{-1}\boldsymbol{\psi}_n^0$ atteint son maximum dans le voisinage de $\boldsymbol{\beta}^{\mathbf{0}}$. Cela signifie que quand $n \to \infty$, $\mathcal{Z}_{nk}(\boldsymbol{\lambda}, \boldsymbol{\gamma}_0, \boldsymbol{\beta})$ a un seul maximum local dans tout le η-voisinage de $\boldsymbol{\beta}^{\mathbf{0}}$ et alors $\widehat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta}_0 = o_{\mathbb{P}}(1)$.

Pour $\mathbf{S}_n^0$, par la loi des grands nombres, quand $n \to \infty$, on a que

$$\frac{n}{k^2}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{z}_i^t(\boldsymbol{\beta}^{\mathbf{0}}) \xrightarrow{p.s.} \mathbb{E}[\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{z}_i^t(\boldsymbol{\beta}^{\mathbf{0}})] \tag{3.30}$$

et

$$\frac{n}{(n-k)^2}\sum_{j\in J}\mathbf{z}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{z}_j^t(\boldsymbol{\beta}^{\mathbf{0}}) \xrightarrow{p.s.} \mathbb{E}[\mathbf{z}_j(\boldsymbol{\beta}^{\mathbf{0}})\mathbf{z}_j^t(\boldsymbol{\beta}^{\mathbf{0}})]. \tag{3.31}$$

En utilisant les relations (3.30) et (3.31) ainsi le fait que par la loi des grands nombres les deux matrices $-\mathbf{V}_{1n}^0$ et $-\mathbf{V}_{2n}^0$ convergent presque sûrement vers $\mathbf{V}$, alors, on obtient que $\mathbf{S}_n^0 = \mathbf{I}_d + O_{\mathbb{P}}(n^{-1/2})$. En utilisant ce fait, ainsi la relation (3.26), on obtient que

$$\widehat{\boldsymbol{\lambda}}_n(k) = \boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(n^{-1/2}).$$

Rappelons que, pour tout $i = 1, \ldots, n$ on a $\mathbb{E}[g_i(\boldsymbol{\beta}^{\mathbf{0}})] = 0$. Alors, par le Théorème Central Limite, chaque terme de $\boldsymbol{\psi}_n^0$ est $O_{\mathbb{P}}(n^{-1/2})$, ce qui implique que $\widehat{\boldsymbol{\lambda}}_n(k) = O_{\mathbb{P}}(n^{-1/2})$. ■

Le résultat suivant est une généralisation de la version non paramétrique du théorème de Wilks pour le rapport de vraisemblance empirique défini par la relation (3.15).

Théorème 3.3.1. *Supposons que (A1)-(A6) sont vérifiées. Sous l'hypothèse nulle* H_0*, la statistique* $\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k))$ *donnée par la relation (3.15) converge vers une loi de chi-deux de degré de liberté d, quand* $n \to \infty$*, où d désigne la dimension de* $\boldsymbol{\beta}$*.*

Preuve. Par la preuve du Proposition 3.3.1, on a

$$\widehat{\boldsymbol{\lambda}}_n(k) = \boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(n^{-1/2})$$

et

$$\mathbf{S}_n^0 = \mathbf{I}_d + O_{\mathbb{P}}(n^{-1/2}).$$

Par le Théorème de Linderberg-Feller, on a que

$$\sqrt{n}\widehat{\boldsymbol{\lambda}}_n(k) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \mathbf{I}_d).$$

Alors, en utilisant la relation (3.26), on obtient

$$\begin{aligned}\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k)) &= n(\boldsymbol{\psi}_n^0)^t(\mathbf{S}_n^0)^{-1}\boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(1) + \xi_n(k)\\ &= n\widehat{\boldsymbol{\lambda}}_n^t(k)\widehat{\boldsymbol{\lambda}}_n(k) + o_{\mathbb{P}}(1) + \xi_n(k)\\ &= \sqrt{n}\widehat{\boldsymbol{\lambda}}_n^t(k)\widehat{\boldsymbol{\lambda}}_n(k)\sqrt{n} + o_{\mathbb{P}}(1) + \xi_n(k),\end{aligned}$$

où $\xi_n(k)$ est donné par la relation (3.28). Alors, puisque $\sqrt{n}\widehat{\boldsymbol{\lambda}}_n(k) \sim \mathcal{N}(0, \mathbf{I}_d)$, on déduit que

$$\mathcal{Z}_{nk}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k)) \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2(d).$$

■

Par le Théorème 3.3.1, pour un niveau fixé α, tel que $\alpha \in (0, 1)$, on peut construire pour $\boldsymbol{\gamma}$ la région de confiance asymptotique suivante :

$$CR_\alpha = \{\boldsymbol{\gamma} \in \mathbb{R}^d : \mathcal{Z}_{nk}(\boldsymbol{\gamma}, \widehat{\boldsymbol{\lambda}}_n, \widehat{\boldsymbol{\beta}}_n) < c_{1-\alpha;d}\},$$

où $c_{1-\alpha;d}$ est le $(1-\alpha)$ quantile de la distribution de chi-deux avec d degré de liberté.

3.4 Modèle avec des données manquantes

Dans cette section, pour le modèle (3.1), on suppose que toutes les $\mathbf{X}_i$ sont observées, par contre la variable réponse Y_i peut être manquante. Soient la suite des variables aléatoires $(\delta_i)_{1\le i\le n}$ définie par $\delta_i = 0$ si Y_i est manquante et $\delta_i = 1$ si Y_i est observée.
On suppose que Y_i est manquante au hasard (missing at random (MAR)), c'est-à-dire pour tout $1 \le i \le n$

$$\mathbb{P}[\delta_i = 1|\mathbf{X}_i, Y_i] = \mathbb{P}[\delta_i = 1|\mathbf{X}_i].$$

On considère les fonctions de probabilités sélectives définies par

$$\begin{cases} \pi_1(\mathbf{x}_i) = \mathbb{P}[\delta_i = 1|\mathbf{X}_i = \mathbf{x}_i], & i \in I,\\ \pi_2(\mathbf{x}_j) = \mathbb{P}[\delta_j = 1|\mathbf{X}_j = \mathbf{x}_j], & j \in J.\end{cases}$$

Pour $i = 1, \ldots, n$, $\pi_1(\mathbf{x}_i)$ désigne la probabilité que la variable Y_i est observée. Nous supposons que $\pi_1(\mathbf{x}_i) > 0$ et $\pi_2(\mathbf{x}_j) > 0$. Ces suppositions sont classiques dans la littérature, voir par example les papiers de [Sun et al. (2009)], [Xue (2009)] et [Ciuperca (2013a)]. De plus, $\lim_{n\to\infty} \sum_{i\in I} \delta_i = \lim_{n\to\infty} \sum_{j\in J} \delta_j = \infty$, ce qui signifie que nous avons un nombre important de valeurs non manquantes pour Y.

La littérature des données manquantes a été apparu dans les analyses statistique depuis le début de l'année 1970 stimulée par les progrès de la technologie de l'informatique qui sert à faciliter les calculs numériques. Dans la pratique, les variables réponses peuvent être manquantes grâce à plusieurs raisons telles que la réticence de certains unités échantillonnées à fournir des informations souhaitées, la perte de l'information à cause des facteurs incontrôlables, l'échec de l'enquêteur à recueillir des informations correctes et ainsi de suite... Les réponses manquantes sont très communes dans les sondages d'opinion, les enquêtes d'études de marché et également dans des nombreuses expériences scientifiques. La manque de certaines réponses rendent les méthodes existantes dans la littérature non-applicables. Pour des travaux concernant des données manquantes on cite les papiers de [Little and Rubin (1983)], [Little and Rubin (2002)], [Wang and Rao (2002)], [Celeux et al. (2006a)], [Celeux et al. (2006b)], [Wang and Chen (2009)], [Xue (2009)], [Xu et al. (2012)].

Le modèle non-linéaire avec données manquantes au hasard (MAR), a été considéré dans la littérature. [Muller (2009)] a construit un estimateur efficace pour l'espérance $\mathbb{E}[h(\mathbf{X}, \mathbf{Y})]$ en utilisant un estimateur efficace de paramètres, tel que h désigne une fonction carré intégrable et connue. Le moyenne du réponse $\mathbb{E}(\mathbf{Y})$ désigne un cas spécial. [Ciuperca (2013a)] a construit le rapport de la vraisemblance empirique en utilisant la méthode d'imputation des données. L'idée principale de l'imputation est de "remplir" les valeurs manquantes de Y par des valeurs "appropriées" pour créer un ensemble des données complètes, ce qui permet d'appliquer les méthodes classiques d'étude.

Dans la suite, une méthode de vraisemblance empirique est utilisée pour étudier le modèle (3.1) en supposant qu'il y a des variables réponse manquantes. Nous sommes intéressés à construire la région de confiance asymptotique pour $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}_1$, basée sur les données $(\mathbf{X}_i, Y_i, \delta_i)_{1\leq i\leq n}$. Dans le même temps, on teste l'hypothèse nulle

$$\mathrm{H}_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0.$$

Nous rappelons que, sous l'hypothèse nulle H_0, le paramètre $\boldsymbol{\beta}^{\mathbf{0}}$ désigne la vraie valeur (inconnue) de $\boldsymbol{\beta}$, où $\boldsymbol{\beta}$ est la valeur générique du paramètre de régression de la première

phase.

3.4.1 Statistique de test

Pour construire la région de confiance asymptotique de $\boldsymbol{\gamma}$, on propose trois statistiques de vraisemblance empirique : vraisemblance empirique basée sur les données complètes, vraisemblance empirique pondérée et la vraisemblance empirique avec des valeurs imputées.

Vraisemblance empirique basée sur le cas complet

Tout d'abord, on donne la vraisemblance empirique basée sur le cas complet. Dans le contexte de régression, cela sert à se restreinte aux individus pour lesquels toutes les variables sont renseignées.

Pour les observations de la variable réponse de la seconde phase, on considère une nouvelle variable, notée $\tilde{Y}_j$ définie d'un façon similaire au cas des données non-manquantes. Plus précisément, pour le modèle (3.1), pour $j \in J$, si Y_j est non-manquante, $\tilde{Y}_j \equiv Y_j - f(\mathbf{X}_j, \boldsymbol{\beta} - \boldsymbol{\gamma}^0) + f(\mathbf{X}_j, \boldsymbol{\beta})$. D'autre part, si Y_j est manquante alors $\tilde{Y}_j$ est égale à une valeur finie. On définit les deux vecteurs aléatoires suivants

$$\begin{cases} \mathbf{g}_{i,C}(\boldsymbol{\beta}) \equiv \delta_i \dot{\mathbf{f}}_i(\boldsymbol{\beta})(Y_i - f_i(\boldsymbol{\beta})), & i \in I, \\ \mathbf{g}_{j,C}(\boldsymbol{\beta}) \equiv \delta_j \dot{\mathbf{f}}_j(\boldsymbol{\beta})(\tilde{Y}_j - f_j(\boldsymbol{\beta})), & j \in J. \end{cases} \tag{3.32}$$

On considère également les deux matrices suivantes

$$\mathbf{V}_{1n,C}(\boldsymbol{\beta}) \equiv \frac{1}{k} \sum_{i \in I} \dot{\mathbf{g}}_{i,C}(\boldsymbol{\beta}), \qquad \mathbf{V}_{2n,C}(\boldsymbol{\beta}) \equiv \frac{1}{n-k} \sum_{j \in J} \dot{\mathbf{g}}_{j,C}(\boldsymbol{\beta}).$$

D'une manière similaire à la Section 3.2 et par le même argument de la Remarque 3.3.1, afin de construire la région de confiance asymptotique pour $\boldsymbol{\gamma}$ ou tester l'hypothèse nulle H_0, la statistique du rapport de vraisemblance empirique pour le cas complet est

$$\mathcal{Z}_{nk,C}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_C, \boldsymbol{\beta}) \equiv 2\Big[\sum_{i \in I} \log(1 + \frac{n}{k} \boldsymbol{\lambda}_C^t \mathbf{z}_{i,C}(\boldsymbol{\beta})) + \sum_{j \in J} \log(1 - \frac{n}{n-k} \boldsymbol{\lambda}_C^t \mathbf{z}_{j,C}(\boldsymbol{\beta}) \Big], \tag{3.33}$$

avec

$$\mathbf{z}_{i,C}(\boldsymbol{\beta}) \equiv \mathbf{M}_{n,C}^{\frac{1}{2}}(\boldsymbol{\beta}) \mathbf{V}_{1n,C}^{-1}(\boldsymbol{\beta}) \mathbf{g}_{i,C}(\boldsymbol{\beta}), \qquad \mathbf{z}_{j,C}(\boldsymbol{\beta}) \equiv \mathbf{M}_{n,C}^{\frac{1}{2}}(\boldsymbol{\beta}) \mathbf{V}_{2n,C}^{-1}(\boldsymbol{\beta}) \mathbf{g}_{j,C}(\boldsymbol{\beta}),$$

$$\mathbf{M}_{n,C}(\boldsymbol{\beta}) \equiv \frac{k(n-k)}{n^2}\mathbf{V}_{1n,C}(\boldsymbol{\beta})\mathbf{H}_{n,C}(\boldsymbol{\beta})\mathbf{V}_{2n,C}(\boldsymbol{\beta})$$

et

$$\begin{aligned}\mathbf{H}_{n,C}(\boldsymbol{\beta}) \equiv & \Big[\frac{n-k}{nk}\sigma_1^2\mathbf{V}_{2n,C}(\boldsymbol{\beta})\mathbf{V}_{1n,C}^{-1}(\boldsymbol{\beta})\sum_{i\in I}\delta_i\pi_1(\mathbf{X}_i)\big[\ddot{\mathbf{f}}_i(\boldsymbol{\beta})(Y_i-f_i(\boldsymbol{\beta}))-\dot{\mathbf{f}}_i(\boldsymbol{\beta})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta})\big] \\ & +\frac{k}{n(n-k)}\sigma_2^2\sum_{j\in J}\delta_j\pi_2(\mathbf{X}_j)\big[\ddot{\mathbf{f}}_j(\boldsymbol{\beta})(\tilde{Y}_j-f_j(\boldsymbol{\beta}))-\dot{\mathbf{f}}_j(\boldsymbol{\beta})\dot{\mathbf{f}}_j^t(\boldsymbol{\beta})\big]\mathbf{V}_{2n,C}^{-1}(\boldsymbol{\beta})\mathbf{V}_{1n,C}(\boldsymbol{\beta})\Big]^{-1}.\end{aligned}$$

Nous rappelons que, σ_1^2 et σ_2^2 désignent respectivement la variance de ε_i pour $i \in I$ et ε_j pour $j \in J$.

Pour le modèle avec des données manquantes, sous les suppositions $\lim_{n\to\infty}\sum_{i\in I}\delta_i \to \infty$ et $\lim_{n\to\infty}\sum_{j\in J}\delta_j \to \infty$, on peut estimer σ_1^2 et σ_2^2 respectivement par $\tilde{\sigma}_1^2$ et $\tilde{\sigma}_2^2$, définis par

$$\tilde{\sigma}_1^2 = \frac{\sum\limits_{i\in I}\delta_i[Y_i - f(\mathbf{X}_i,\widehat{\boldsymbol{\beta}})]^2}{\sum\limits_{i\in I}\delta_i}, \qquad \tilde{\sigma}_2^2 = \frac{\sum\limits_{j\in J}\delta_j[Y_j - f(\mathbf{X}_j,\widehat{\boldsymbol{\beta}}_1)]^2}{\sum\limits_{j\in J}\delta_j}, \tag{3.34}$$

avec

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\sum_{i\in I}\delta_i(Y_i - f(\mathbf{X}_i,\boldsymbol{\beta}))^2, \quad \widehat{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{\beta}_1}\sum_{j\in J}\delta_j(Y_j - f(\mathbf{X}_j,\boldsymbol{\beta}_1))^2. \tag{3.35}$$

Les fonctions scores correspondantes à la statistique de test donnée par la relation (3.33) sont

$$\begin{cases} \boldsymbol{\phi}_{1n,C}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_C,\boldsymbol{\beta}) = \sum\limits_{i\in I}\dfrac{\mathbf{z}_{i,C}(\boldsymbol{\beta})}{\frac{k}{n}+\boldsymbol{\lambda}_C^t\mathbf{z}_{i,C}(\boldsymbol{\beta})} - \sum\limits_{j\in J}\dfrac{\mathbf{z}_{j,C}(\boldsymbol{\beta})}{\frac{n-k}{n}-\boldsymbol{\lambda}_C^t\mathbf{z}_{j,C}(\boldsymbol{\beta})}, \\ \boldsymbol{\phi}_{2n,C}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_C,\boldsymbol{\beta}) = \sum\limits_{i\in I}\dfrac{\dot{\mathbf{z}}_{i,C}(\boldsymbol{\beta})}{\frac{k}{n}+\boldsymbol{\lambda}_C^t\mathbf{z}_{i,C}(\boldsymbol{\beta})} - \sum\limits_{j\in J}\dfrac{\dot{\mathbf{z}}_{j,C}(\boldsymbol{\beta})}{\frac{n-k}{n}-\boldsymbol{\lambda}_C^t\mathbf{z}_{j,C}(\boldsymbol{\beta})}, \end{cases} \tag{3.36}$$

où $\dot{\mathbf{z}}_{i,C}(\boldsymbol{\beta})$ et $\dot{\mathbf{z}}_{j,C}(\boldsymbol{\beta})$ sont respectivement les dérivées de $\mathbf{z}_{i,C}(\boldsymbol{\beta})$ et $\mathbf{z}_{j,C}(\boldsymbol{\beta})$ par rapport à $\boldsymbol{\beta}$.

Alors, en résolvant le système $\boldsymbol{\phi}_{1n,C}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_C,\boldsymbol{\beta}) = \mathbf{0}_d$ et $\boldsymbol{\phi}_{2n,C}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_C,\boldsymbol{\beta}) = \mathbf{0}_d$ donné par (3.36), on obtient les valeurs de $\widehat{\boldsymbol{\lambda}}_{n,C}(k)$ et $\widehat{\boldsymbol{\beta}}_{n,C}(k)$ qui maximisent la statistique du (3.33).

On considère les probabilités sélectives $\pi_1(\mathbf{X}_i)$ et $\pi_2(\mathbf{X}_j)$ comme connues. Dans le cas ou elles sont inconnues, on peut considérer pour $\pi_1(\mathbf{X}_i)$ et $\pi_2(\mathbf{X}_j)$, respectivement les estimateurs non-paramétriques à noyau $\widehat{\pi}_1(\mathbf{X}_i)$ et $\widehat{\pi}_2(\mathbf{X}_j)$ donnés par

$$\begin{aligned}\widehat{\pi}_1(\mathbf{X}_i) &= \frac{\sum_{l\in I}\delta_l K_1((\mathbf{X}_l-\mathbf{X}_i)/h_{1n})}{\max\{1,\sum_{l\in I}K_1((\mathbf{X}_l-\mathbf{X}_i)/h_{1n})\}}, \qquad i\in I,\\ \widehat{\pi}_2(\mathbf{X}_j) &= \frac{\sum_{l\in J}\delta_l K_2((\mathbf{X}_l-\mathbf{X}_j)/h_{2n})}{\max\{1,\sum_{l\in J}K_2((\mathbf{X}_l-\mathbf{X}_j)/h_{2n})\}}, \qquad j\in J.\end{aligned} \tag{3.37}$$

Ici, h_{1n} et h_{2n} sont des suites positives qui tendent vers zero quand $n\to\infty$. K_1, K_2 sont des fonctions noyaux définies dans $\mathbb{R}^d$. L'estimateur à noyau est aussi utilisé par [Dabo-Niang et al. (2015)] pour estimer une régression fonctionnelle.

Les fenêtres h_{1n} et h_{2n} satisfont les conditions suivantes :

(A7) $kh_{1n}^{4\max\{2,d-1\}}\to 0$ et $(n-k)h_{2n}^{4\max\{2,d-1\}}\to 0$, quand $n\to\infty$.

Les fonctions noyaux K_1 et K_2 satisfont la condition classique suivante :

(A8) Il existe deux constantes positives $C_1, C_2, C_3, C_4, \rho_1$ et ρ_2, tel que, pour chaque vecteur $\mathbf{v}$, $C_1\mathbb{1}_{\|\mathbf{v}\|\le\rho_1}\le K_1(\mathbf{v})\le C_2\mathbb{1}_{\|\mathbf{v}\|\le\rho_1}$ et $C_3\mathbb{1}_{\|\mathbf{v}\|\le\rho_1}\le K_2(\mathbf{v})\le C_4\mathbb{1}_{\|\mathbf{v}\|\le\rho_2}$.

La condition (A8) est également imposée dans les papiers de [Wei et al. (2013)] et [Xue (2009)], où les modèles linéaires avec et sans point de rupture sont respectivement considérés.

Concernant les fonctions de probabilités sélectives $\pi_1(\mathbf{x})$ et $\pi_2(\mathbf{x})$, soit la supposition de régularité suivante :

(A9) $\pi_1(\mathbf{x})$ et $\pi_2(\mathbf{x})$ ont des dérivées partielles bornées, par rapport à $\mathbf{x}$, jusqu'à l'ordre $\max(2, d-1)$, presque partout.

Les conditions (A7)-(A9) sont des suppositions usuelles pour prouver le taux de convergence de l'estimateur non-paramétrique à noyau. Voir par exemple le papier de [Wei et al. (2013)].

Vraisemblance empirique pondérée

Maintenant, nous présentons la méthode de vraisemblance empirique pondérée. En générale, les analyses basés sur le cas-complet ne fournit pas des estimateurs très efficaces. Une approche alternative basée sur l'inverse de la probabilité pondérée a été considéré. La pondération est une méthode corrigeant l'importance de chacune des observations lors d'une analyse. Dans le cas de l'analyse d'un échantillon avec données manquantes, la pondération peut être utilisée pour donner plus d'importance aux valeurs observées en fonction de certaines caractéristiques associées à l'absence des données. Ces poids peuvent être calculés par l'inverse de la probabilité qu'une donnée soit manquante. L'utilisation de ces poids est fondée sur la théorie du score de propension développé par [Rosenbaum and Rubin (1983)] et a été considéré par [Horvitz and Thompson (1952)].

En utilisant cette même approche, on définit les deux vecteurs aléatoires suivants

$$\begin{cases} \mathbf{g}_{i,W}(\boldsymbol{\beta}) \equiv \dfrac{\delta_i}{\pi_1(\mathbf{X}_i)}\dot{\mathbf{f}}_i(\boldsymbol{\beta})(Y_i - f_i(\boldsymbol{\beta})), & i \in I, \\ \mathbf{g}_{j,W}(\boldsymbol{\beta}) \equiv \dfrac{\delta_j}{\pi_2(\mathbf{X}_j)}\dot{\mathbf{f}}_j(\boldsymbol{\beta})(\tilde{Y}_j - f_j(\boldsymbol{\beta})), & j \in J. \end{cases} \tag{3.38}$$

Nous rappelons que, pour $j \in J$, on prend $\tilde{Y}_j \equiv Y_j - f(\mathbf{X}_j, \boldsymbol{\beta} - \boldsymbol{\gamma}^0) + f(\mathbf{X}_j, \boldsymbol{\beta})$ si Y_j est observée et $\tilde{Y}_j$ égale à une valeur finie si Y_j est manquante.

Dans ce cas, on considère les deux matrices suivantes

$$\mathbf{V}_{1n,W}(\boldsymbol{\beta}) \equiv \frac{1}{k}\sum_{i\in I}\dot{\mathbf{g}}_{i,W}(\boldsymbol{\beta}), \qquad \mathbf{V}_{2n,W}(\boldsymbol{\beta}) \equiv \frac{1}{n-k}\sum_{j\in J}\dot{\mathbf{g}}_{j,W}(\boldsymbol{\beta}).$$

De la même façon que pour les données complètes, en utilisant un argument similaire de la Remarque 3.3.1, la statistique de test pour la méthode pondérée est

$$\mathcal{Z}_{nk,W}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_W, \boldsymbol{\beta}) \equiv 2\Big[\sum_{i\in I}\log(1 + \frac{k}{n}\boldsymbol{\lambda}_W^t\mathbf{z}_{i,W}(\boldsymbol{\beta})) + \sum_{j\in J}\log(1 - \frac{n}{n-k}\boldsymbol{\lambda}_W^t\mathbf{z}_{j,W}(\boldsymbol{\beta})\Big], \tag{3.39}$$

où

$$\mathbf{z}_{i,W}(\boldsymbol{\beta}) \equiv \mathbf{M}_{n,W}^{\frac{1}{2}}(\boldsymbol{\beta})\mathbf{V}_{1n,W}^{-1}(\boldsymbol{\beta})\mathbf{g}_{i,W}(\boldsymbol{\beta}), \qquad \mathbf{z}_{j,W}(\boldsymbol{\beta}) \equiv \mathbf{M}_{n,W}^{\frac{1}{2}}(\boldsymbol{\beta})\mathbf{V}_{2n,W}^{-1}(\boldsymbol{\beta})\mathbf{g}_{j,W}(\boldsymbol{\beta}),$$

$$\mathbf{M}_{n,W}(\boldsymbol{\beta}) \equiv \frac{k(n-k)}{n^2}\mathbf{V}_{1n,W}(\boldsymbol{\beta})\mathbf{H}_{n,W}(\boldsymbol{\beta})\mathbf{V}_{2n,W}(\boldsymbol{\beta})$$

et

$$\mathbf{H}_{n,W}(\boldsymbol{\beta}) \equiv \Big[\frac{n-k}{nk}\sigma_1^2\mathbf{V}_{2n,W}(\boldsymbol{\beta}) + \frac{k}{n(n-k)}\sigma_2^2\mathbf{V}_{1n,W}(\boldsymbol{\beta})\Big]^{-1}.$$

Les variances σ_1^2 et σ_2^2 peuvent être estimées respectivement par $\tilde{\sigma}_1^2$ et $\tilde{\sigma}_2^2$, données par (3.34).

Les fonctions scores de la statistique (3.39) sont

$$\begin{cases} \boldsymbol{\phi}_{1n,W}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_W, \boldsymbol{\beta}) = \displaystyle\sum_{i\in I} \frac{\mathbf{z}_{i,W}(\boldsymbol{\beta})}{\frac{k}{n} + \boldsymbol{\lambda}_W^t\mathbf{z}_{i,W}(\boldsymbol{\beta})} - \sum_{j\in J} \frac{\mathbf{z}_{j,W}(\boldsymbol{\beta})}{\frac{n-k}{n} - \boldsymbol{\lambda}_W^t\mathbf{z}_{j,W}(\boldsymbol{\beta})}, \\ \boldsymbol{\phi}_{2n,W}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_W, \boldsymbol{\beta}) = \displaystyle\sum_{i\in I} \frac{\dot{\mathbf{z}}_{i,W}(\boldsymbol{\beta})}{\frac{k}{n} + \boldsymbol{\lambda}_W^t\mathbf{z}_{i,W}(\boldsymbol{\beta})} - \sum_{j\in J} \frac{\dot{\mathbf{z}}_{j,W}(\boldsymbol{\beta})}{\frac{n-k}{n} - \boldsymbol{\lambda}_W^t\mathbf{z}_{j,W}(\boldsymbol{\beta})}, \end{cases} \tag{3.40}$$

où $\dot{\mathbf{z}}_{i,W}(\boldsymbol{\beta})$ et $\dot{\mathbf{z}}_{j,W}(\boldsymbol{\beta})$ sont respectivement les dérivées de $\mathbf{z}_{i,W}(\boldsymbol{\beta})$ et $\mathbf{z}_{j,W}(\boldsymbol{\beta})$ par rapport à $\boldsymbol{\beta}$.

Alors, en résolvant le système $\boldsymbol{\phi}_{1n,W}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_W, \boldsymbol{\beta}) = \mathbf{0}_d$ et $\boldsymbol{\phi}_{2n,W}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_W, \boldsymbol{\beta}) = \mathbf{0}_d$ donné par la relation (3.40), on obtient $\widehat{\boldsymbol{\lambda}}_{n,W}(k)$ et $\widehat{\boldsymbol{\beta}}_{n,W}(k)$ qui maximisent la statistique (3.39).

Vraisemblance empirique avec des valeurs imputées

Pour le profil de la vraisemblance empirique avec des données complètes et pondérés, les informations contenues dans les données ne sont pas suffisantes. Puisque les observations avec des données manquantes sont exclus quand on a construit le rapport de vraisemblance empirique, les taux de recouvrement de la région de confiance sont réduits quand beaucoup des valeurs manquantes ont eu lieu.
Afin de résoudre ce problème, on utilise la méthode d'imputation pour reconstituer les variables réponse manquantes. Cette méthode consiste à attribuer une valeur pour chaque réponse manquante afin d'obtenir un ensemble de données complet et ensuite appliquer les méthodes statistiques standards (Voir par exemple les papiers de [Yates (1993)], [Healy and Westmacott (1956)] et [Cheng (1994)]). Parmi les méthodes d'imputation, on considère la méthode d'imputation en utilisant la régression non-linéaire. On impute Y_i (ou Y_j) si Y_i (ou Y_j) est manquante et on donne des nouvelles valeurs pour les variables observées, en utilisant l'estimateur des moindres carrés pour les paramètres de régression $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_1$ et les estimateurs non paramétriques pour les probabilités $\pi_1(\mathbf{X}_i)$ et $\pi_2(\mathbf{X}_j)$. Le but de cette méthode est d'utiliser de manière suffisante toutes les informations qui contiennent dans les

données [1].
On définit :

$$\begin{aligned} Y_{i,R} &\equiv \frac{\delta_i}{\widehat{\pi}_1(\mathbf{X}_i)} Y_i + \Big(1 - \frac{\delta_i}{\widehat{\pi}_1(\mathbf{X}_i)}\Big) f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}), \qquad i \in I, \\ Y_{j,R} &\equiv \frac{\delta_j}{\widehat{\pi}_2(\mathbf{X}_j)} Y_j + \Big(1 - \frac{\delta_j}{\widehat{\pi}_2(\mathbf{X}_j)}\Big) f(\mathbf{X}_j, \widehat{\boldsymbol{\beta}}_1), \qquad j \in J, \end{aligned} \tag{3.41}$$

où $\widehat{\boldsymbol{\beta}}$ et $\widehat{\boldsymbol{\beta}}_1$ sont donnés par la relation (3.35).

On considère la vraisemblance empirique en tenant compte de toutes les valeurs imputées. Alors, pour $j \in J$, on considère $Y^*_{j,R} \equiv Y_{j,R} - f(\mathbf{X}_j, \boldsymbol{\beta} - \boldsymbol{\gamma}^0) + f(\mathbf{X}_j, \boldsymbol{\beta})$.

Dans ce cas, les vecteurs aléatoires auxiliaires sont définis par

$$\begin{cases} \mathbf{g}_{i,R}(\boldsymbol{\beta}) \equiv \dot{\mathbf{f}}_i(\boldsymbol{\beta})(Y_{i,R} - f_i(\boldsymbol{\beta})), & i \in I, \\ \mathbf{g}_{j,R}(\boldsymbol{\beta}) \equiv \dot{\mathbf{f}}_j(\boldsymbol{\beta})(Y^*_{j,R} - f_j(\boldsymbol{\beta})), & j \in J. \end{cases}$$

Soient les matrices suivantes

$$\mathbf{V}_{1n,R}(\boldsymbol{\beta}) \equiv \frac{1}{k} \sum_{i \in I} \dot{\mathbf{g}}_{i,R}(\boldsymbol{\beta}), \qquad \mathbf{V}_{2n,R}(\boldsymbol{\beta}) \equiv \frac{1}{n-k} \sum_{j \in J} \dot{\mathbf{g}}_{j,R}(\boldsymbol{\beta}).$$

Comme pour les deux premières méthodes, la statistique de test pour la méthode d'imputation est

$$\mathcal{Z}_{nk,R}(\boldsymbol{\gamma}_0, \boldsymbol{\lambda}_R, \boldsymbol{\beta}) \equiv 2\Big[\sum_{i \in I} \log(1 + \frac{n}{k} \boldsymbol{\lambda}_R^t \mathbf{z}_{i,R}(\boldsymbol{\beta})) + \sum_{j \in J} \log(1 - \frac{n}{n-k} \boldsymbol{\lambda}_R^t \mathbf{z}_{j,R}(\boldsymbol{\beta}))\Big], \tag{3.42}$$

où

$$\mathbf{z}_{i,R}(\boldsymbol{\beta}) \equiv \mathbf{M}^{\frac{1}{2}}_{n,R}(\boldsymbol{\beta}) \mathbf{V}^{-1}_{1n,R}(\boldsymbol{\beta}) \mathbf{g}_{i,R}(\boldsymbol{\beta}), \qquad \mathbf{z}_{j,R}(\boldsymbol{\beta}) \equiv \mathbf{M}^{\frac{1}{2}}_{n,R}(\boldsymbol{\beta}) \mathbf{V}^{-1}_{2n,R}(\boldsymbol{\beta}) \mathbf{g}_{j,R}(\boldsymbol{\beta}),$$

$$\mathbf{M}_{n,R}(\boldsymbol{\beta}) \equiv \frac{k(n-k)}{n^2} \mathbf{V}_{1n,R}(\boldsymbol{\beta}) \mathbf{H}_{n,R}(\boldsymbol{\beta}) \mathbf{V}_{2n,R}(\boldsymbol{\beta})$$

1. [Wang and Rao (2002)] a considéré le cas où la probabilité que Y_i est manquante est égale à 1 pour tout $i = 1, \ldots, n$. Dans ce cas, les valeurs observés sont conservées et les valeurs manquantes sont imputées avec la régression linéaire.

et

$$\mathbf{H}_{n,R}(\boldsymbol{\beta}) \equiv \Big[\frac{n-k}{nk}\sigma_1^2\mathbf{V}_{2n,R}(\boldsymbol{\beta})\mathbf{V}_{1n,R}^{-1}(\boldsymbol{\beta})\sum_{i\in I}\frac{1}{\pi_1(\mathbf{X}_i)}\big[\ddot{\mathbf{f}}_i(\boldsymbol{\beta})(Y_{i,R}-f_i(\boldsymbol{\beta}))-\dot{\mathbf{f}}_i(\boldsymbol{\beta})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta})\big] + \frac{k}{n(n-k)}\sigma_2^2\sum_{j\in J}\frac{1}{\pi_2(\mathbf{X}_j)}\big[\ddot{\mathbf{f}}_j(\boldsymbol{\beta})(Y_{j,R}^*-f_j(\boldsymbol{\beta}))-\dot{\mathbf{f}}_j(\boldsymbol{\beta})\dot{\mathbf{f}}_j^t(\boldsymbol{\beta})\big]\mathbf{V}_{2n,R}^{-1}(\boldsymbol{\beta})\mathbf{V}_{1n,R}(\boldsymbol{\beta})\Big]^{-1}.$$

Les fonctions scores de la statistique de test donnée par la relation (3.42) sont

$$\begin{cases} \boldsymbol{\phi}_{1n,R}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_R,\boldsymbol{\beta}) = \displaystyle\sum_{i\in I}\frac{\mathbf{z}_{i,R}(\boldsymbol{\beta})}{\frac{k}{n}+\boldsymbol{\lambda}_R^t\mathbf{z}_{i,R}(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\mathbf{z}_{j,R}(\boldsymbol{\beta})}{\frac{n-k}{n}-\boldsymbol{\lambda}_R^t\mathbf{z}_{j,R}(\boldsymbol{\beta})}, \\ \boldsymbol{\phi}_{2n,R}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_R,\boldsymbol{\beta}) = \displaystyle\sum_{i\in I}\frac{\dot{\mathbf{z}}_{i,R}(\boldsymbol{\beta})}{\frac{k}{n}+\boldsymbol{\lambda}_R^t\mathbf{z}_{i,R}(\boldsymbol{\beta})} - \sum_{j\in J}\frac{\dot{\mathbf{z}}_{j,R}(\boldsymbol{\beta})}{\frac{n-k}{n}-\boldsymbol{\lambda}_R^t\mathbf{z}_{j,R}(\boldsymbol{\beta})}, \end{cases} \tag{3.43}$$

où $\dot{\mathbf{z}}_{i,R}(\boldsymbol{\beta})$ et $\dot{\mathbf{z}}_{j,R}(\boldsymbol{\beta})$ sont respectivement les dérivées de $\mathbf{z}_{i,R}(\boldsymbol{\beta})$ et $\mathbf{z}_{j,R}(\boldsymbol{\beta})$ par rapport à $\boldsymbol{\beta}$.

Alors, en résolvant le système $\boldsymbol{\phi}_{1n,R}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_R,\boldsymbol{\beta}) = \mathbf{0}_d$ et $\boldsymbol{\phi}_{2n,R}(\boldsymbol{\gamma}_0,\boldsymbol{\lambda}_R,\boldsymbol{\beta}) = \mathbf{0}_d$ donnés par (3.43), on obtient $\widehat{\boldsymbol{\lambda}}_{n,R}(k)$ et $\widehat{\boldsymbol{\beta}}_{n,R}(k)$, qui maximisent la statistique du (3.42).

3.4.2 Comportement asymptotique de $\mathcal{Z}_{nk,C}, \mathcal{Z}_{nk,W}$ et $\mathcal{Z}_{nk,R}$

Dans cette section, on étudie les distributions asymptotiques des statistiques de vraisemblance empirique $\mathcal{Z}_{nk,C}, \mathcal{Z}_{nk,W}$ et $\mathcal{Z}_{nk,R}$, données respectivement par les relations (3.33), (3.39) et (3.42). Le résultat principal est donné par le Théorème 3.4.1, où on montre que sous l'hypothèse nulle H_0, la distribution asymptotique de ces trois statistiques est une loi de chi-deux.

Dans le cas des données manquantes, on a besoin de la supposition suivante :

(A10) Les matrices $\mathbf{V}_{1n,C}$, $\mathbf{V}_{2n,C}$, $\mathbf{V}_{1n,W}$, $\mathbf{V}_{2n,W}$, $\mathbf{V}_{1n,R}$ et $\mathbf{V}_{2n,R}$, sont inversibles, pour chaque $\mathbf{X}_i \in \Upsilon$ et pour tout $\boldsymbol{\beta}$ dans le voisinage de $\boldsymbol{\beta}^0$, leurs déterminants sont bornés pour une valeur de n assez grande.

La proposition suivante est similaire à la Proposition 3.3.1, pour chacune des trois statistiques du rapport de vraisemblance empirique définies dans la Section 3.4.1 dans le cas où on a des variables réponse manquantes. Alors, on montre que $\|\widehat{\boldsymbol{\lambda}}_{n,C}(k)\| \to 0$,

$\|\widehat{\boldsymbol{\lambda}}_{n,W}(k)\| \to 0$ et $\|\widehat{\boldsymbol{\lambda}}_{n,R}(k)\| \to 0$, quand $n \to \infty$ et que $\widehat{\boldsymbol{\beta}}_{n,C}(k)$, $\widehat{\boldsymbol{\beta}}_{n,W}(k)$ et $\widehat{\boldsymbol{\beta}}_{n,R}(k)$ sont des estimateurs consistants de $\boldsymbol{\beta}^0$, sous l'hypothèse H_0.

Proposition 3.4.1. *Sous l'hypothèse nulle* H_0*, si les suppositions (A1)-(A4), (A6)-(A10) sont vérifiées, alors, pour les estimateurs* $(\widehat{\boldsymbol{\lambda}}_{n,C}(k), \widehat{\boldsymbol{\beta}}_{n,C}(k))$, $(\widehat{\boldsymbol{\lambda}}_{n,W}(k), \widehat{\boldsymbol{\beta}}_{n,W}(k))$ *et* $(\widehat{\boldsymbol{\lambda}}_{n,R}(k), \widehat{\boldsymbol{\beta}}_{n,R}(k))$ *obtenus en résolvant respectivement les systèmes (3.36),(3.40) et (3.43), on a*

$$\begin{aligned} &\widehat{\boldsymbol{\lambda}}_{n,C}(k) = O_{\mathbb{P}}(n^{-1/2}), \quad && \widehat{\boldsymbol{\lambda}}_{n,W}(k) = O_{\mathbb{P}}(n^{-1/2}), \quad && \widehat{\boldsymbol{\lambda}}_{n,R}(k) = O_{\mathbb{P}}(n^{-1/2}), \\ &\widehat{\boldsymbol{\beta}}_{n,C}(k) - \boldsymbol{\beta}^{\mathbf{0}} = o_{\mathbb{P}}(1), && \widehat{\boldsymbol{\beta}}_{n,W}(k) - \boldsymbol{\beta}^{\mathbf{0}} = o_{\mathbb{P}}(1), && \widehat{\boldsymbol{\beta}}_{n,R}(k) - \boldsymbol{\beta}^{\mathbf{0}} = o_{\mathbb{P}}(1). \end{aligned}$$

Preuve. La preuve de cette proposition est similaire à la preuve de la Proposition 3.3.1 donnée dans le cas où la variable réponse Y_i est observée pour tout $i = 1, \ldots, n$. ■

Le théorème suivant donne la distribution asymptotique des trois statistiques de vraisemblance empirique données par les relations (3.33), (3.39) et (3.42).

Théorème 3.4.1. *Supposons que (A1)-(A4), (A6)-(A10) sont vérifiées. Sous l'hypothèse nulle* H_0*, les statistiques* $\mathcal{Z}_{nk,C}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_{n,C}(k), \widehat{\boldsymbol{\beta}}_{n,C}(k))$, $\mathcal{Z}_{nk,W}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_{n,W}(k), \widehat{\boldsymbol{\beta}}_{n,W}(k))$ *et* $\mathcal{Z}_{nk,R}(\boldsymbol{\gamma}_0, \widehat{\boldsymbol{\lambda}}_{n,R}(k), \widehat{\boldsymbol{\beta}}_{n,R}(k))$ *ont une distribution asymptotique* $\chi^2(d)$*, où* d *désigne la dimension de* $\boldsymbol{\beta}$*.*

Pour un niveau fixé $\alpha \in (0, 1)$, les $(1 - \alpha)$ régions de confiance asymptotiques pour $\boldsymbol{\gamma}$, basées sur la statistique de vraisemblance empirique, pour les trois méthodes proposés, sont

$$\begin{aligned} CR_{\alpha,C} &= \{\boldsymbol{\gamma} \in \mathbb{R}^d : \mathcal{Z}_{nk,C}(\boldsymbol{\gamma}, \widehat{\boldsymbol{\lambda}}_{n,C}(k), \widehat{\boldsymbol{\beta}}_{n,C}(k)) < c_{1-\alpha;d}\}, \\ CR_{\alpha,W} &= \{\boldsymbol{\gamma} \in \mathbb{R}^d : \mathcal{Z}_{nk,W}(\boldsymbol{\gamma}, \widehat{\boldsymbol{\lambda}}_{n,W}(k), \widehat{\boldsymbol{\beta}}_{n,W}(k)) < c_{1-\alpha;d}\}, \\ CR_{\alpha,R} &= \{\boldsymbol{\gamma} \in \mathbb{R}^d : \mathcal{Z}_{nk,R}(\boldsymbol{\gamma}, \widehat{\boldsymbol{\lambda}}_{n,R}(k), \widehat{\boldsymbol{\beta}}_{n,R}(k)) < c_{1-\alpha;d}\}, \end{aligned}$$

où $c_{1-\alpha;d}$ est le quantile d'ordre $1 - \alpha$ de la distribution de χ^2 avec d degrés de liberté.
Preuve. La preuve de cette proposition est similaire à la preuve du Théorème 3.3.1 donnée dans le cas où la variable réponse Y_i est observée pour tout $i = 1, \ldots, n$. ■

3.5 Simulations

Dans cette section, on présente des simulations pour évaluer la performance de la méthode de vraisemblance empirique proposée, en utilisant la technique Monte-Carlo. Tout

d'abord, on considère le cas où pour toutes les observations $i = 1, \ldots, n$, la variable Y_i est observée, ensuite si des valeurs de $\mathbf{Y}$ sont manquantes.

On considère la fonction non-linéaire suivante

$$f(x, \boldsymbol{\beta}) = a\frac{1 - x^b}{b}, \tag{3.44}$$

avec $\boldsymbol{\beta} = (a, b) \in [-100, 100] \times [0.1, 20]$.

Pour l'erreur, on considère les trois cas suivants :

Cas a : $\varepsilon_i = \mathcal{N}(0, 1)$ et $\varepsilon_j = \mathcal{N}(0, 1)$,

Cas b : $\varepsilon_i = 1/\sqrt{6}(\chi^2(3) - 3)$ et $\varepsilon_j = 2/\sqrt{6}t(6)$,

Cas c : $\varepsilon_i = 2\mathcal{E}xp(2) - 1$ et $\varepsilon_j = \mathcal{N}(0, 1)$,

où $\mathcal{N}(0, 1)$, $\mathcal{E}xp(2)$, $\chi^2(3)$ et $t(6)$ sont respectivement la distribution normale standarde, la distribution exponentielle de moyenne $1/2$, la distribution chi-deux à 3 degré de liberté et la distribution student avec 6 degré de liberté.

On considère les deux modèles non-linéaires suivants :

Modèle 1. On génère l'ensemble des données avec le modèle suivant

$$\begin{cases} Y_i = a_1^0\dfrac{1 - X_i^{b_1^0}}{b_1^0} + \varepsilon_i, & i = 1, \ldots, k, \\ Y_j = a_1^0\dfrac{1 - X_i^{b_1^0}}{b_1^0} + \varepsilon_j, & j = k + 1, \ldots, n, \end{cases} \tag{3.45}$$

où $a_1^0 = 10$, $b_1^0 = 2$ et $\mathbf{X} \sim \chi^2(2)$. Dans cet exemple, l'hypothèse nulle est $\mathrm{H}_0 : \gamma_0 = (0, 0)^t$.

Modèle 2. On génère l'ensemble des données avec le modèle suivant

$$\begin{cases} Y_i = a_1^0\dfrac{1 - X_i^{b_1^0}}{b_1^0} + \varepsilon_i, & i = 1, \ldots, k, \\ Y_j = a_2^0\dfrac{1 - X_i^{b_2^0}}{b_2^0} + \varepsilon_j, & j = k + 1, \ldots, n, \end{cases} \tag{3.46}$$

où $a_1^0 = 10$, $b_1^0 = 2$, $a_2^0 = 7$, $b_2^0 = 1.75$ et $\mathbf{X} \sim \chi^2(2)$. Dans cet exemple, l'hypothèse nulle est $\mathrm{H}_0 : \gamma_0 = (3, 0.25)^t$.

3.5.1 Modèle avec des données complètes

Pour le niveau de confiance nominal $1 - \alpha = 0.95$ et M= 1000 répliques Monte-Carlo pour chaque modèle/cas, on présente dans le Tableau 3.5.1 les probabilités de recouvrement (PR) et les longueurs de l'intervalle de confiance (LIC) obtenues par la méthode de log-vraisemblance empirique pour les deux modèles sans données manquantes.

Ces résultats sont obtenus en utilisant le Théorème 3.3.1, pour une valeur fixée de n, telle que $n = 1000$ et pour différentes positions pour le point de rupture. Alors, la région de confiance asymptotique est $CR_\alpha = \{\boldsymbol{\gamma} \in \mathbb{R}^d : \mathcal{Z}_{nk}(\boldsymbol{\gamma}, \widehat{\boldsymbol{\lambda}}_n(k), \widehat{\boldsymbol{\beta}}_n(k)) < c_{1-\alpha;d}\}$, où $c_{1-\alpha;d}$ est le $(1 - \alpha)$ quantile de la distribution de chi-deux avec d degré de liberté.

Afin de calculer la probabilité de recouvrement (PR), on génère le modèle sous l'hypothèse nulle H_0 donnée par la relation (3.2) et on compte les nombres de fois parmi les M répliques Monte-Carlo, que la valeur obtenue de la statistique ne dépasse pas la valeur critique $c_{1-\alpha;d}$.

Pour chaque modèle, la longueur de l'intervalle de confiance (LIC), désigne la différence entre la plus grande valeur et la plus petite valeur obtenues pour la statistique $\mathcal{Z}_{nk}$. On voit que toutes les probabilités de recouvrement (PR) sont très proches de 0.95, ce qui indique la performance de la région du confiance proposée. Les résultats ne varient pas avec la localisation du point k, ni avec la distribution de l'erreur.

3.5.2 Modèle avec des données manquantes

Dans cette section, on suppose que la variable réponse Y_i peut être manquante au hasard (MAR).

Tout à long de cette section, les fonctions noyaux considérées sont des noyaux Epanechnikov, telles que

$$\begin{cases} K_1(X_i) = 0.75(1 - X_i^2)\mathbb{1}_{|X_i| \leq 1}, \\ K_2(X_j) = 0.75(1 - X_j^2)\mathbb{1}_{|X_j| \leq 1}. \end{cases}$$

On considère les fenêtres $h_{1n} = k^{-1/7}$ et $h_{2n} = (n - k)^{-1/7}$, qui satisfont la condition (A7).

Tableau 3.5.1 – PR et LIC pour les modèles 1 and 2, avec les cas a, b et c. $M = 1000$

		Modèle 1			**Modèle 2**		
k		**Cas a**	**Cas b**	**Cas c**	**Cas a**	**Cas b**	**Cas c**
300	PR	0.961	0.926	0.954	0.949	0.957	0.961
	LIC	5.718	5.157	5.422	4.978	5.637	5.229
500	PR	0.970	0.949	0.943	0.960	0.939	0.955
	LIC	4.813	5.173	5.438	5.119	5.276	5.393
700	PR	0.963	0.944	0.940	0.947	0.950	0.954
	LIC	5.577	5.913	5.817	5.788	5.909	5.949

Nous considérons les trois études suivantes de probabilités de réponse sous la supposition MAR :

Étude 1 :

$$\pi_1(X_i) = \begin{cases} 0.8 + 0.2|X_i - 1| & \text{si } |X_i - 1| \leq 1 \\ 0.95 & \text{sinon.} \end{cases}$$

$$\pi_2(X_i) = \begin{cases} 0.8 + 0.2|X_j - 1| & \text{si } |X_j - 1| \leq 1 \\ 0.95 & \text{sinon.} \end{cases}$$

Étude 2 :

$$\pi_1(X_i) = \{\ 0.8 \quad \text{pour tout } X_i.$$

$$\pi_2(X_j) = \{\ 0.8 \quad \text{pour tout } X_j.$$

Étude 3 :

$$\pi_1(X_i) = \begin{cases} 0.8 + 0.2|X_i - 1| & \text{si } |X_i - 1| \leq 1 \\ 0.95 & \text{sinon.} \end{cases}$$

$$\pi_2(X_j) = \{\ 0.8 \quad \text{pour tout } X_j.$$

Pour les études **1**, **2** et **3**, les Tableaux 3.5.2, 3.5.3 et 3.5.4, présentent les probabilités de recouvrement (PR) et les longueurs de l'intervalle de confiance (LIC) en utilisant les statistiques $\mathcal{Z}_{nk,C}, \mathcal{Z}_{nk,W}$ et $\mathcal{Z}_{nk,R}$ données respectivement par les relations (3.33), (3.39) et (3.42). On considère pour chaque étude $1 - \alpha = 0.95$, $n = 1000$ et on applique les trois cas **a**, **b** et **c**. On réalise $M = 1000$ répliques pour chaque simulation.

Pour le **Modèle 1**, les probabilités de recouvrement (PR) obtenues sont proches de 0.95. Les LIC obtenus en utilisant une méthode d'imputation sont moins larges que ceux obtenus pour les méthodes cas-complet et pondérée.
Pour le **Modèle 2**, on obtient que toutes les probabilités de recouvrement (PR) obtenues sont proches de 0.95. Les LIC obtenus en utilisant une méthode d'imputation sont moins larges que ceux obtenus pour les méthodes cas-complet et pondérée.

Tableau 3.5.2 – PR et LIC, pour l'étude **1**, avec les cas **a**, **b** et **c**, $n = 1000$, $k = 600$.

		Modèle 1			**Modèle 2**		
Y		**Cas a**	**Cas b**	**Cas c**	**Cas a**	**Cas b**	**Cas c**
Cas-complet	PR	0.939	0.942	0.958	0.950	0.951	0.967
	LIC	5.490	5.573	5.183	5.493	5.607	5.939
Pondérée	PR	0.948	0.943	0.933	0.949	0.926	0.953
	LIC	5.551	5.820	5.137	5.726	5.483	5.357
Imputée	PR	0.929	0.937	0.951	0.940	0.933	0.950
	LIC	5.420	5.324	4.892	5.207	5.229	5.233

Tableau 3.5.3 – PR et LIC, pour l'étude **2**, avec les cas **a**, **b** et **c**, $n = 1000$, $k = 600$.

		Modèle 1			**Modèle 2**		
Y		**Cas a**	**Cas b**	**Cas c**	**Cas a**	**Cas b**	**Cas c**
Cas-complet	PR	0.937	0.942	0.939	0.944	0.953	0.946
	LIC	5.962	5.258	5.414	5.426	5.382	5.547
Pondérée	PR	0.925	0.930	0.927	0.962	0.942	0.950
	LIC	5.482	5.679	5.519	5.425	5.437	5.493
Imputée	PR	0.929	0.934	0.959	0.960	0.943	0.953
	LIC	5.355	5.257	4.594	5.380	5.422	5.463

Tableau 3.5.4 – PR et LIC, pour l'étude **3**, avec les cas **a**, **b** et **c**, $n = 1000$, $k = 600$.

Y		**Modèle 1**			**Modèle 2**		
		Cas a	**Cas b**	**Cas c**	**Cas a**	**Cas b**	**Cas c**
Cas-complet	PR	0.951	0.938	0.929	0.953	0.947	0.936
	LIC	5.352	5.637	5.429	5.189	5.293	5.388
Pondérée	PR	0.927	0.939	0.941	0.950	0.943	0.959
	LIC	5.854	5.680	5.705	5.482	5.379	5.523
Imputée	PR	0.950	0.960	0.945	0.944	0.928	0.937
	LIC	5.225	5.412	5.5379	4.573	4.988	5.401

3.6 Annexe

Lemme 3.6.1. *On suppose que (A1)-(A4) sont vérifiées. Il existe $\eta_0 > 0$, tel que pour tout $\eta \in [0, \eta_0]$ et tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$, on a*

(i) $\mathbf{V}_{1n}(\boldsymbol{\beta}) = \mathbf{V}^0_{1n} + o_{\mathbb{P}}(\boldsymbol{\beta} - \boldsymbol{\beta^0})$.

(ii) $\mathbf{V}_{2n}(\boldsymbol{\beta}) = \mathbf{V}^0_{2n} + o_{\mathbb{P}}(\boldsymbol{\beta} - \boldsymbol{\beta^0})$.

On rappelle que, $\mathbf{V}_{1n}(\boldsymbol{\beta}) = \dfrac{1}{k}\sum_{i\in I} \dot{\mathbf{g}}_i(\boldsymbol{\beta})$ *et* $\mathbf{V}_{2n}(\boldsymbol{\beta}) = \dfrac{1}{n-k}\sum_{j\in J} \dot{\mathbf{g}}_j(\boldsymbol{\beta})$.

Preuve. (i) Par le développement de Taylor de $\dot{\mathbf{g}}_i(\boldsymbol{\beta})$ en $\boldsymbol{\beta} = \boldsymbol{\beta^0}$ jusqu'à l'ordre 2, on obtient

$$\begin{aligned}
\mathbf{V}_{1n}(\boldsymbol{\beta}) \;=\; & \left[\frac{1}{k}\sum_{i\in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i - \frac{1}{k}\sum_{i\in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0})\right] \\
& + \frac{1}{k}\sum_{i\in I} \mathbf{M}_i\varepsilon_i - \frac{1}{k}\sum_{i\in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(1)})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t \\
& - \frac{1}{k}\sum_{i\in I} \mathbf{M}_i\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(1)})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t - \frac{1}{k}\sum_{i\in I} \mathbf{M}_{1i}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t \\
& + \frac{1}{k}\sum_{i\in I} \mathbf{M}_{2i}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t + \frac{1}{k}\sum_{i\in I} \mathbf{M}_{2i}\mathbf{M}_{3i}(\boldsymbol{\beta} - \boldsymbol{\beta^0})(\boldsymbol{\beta} - \boldsymbol{\beta^0})^t, \quad (3.47)
\end{aligned}$$

où, $\mathbf{M}_i$, $\mathbf{M}_{1i}$ et $\mathbf{M}_{2i}$ sont des matrices de dimension $d \times d$, définies par $\mathbf{M}_i = \sum_{r=1}^{d} \frac{\partial \ddot{\mathbf{f}}_i(\boldsymbol{\beta}_{i,r}^{(2)})}{\partial \beta_r}(\beta_r - \beta_r^0)$, $\mathbf{M}_{1i} = \big(\frac{\partial^2 f_i(\boldsymbol{\beta}_{i,rl}^{(3)})}{\partial \beta_r \partial \beta_l}\big)_{1\leq r,l\leq d}$ et $\mathbf{M}_{2i} = \big(\frac{\partial^2 f_i(\boldsymbol{\beta}_{i,rl}^{(4)})}{\partial \beta_r \partial \beta_l}\big)_{1\leq r,l\leq d}$.
Pour $1 \leq r, l \leq d$, β_r désigne la r-ème composante du vecteur $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_{i,rl}^{(a)} = \boldsymbol{\beta^0} + u_{i,rl}^{(a)}(\boldsymbol{\beta} - \boldsymbol{\beta^0})$, avec $u_{i,rl}^{(a)} \in [0,1]$ et $a \in \{2,3,4\}$.

Pour le premier terme du membre droit de la relation (3.47), on a

$$\frac{1}{k}\sum_{i\in I} \ddot{\mathbf{f}}_i(\boldsymbol{\beta^0})\varepsilon_i - \frac{1}{k}\sum_{i\in I} \dot{\mathbf{f}}_i(\boldsymbol{\beta^0})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0}) = \mathbf{V}_{1n}^0. \tag{3.48}$$

Par l'inégalité de Bienaymé-Tchebychev et la supposition (A1), pour tout $C_1 > 0$ et $i \in I$, on obtient que

$$\mathbb{P}\Big[|\varepsilon_i| > C_1\Big] \leq \frac{\sigma_1^2}{C_1}. \tag{3.49}$$

Pour $1 \leq r, l \leq d$ et pour i fixé, tel que $i \in I$, $M_{i,sl}$ est une variable aléatoire qui désigne le terme (s,l) de la matrice $\mathbf{M}_i$ tel que

$$M_{i,sl} = \sum_{r=1}^{d} \frac{\partial f_i(\boldsymbol{\beta}_{i,r}^{(2)})}{\partial \beta_r \partial \beta_s \partial \beta_l}(\beta_r - \beta_r^0).$$

En utilisant la supposition (A3), on a que $|M_{i,sl}| \leq \|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|$ avec une probabilité égale à 1. En utilisant le Lemme 2.6.1(iii), pour tout $C_2 > 0$ et $i \in I$, on obtient que

$$\|\mathbf{M}_i\|_1 \leq C_2\|\boldsymbol{\beta} - \boldsymbol{\beta^0}\|. \tag{3.50}$$

Pour le deuxième terme du membre droit de la relation (3.47), on a

$$\|\frac{1}{k}\sum_{i\in I} \mathbf{M}_i\varepsilon_i\|_1 \leq \frac{1}{k}\sum_{i\in I} \|\mathbf{M}_i\varepsilon_i\|_1 \leq \frac{1}{k}\sum_{i\in I} \|\mathbf{M}_i\|_1|\varepsilon_i|.$$

En utilisant les relations (3.49) et (3.50), on obtient que

$$\frac{1}{k}\sum_{i\in I} \mathbf{M}_i\varepsilon_i = o_{\mathbb{P}}(\boldsymbol{\beta} - \boldsymbol{\beta^0}). \tag{3.51}$$

Par l'inégalité de Markov et en tenant compte de la supposition (A4), alors pour une valeur fixée de i et pour $1 \leq r \leq l$, on a que

$$\mathbb{P}\left[\Big|\frac{\partial f_i(\boldsymbol{\beta}_{i,r}^{(2)})}{\partial \beta_r}\Big| > \epsilon_1\right] \leq \frac{1}{\epsilon_1}\mathbb{E}\left[\Big|\frac{\partial f_i(\boldsymbol{\beta}_{i,r}^{(2)})}{\partial \beta_r}\Big|\right].$$

Pour $\epsilon_1 = \mathbb{E}(|\partial f_i(\boldsymbol{\beta}_{i,r}^{(2)})/\partial\beta_r|)/\epsilon$, on obtient que

$$\mathbb{P}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta}_{i,r}^{(2)})}{\partial\beta_r}\Big| > \frac{1}{\epsilon}\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta}_{i,r}^{(2)})}{\partial\beta_r}\Big|\Big] \leq \epsilon.$$

La dernière relation ainsi le Lemme 2.6.1*(i)* impliquent

$$\mathbb{P}\Big[\|\dot{\mathbf{f}}_i(\boldsymbol{\beta}_{i,r}^{(2)})\|_1 \geq \frac{d}{\epsilon}\max\mathbb{E}\Big[\Big|\frac{\partial f_i(\boldsymbol{\beta}_{i,r}^{(2)})}{\partial\beta_r}\Big|\Big]\Big] \leq \epsilon. \tag{3.52}$$

Pour le troisième terme du membre droit de la relation (3.47), par la supposition (A3) et en utilisant la relation (3.52), on obtient que

$$\begin{aligned}\frac{1}{k}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(1)})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\|_1 &\leq (\frac{1}{k}\sum_{i\in I}\|\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(1)})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\|_1 \\ &\leq \frac{1}{k}\sum_{i\in I}\|\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\|_1\|\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(1)})\|_1\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_1, \end{aligned}\tag{3.53}$$

ce qui donne

$$\frac{1}{k}\sum_{i\in I}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(1)})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{3.54}$$

En utilisant la supposition (A3), pour $1 \leq s, r \leq d$, pour tout $\epsilon > 0$ il existe $\epsilon_3 > 0$ tel que

$$\mathbb{P}\Big[\Big|\frac{\partial^2 f_i(\boldsymbol{\beta}_{i,sr}^{(3)})}{\partial\beta_s\partial\beta_r}\Big| \geq \epsilon_2\Big] \leq \epsilon.$$

Par le Lemme 2.6.1 (iii), pour tout $\epsilon > 0$, on a que

$$\mathbb{P}\Big[\|\mathbf{M}_{1i}\|_1 \geq \epsilon_3\Big] \leq \epsilon. \tag{3.55}$$

En utilisant la supposition (A3) et par un argument similaire pour $\mathbf{M}_{1i}$, pour $1 \leq s, r \leq d$, pour tout $\epsilon > 0$ il existe $\epsilon_4 > 0$ tel que

$$\mathbb{P}\Big[\|\mathbf{M}_{2i}\|_1 \geq \epsilon_4\Big] \leq \epsilon. \tag{3.56}$$

Pour le quatrième terme du membre droit de la relation (3.47), en utilisant les relations (3.50), (3.52) et le fait que $\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|_2 \leq \eta$, on obtient que

$$\frac{1}{k}\sum_{i\in I}\mathbf{M}_i\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{(1)})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{3.57}$$

En utilisant les suppositions (A3) et les relations (3.52), (3.55), pour le cinquième terme du membre droit de la relation (3.47), on a

$$\frac{1}{k}\sum_{i\in I}\mathbf{M}_{1i}\dot{\mathbf{f}}_i(\boldsymbol{\beta^0})(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{3.58}$$

De même, en utilisant la supposition (A3) et les relations (3.52), (3.55) et (3.56), on obtient que

$$\frac{1}{k}\sum_{i\in I}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\mathbf{M}_{2i}\dot{\mathbf{f}}_i^t(\boldsymbol{\beta^0}) = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) \tag{3.59}$$

et

$$\frac{1}{k}\sum_{i\in I}(\boldsymbol{\beta}-\boldsymbol{\beta^0})\mathbf{M}_{2i}\mathbf{M}_{3i}(\boldsymbol{\beta}-\boldsymbol{\beta^0})^t = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{3.60}$$

En combinant les relations (3.48), (3.51), (3.54) et (3.57)-(3.60), on obtient que

$$\mathbf{V}_{1n}(\boldsymbol{\beta}) = \mathbf{V}_{1n}^0 + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}).$$

(ii) De la même façon que *(i)*, on peut démontrer que

$$\mathbf{V}_{2n}(\boldsymbol{\beta}) = \mathbf{V}_{2n}^0 + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}).$$

■

Conséquence de ce Lemme, on obtient $\mathbf{M}_n(\boldsymbol{\beta}) = \mathbf{M}_n(\boldsymbol{\beta^0}) + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0})$, où la matrice $\mathbf{M}_n(\boldsymbol{\beta})$ est donnée par la Remarque 3.3.1. Pour simplifier la notation, on écrit $\mathbf{M}_n(\boldsymbol{\beta^0}) = \mathbf{M}_n^0$.

Par le lemme suivant, on donne les comportements de $\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})$ dans le voisinage de $\boldsymbol{\beta^0}$.

Lemme 3.6.2. *On suppose que (A1)-(A6) sont vérifiées. Il existe $\eta_0 > 0$, tel que pour tout $\eta \in [0, \eta_0]$ et tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta^0})$ on a*

$$\frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}) = \frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta^0}) - \frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}) + o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta^0}). \tag{3.61}$$

Preuve. Sous la supposition (A5), en utilisant le Lemme 3.6.1, le développement de Taylor de $k^{-1}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta})$ dans le voisinage de $\boldsymbol{\beta}=\boldsymbol{\beta}^0$ jusqu'à l'ordre 2, est

$$\begin{aligned}\frac{1}{k}\sum_{i\in I}\mathbf{z}_i(\boldsymbol{\beta}) &= \frac{1}{k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i+\frac{1}{2k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i\\
&-\frac{1}{2k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\
&-\frac{1}{6k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{4i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\
&-\frac{1}{4k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\
&-\frac{1}{12k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{4i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\
&-\frac{1}{12k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{D}_i(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{4i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\\
&+\frac{1}{6k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{D}_i\varepsilon_i, \qquad (3.62)\end{aligned}$$

où $\mathbf{M}_{4i}=\big(\frac{\partial^2 f_i(\boldsymbol{\beta}_{i,ul}^{(5)})}{\partial\beta_u\partial\beta_l}\big)_{1\le u,l\le d}$ est une matrice de dimension $d\times d$ et

$$\mathbf{D}_i=\Big(\sum_{l=1}^{d}\sum_{u=1}^{d}\frac{\partial^2\dot{\mathbf{f}}_i(\boldsymbol{\beta}_{i,ul}^{(6)})}{\partial\beta_u\partial\beta_l}\cdot(\beta_u-\beta_u^0)(\beta_l-\beta_l^0)\Big)$$

est un vecteur de dimension d. Pour $1\le u,l\le d$, β_u désigne la u-ème composante du vecteur $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_{i,ul}^{(a)}=\boldsymbol{\beta}^{\mathbf{0}}+v_{i,ul}^{(a)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})$, avec $v_{i,ul}^{(a)}\in[0,1]$ et $a\in\{5,6\}$.

Pour le deuxième terme du membre droit de la relation (3.62), par la loi des grands nombres, le terme $(2k)^{-1}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i$ converge presque sûrement vers l'espérance de $(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i$, quand $n\to\infty$. Puisque ε_i est indépendant de $\mathbf{X}_i$ et $\mathbb{E}[\varepsilon_i]=0$, on a

$$\frac{1}{2k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\varepsilon_i=o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \qquad (3.63)$$

Pour le troisième terme du membre droit de la relation (3.62), par des calculs simples, on peut obtenir

$$\frac{1}{2k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = -\frac{1}{2}(\mathbf{M}_n^0)^{\frac{1}{2}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{3.64}$$

En utilisant les suppositions (A3), (A4) et par un argument similaire à celui utilisé pour les relations (3.52) et (3.55), pour les quatrième, cinquième et sixième termes du membre droit de la relation (3.62), on obtient respectivement, que

$$\frac{1}{6k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\dot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{4i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}), \tag{3.65}$$

$$\frac{1}{4k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})\dot{\mathbf{f}}_i^t(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) \tag{3.66}$$

et

$$\frac{1}{12k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\ddot{\mathbf{f}}_i(\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{4i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{3.67}$$

Pour chaque i, tel que $1 \leq i \leq k$ et pour $1 \leq s \leq d$, on note par D_{is} la variable aléatoire qui désigne la s-ème composante du vecteur $\mathbf{D}_i$, telle que

$$\mathbf{D}_{is} = \sum_{l=1}^{d}\sum_{u=1}^{d}\frac{\partial^3 f_i(\boldsymbol{\beta}_{i,ul}^{(6)})}{\partial\beta_u\partial\beta_l\partial\beta_s}(\beta_u-\beta_u^0)(\beta_l-\beta_l^0).$$

En utilisant le Lemme 2.6.1, ainsi que la supposition (A3), pour tout $C_3 > 0$, on a

$$\|\mathbf{D}\|_1 \leq C_3\|\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}\|^2. \tag{3.68}$$

La dernière relation, avec la relation (3.49), impliquent que

$$\frac{1}{6k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{D}_i\varepsilon_i = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{3.69}$$

Finalement, pour le terme $(12k)^{-1}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{D}_i(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{4i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})$, la supposition (A3) et la relation (3.68) donnent

$$\frac{1}{12k}\sum_{i\in I}(\mathbf{M}_n^0)^{\frac{1}{2}}(\mathbf{V}_{1n}^0)^{-1}\mathbf{D}_i(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}})^t\mathbf{M}_{4i}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}) = o_{\mathbb{P}}(\boldsymbol{\beta}-\boldsymbol{\beta}^{\mathbf{0}}). \tag{3.70}$$

En combinant les relations (3.63)-(3.67), (3.69) et (3.70), le lemme est obtenu. ■

Deuxième partie

Vraisemblance empirique pour un modèle à deux phases en grande dimension

Chapitre 4

Vraisemblance empirique pour un modèle à deux phases en grande dimension

Sommaire

4.1 Introduction 124
4.2 Préliminaires 127
4.2.1 Notations 130
4.2.2 Suppositions 131
4.3 Résultat principal 132
4.4 Simulations 157
4.4.1 L'importance des suppositions (A5) et (A6) 157
4.5 Annexe 162

Abstract A non parametric method based on the empirical likelihood is proposed for detecting the change in the coefficients of high-dimensional linear model where the number of model variables may increase as the sample size increases. This amounts to testing the nulle hypothesis of no change against the alternative of one change in the regression coefficients. Based on the theoretical asymptotic behaviour of the empirical likelihood ratio statistic, we propose, for a fixed design, a simpler test statistic, easier to use in practice. The asymptotic normality of the proposed test statistic under the null hypothesis is proved, a result which is different from the χ^2 law for a model with a fixed variable number. Under alternative hypothesis, the test statistic diverges. Some Monte-Carlo simulations study the

behaviour of the proposed test statistic.

Key-words Change-point ; High-dimension ; Linear model ; Empirical likelihood test.

Résumé Une méthode non-paramétrique basée sur la vraisemblance empirique est proposée afin de détecter le changement dans les coefficients d'un modèle linéaire en grande dimension avec des régresseurs déterministes, quand le nombre des variables du modèle peut augmenter avec la taille de l'échantillon. Ceci nous permet de tester l'hypothèse nulle de non-changement contre l'hypothèse alternative d'un seul changement dans les coefficients de régression. Basée sur le comportement asymptotique théorique de la statistique du rapport de vraisemblance empirique, on propose une statistique de test qui sera utilisée facilement dans la pratique. La normalité asymptotique de la statistique de test proposée sous l'hypothèse nulle est prouvée en obtenant des résultats différents à ceux obtenus quand le nombre des variables est fixé et la loi asymptotique est le χ^2. Sous l'hypothèse alternative, la statistique de test diverge. Des simulations de Monte-Carlo sont considérées afin d'étudier le comportement de la statistique de test proposée.

Mots-clefs Point de rupture. Grande dimension. Modèle linéaire. Test de la vraisemblance empirique.

4.1 Introduction

Le développement de la technologie ainsi que des techniques numériques rapides permettent de considérer et d'étudier des modèles statistiques avec un grand nombre de variables. Pour des travaux concernant des données en grande dimension, on cite les papiers de [Yin et al. (1983)], [Portnoy (1984)], [Portnoy (1985)], [Silverstein (1985)], [Yin et al. (1988)], [Saranadasa (1993)], [Samarov et al. (2005)], [Dabo-Niang and Yao (2013)], [Butucea and Gayraud (2015)] et [Butucea et al. (2015)].

Le modèle en grande dimension se réfère à un modèle dont le nombre de variables explicatives p augmente à l'infini quand le nombre des observations n converge vers l'infini. Lorsque p diverge, les méthodes statistiques classiques ne fonctionnent pas.

Afin de sélectionner automatiquement les variables significatives, la plupart des travaux de la littérature utilisent les méthodes de type LASSO (Least Absolute Shrinkage and Selection Operator). Cette méthode a été introduite par [Tibshirani (1996)]. Si le modèle

contient des valeurs aberrantes, les paramètres estimés par les moindres carrées avec la pénalité LASSO ont une grande variabilité.

Ainsi, [Dicker et al. (2013)] a considéré un modèle quantile avec une pénalité seamless-L_0 lorsque le nombre de variables explicatives p est tel que $p \to \infty$, $p/n \to 0$ quand $n \to \infty$. Pour une régression quantile générale, [Wu and Liu (2009)] a proposé la pénalité SCAD, alors que, [Zou and Yuan (2008)] a considéré la régression quantile composite avec la pénalité LASSO. Le cas $p \to \infty$ a été également considéré par [Fan and Peng (2004)] pour une méthode de vraisemblance pénalisée non concave, lorsque $p^5/n \to \infty$.

Toutes ces méthodes sont basées sur le principe de sélection automatiques des variables significatives. Afin d'avoir des estimateurs de paramètres plus précis et un meilleur ajustement de la variable dépendante, cette dernière variable doit être modélisée seulement en fonction des variables importantes.

Pour un modèle avec des données aberrantes, pour avoir une prédiction plus précise et également robuste, une possibilité est de considérer la méthode de vraisemblance empirique. Pour ce type des modèles, la plupart des travaux ont étudié le cas où le nombre de variables explicatives p est fixé. Pour le modèle de régression linéaire en très grande dimension, on cite le papier de [Guo et al. (2013)] qui a étudié le cas quand les régresseurs sont déterministes. [Liu et al(2013)] a également considéré la méthode de vraisemblance empirique pour une suite des vecteurs aléatoires indépendants et identiquement distribués de dimension p, tel que $p \to \infty$ quand $n \to \infty$.

Dans ce chapitre, nous sommes intéressés a étudier si un changement dans les coefficients de régression a eu lieu. Le nombre p de variables explicatives varie avec la taille de l'échantillon n, et p converge vers l'infini si $n \to \infty$.

Puisque l'étude des données en grande dimension est une technique statistique récente, le nombre de publications qui traite le problème de point de rupture dans les modèles en grande dimension est limité. [Lung-Yut-Fong et al.(2013)] a proposé une approche pour détecter le point de rupture dans le réseau trafic en grand volume. Sous l'hypothèse nulle de non-changement, la distribution asymptotique de la statistique de test proposée dans le papier de [Lung-Yut-Fong et al.(2013)] est le *argsup* d'un pont Brownien.

Plusieurs papiers ont utilisés les méthodes de type LASSO. [Lee et al. (2015)] a considéré un point de rupture possible dans la régression en grande dimension avec des erreurs

gaussiennes. Dans [Ciuperca (2014)], les estimateurs LASSO adaptatifs et type LASSO sont étudiés, alors que dans [Ciuperca (2013b)], le modèle quantile avec pénalité SCAD est considéré. Ces deux papiers ont étudié des modèles avec une valeur de p fixée. Afin de choisir le nombre de points de rupture, [Ciuperca (2014)] a proposé un critère de sélection de modèle.

La technique de la vraisemblance empirique n'a pas encore été utilisée pour étudier le point de rupture dans des modèles en grande dimension.

Dans ce chapitre, on considère la méthode de vraisemblance empirique pour un modèle linéaire en grande dimension avec un point de rupture. On étudie le comportement asymptotique de la statistique de test du rapport de vraisemblance empirique dans le cas où les regresseus sont déterministes.

On considère d'abord le premier modèle linéaire

$$Y_i = \mathbf{X}_i^t \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \dots, n. \tag{4.1}$$

Maintenant, on considère un deuxième modèle linéaire qui subit un changement à l'observation k. C'est un modèle à deux phase (avec un point de rupture) :

$$Y_i = \begin{cases} \mathbf{X}_i^t \boldsymbol{\beta} + \varepsilon_i, & 1 \leq i \leq k, \\ \mathbf{X}_i^t \boldsymbol{\beta}_2 + \varepsilon_i, & k < i \leq n, \end{cases} \tag{4.2}$$

où $\mathbf{X}_i$ est un vecteur de p variables explicatives, $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_2$ sont des vecteurs de p paramètres inconnus et ε_i désigne l'erreur du modèle.
Le paramètre $\boldsymbol{\beta}$ de la première phase de l'équation (4.2) coïncide avec celle de l'équation (4.1). Pour les modèles (4.1) et (4.2), Y_i est la variable réponse de l'observation i. Les erreurs ε_i sont supposées indépendantes et identiquement distribuées (i.i.d), d'espérance zero et de variance finie σ^2.
On suppose que le nombre des variables explicatives p dépend de la taille de l'échantillon n : $p = p_n$, et $p_n \to \infty$ quand $n \to \infty$. Le point de rupture k de (4.2) dépend également de n. Le changement dans le modèle (4.2) a lieu assez loin de la première observation et suffisamment précédente de la dernière observation. Donc, on suppose que $\lim_{n\to\infty} k/n \in (0, 1)$.
Sous l'hypothèse H_0, on suppose qu'aucun changement n'a eu lieu dans les paramètres du modèle, donc le modèle a la forme (4.1). Alors, l'hypothèse nulle peut s'écrire

$$\mathrm{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}, \tag{4.3}$$

L'hypothèse alternative qui suppose la présence d'un seul changement dans les paramètres de régression est

$$\mathrm{H}_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}, \tag{4.4}$$

alors, le modèle a la forme (4.2).

Le chapitre est organisé comme suit. Des notations utilisées tout au long du chapitre sont définies et des suppositions utilisées pour l'étude théorique sont également énoncées dans la Section 4.2. Dans la Section 4.3, on construit la statistique de test du rapport de vraisemblance empirique et on étudie ces comportements asymptotiques. La distribution asymptotique sous l'hypothèse nulle H_0 de la statistique de test est obtenue, alors que cette statistique diverge sous l'hypothèse H_1. Dans la Section 4.4, on analyse le niveau empirique et la puissance empirique par des simulations qui confirment la performance de la statistique proposée. Des lemmes, avec leurs preuves, nécessaires pour les démonstrations des théorèmes et des propositions sont donnés dans la Section 4.5.

4.2 Préliminaires

Dans cette section, on introduit la méthode de vraisemblance empirique pour le modèle linéaire en grande dimension. Des notations et des suppositions sont également données.

Sous l'hypothèse nulle H_0, c'est-à-dire le modèle (4.1), on note par $\boldsymbol{\beta^0}$ la vraie valeur du paramètre $\boldsymbol{\beta}$. Sous l'hypothèse alternative H_1, c'est-à-dire le modèle (4.2), les vraies valeurs de $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_2$, sont respectivement $\boldsymbol{\beta}^0$ et $\boldsymbol{\beta}_2^0$.

Afin de définir le profil de la vraisemblance empirique (sous H_0 et ensuite sous H_1), pour tout $\boldsymbol{\beta} \in \mathbb{R}^p$ et $i = 1, \ldots, n$, on introduit le p-vecteur aléatoire suivant :

$$\mathbf{z}_i(\boldsymbol{\beta}) \equiv \mathbf{X}_i(Y_i - \mathbf{X}_i^t\boldsymbol{\beta}).$$

On considère également le vecteur

$$\mathbf{z}_i^0 \equiv \mathbf{X}_i\varepsilon_i.$$

Nous remarquons que, sous l'hypothèse H_0, pour tout $i = 1, \ldots, n$, on a $\mathbf{z}_i^0 = \mathbf{z}_i(\boldsymbol{\beta^0})$ et $\mathbb{E}[\mathbf{z}_i^0] = \mathbf{0}_p$. D'autre part, pour les $(\mathbf{X}_i)_{1\leq i\leq n}$ déterministes, les variables aléatoires $\mathbf{z}_i^0$ sont indépendantes mais non identiquement distribuées.

Pour le point de rupture, on considère la notation $\theta_{nk} = k/n$. Ainsi, en vue de la remarque donnée dans l'introduction, on suppose que $\theta_{nk} \to \theta^0 \in (0, 1)$ quand $n \to \infty$.

On définit les vecteurs de probabilité $(q_1, \ldots, q_k)$ et $(q_{k+1}, \ldots, q_n)$. Évidement, ces probabilités satisfont la relation

$$\sum_{i=1}^{k} q_i = 1; \quad \sum_{j=k+1}^{n} q_j = 1.$$

Sous l'hypothèse H_0 donnée par (4.3), le profil de la vraisemblance empirique pour $\boldsymbol{\beta}$ est

$$\mathcal{R}_{nk}(\boldsymbol{\beta}) \equiv \sup_{(q_1,\ldots,q_k)} \sup_{(q_{k+1},\ldots,q_n)} \left\{ \prod_{i=1}^{k} q_i \prod_{j=k+1}^{n} q_j; \sum_{i=1}^{k} q_i = \sum_{j=k+1}^{n} q_j = 1, \right.$$
$$\left. \sum_{i=1}^{k} q_i \mathbf{z}_i(\boldsymbol{\beta}) = \sum_{j=k+1}^{n} q_j \mathbf{z}_j(\boldsymbol{\beta}) = \mathbf{0}_p \right\}.$$

De même, sous l'hypothèse H_1 donnée par (4.4), le profil de la vraisemblance empirique est

$$\mathcal{R}_{nk}(\boldsymbol{\beta}, \boldsymbol{\beta}_2) \equiv \sup_{(q_1,\ldots,q_k)} \sup_{(q_{k+1},\ldots,q_n)} \left\{ \prod_{i=1}^{k} q_i \prod_{j=k+1}^{n} q_j; \sum_{i=1}^{k} q_i = \sum_{j=k+1}^{n} q_j = 1, \right.$$
$$\left. \sum_{i=1}^{k} q_i \mathbf{z}_i(\boldsymbol{\beta}) = \sum_{j=k+1}^{n} q_j \mathbf{z}_j(\boldsymbol{\beta}_2) = \mathbf{0}_p \right\}.$$

Puisque le dimension (p) de $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_2$ dépend de la taille de l'échantillon n et converge vers l'infini quand n converge vers l'infini, alors le Théorème 2.3.1 du Chapitre 2 ne sera plus applicable ainsi que la distribution asymptotique du profil du rapport de vraisemblance empirique pour $\boldsymbol{\beta}$ et $\boldsymbol{\beta}_2$ n'est plus une loi χ^2 .
Alors, en utilisant une idée similaire au test du maximum de la vraisemblance pour tester H_0 contre H_1, on considère le profil du rapport de vraisemblance empirique $\mathcal{R}_{nk}(\boldsymbol{\beta}) / \sup_{\beta,\beta_2} \mathcal{R}_{nk}(\boldsymbol{\beta}, \boldsymbol{\beta}_2)$. Sans la contrainte $\sum_{i=1}^{k} q_i \mathbf{g}_i(\boldsymbol{\beta}) = \sum_{j=k+1}^{n} q_j \mathbf{g}_j(\boldsymbol{\beta}_2) = \mathbf{0}_p$, les maximums de $\prod_{i=1}^{k} q_i$, $\prod_{j=k+1}^{n} q_j$ sont atteints respectivement pour $q_i = k^{-1}$ et $q_j = (n-k)^{-1}$. Alors, on obtient que $\sup_{\beta,\beta_2} \mathcal{R}_{nk}(\boldsymbol{\beta}, \boldsymbol{\beta}_2) = k^{-k}(n-k)^{-(n-k)}$ [1], donc, le rapport du log-vraisemblance empirique est

$$-2 \sup_{(q_1,\ldots,q_k)} \sup_{(q_{k+1},\ldots,q_n)} \left\{ \sum_{i=1}^{k} \log(kq_i) + \sum_{j=k+1}^{n} \log((n-k)q_j); \sum_{i=1}^{k} q_i = \sum_{j=k+1}^{n} q_j = 1, \right.$$
$$\left. \sum_{i=1}^{k} q_i \mathbf{z}_i(\boldsymbol{\beta}) = \sum_{j=k+1}^{n} q_j \mathbf{z}_j(\boldsymbol{\beta}) = \mathbf{0}_p \right\}.$$

1. voir le papier de [Shi (2000)])

En utilisant la méthode des multiplicateurs de Lagrange, les probabilités optimales q_i et q_j sont

$$q_i = \frac{1}{k + n\boldsymbol{\lambda}_1^t \mathbf{z}_i(\boldsymbol{\beta})}, \qquad q_j = \frac{1}{n - k - n\boldsymbol{\lambda}_2^t \mathbf{z}_j(\boldsymbol{\beta})}, \tag{4.5}$$

où $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathbb{R}^p$ sont les multiplicateurs de Lagrange. Par conséquence, la fonction de la log-vraisemblance empirique correspondante peut s'écrire

$$2\sum_{i=1}^{k} \log\left[1 + \frac{n}{k}\boldsymbol{\lambda}_1^t \mathbf{z}_i(\boldsymbol{\beta})\right] + 2\sum_{j=k+1}^{n} \log\left[1 - \frac{n}{n-k}\boldsymbol{\lambda}_2^t \mathbf{z}_j(\boldsymbol{\beta})\right]. \tag{4.6}$$

Tenant compte de la relation (4.5), la dérivée de (4.6) par rapport à $\boldsymbol{\beta}$ est

$$2\sum_{i=1}^{k} q_i \mathbf{X}_i \mathbf{X}_i^t \boldsymbol{\lambda}_1 - 2\sum_{j=k+1}^{n} q_j \mathbf{X}_j \mathbf{X}_j^t \boldsymbol{\lambda}_2 = \mathbf{0}_p.$$

On applique le Lemme 4 de [Guo et al. (2013)] sur chaque phase du modèle, ce qui donne

$$\|\boldsymbol{\lambda}_1\| = O_{\mathbb{P}}(d^{1/2} k^{-1/2})$$

et

$$\|\boldsymbol{\lambda}_2\| = O_{\mathbb{P}}(d^{1/2} (n-k)^{-1/2}).$$

Alors, les probabilités q_i et q_j données par la relation (4.5) sont respectivement approximées par k^{-1} et $(n-k)^{-1}$. Ainsi, on peut restreindre $\boldsymbol{\lambda}_1$ et $\boldsymbol{\lambda}_2$ tel que

$$\frac{1}{k}\sum_{i=1}^{k} \mathbf{X}_i \mathbf{X}_i^t \boldsymbol{\lambda}_1 = \frac{1}{n-k}\sum_{j=k+1}^{n} \mathbf{X}_j \mathbf{X}_j^t \boldsymbol{\lambda}_2. \tag{4.7}$$

La plus petite valeur propre de chacune des matrices $k^{-1}\sum_{i=1}^{k} \mathbf{X}_i \mathbf{X}_i^t$ et $\frac{1}{n-k}\sum_{j=k+1}^{n} \mathbf{X}_j \mathbf{X}_j^t$ converge vers un nombre strictement positif, pour $n \to \infty$, alors la relation (4.7) peut s'écrire

$$\boldsymbol{\lambda}_1 = \left(\frac{1}{k}\sum_{i=1}^{k} \mathbf{X}_i \mathbf{X}_i^t\right)^{-1} \left(\frac{1}{n-k}\sum_{j=k+1}^{n} \mathbf{X}_j \mathbf{X}_j^t\right) \boldsymbol{\lambda}_2.$$

On note

$$\tilde{\boldsymbol{\lambda}}_2 \equiv \left(\frac{1}{k}\sum_{i=1}^{k} \mathbf{X}_i \mathbf{X}_i^t\right)^{-1} \left((n-k)^{-1}\sum_{j=k+1}^{n} \mathbf{X}_j \mathbf{X}_j^t\right) \boldsymbol{\lambda}_2,$$

et on obtient les nouveaux multiplicateurs de Lagrange, tel que $\boldsymbol{\lambda}_1 = \tilde{\boldsymbol{\lambda}}_2$. Pour des raisons de lisibilité, on note $\tilde{\boldsymbol{\lambda}}_2$ par $\boldsymbol{\lambda}_2$.

Avec cette remarque, on limite l'étude au cas particulier quand $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = \boldsymbol{\lambda}$. En considérant cette contrainte, à la place de la statistique donnée par la relation (4.6), on considère la statistique du rapport de vraisemblance empirique suivante

$$\mathrm{EL}_{nk}(\boldsymbol{\beta}) \equiv 2\sum_{i=1}^{k} \log\left(1 + \frac{n}{k}\boldsymbol{\lambda}^t \mathbf{z}_i(\boldsymbol{\beta})\right) + 2\sum_{j=k+1}^{n} \log\left(1 - \frac{n}{n-k}\boldsymbol{\lambda}^t \mathbf{z}_j(\boldsymbol{\beta})\right), \tag{4.8}$$

où le multiplicateur de Lagrange $\boldsymbol{\lambda} \in \mathbb{R}^p$ satisfait

$$\sum_{i=1}^{k} \frac{\mathbf{z}_i(\boldsymbol{\beta})}{k/n + \boldsymbol{\lambda}^t \mathbf{z}_i(\boldsymbol{\beta})} - \sum_{j=k+1}^{n} \frac{\mathbf{z}_j(\boldsymbol{\beta})}{1 - k/n - \boldsymbol{\lambda}^t \mathbf{z}_j(\boldsymbol{\beta})} = \mathbf{0}_p. \tag{4.9}$$

4.2.1 Notations

Tout d'abord, on introduit la notation $\theta_{nk} \equiv k/n$. Pour simplifier les notations, on utilise la notation θ à la place de θ_{nk}. On considère la notation suivante : $\mathrm{Mx}(\mathbf{A}) = \max_{1\leq i,j\leq p} |a_{ij}|$.

Sous H_0, on définit la matrice carrée suivante de dimension $p \times p$

$$\mathbf{S}_n(\boldsymbol{\beta^0}) \equiv \frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{z}_i(\boldsymbol{\beta}^0)\mathbf{z}_i^t(\boldsymbol{\beta}^0) + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{z}_j(\boldsymbol{\beta}^0)\mathbf{z}_j^t(\boldsymbol{\beta}^0) \tag{4.10}$$

et le $p-$vecteur suivant

$$\boldsymbol{\psi}_n(\boldsymbol{\beta^0}) \equiv \frac{1}{n\theta}\sum_{i=1}^{k} \mathbf{z}_i(\boldsymbol{\beta}^0) - \frac{1}{n(1-\theta)}\sum_{j=k+1}^{n} \mathbf{z}_j(\boldsymbol{\beta}^0), \tag{4.11}$$

où $\boldsymbol{\beta^0}$ désigne la vraie valeur du paramètre $\boldsymbol{\beta}$ pour la première phase, pour des observations de 1 à k.

Sous l'hypothèse nulle, pour la vraie valeur $\boldsymbol{\beta^0}$ de $\boldsymbol{\beta}$, la moyenne de la matrice aléatoire $\mathbf{S}_n(\boldsymbol{\beta^0})$ est la matrice carré suivante de dimension $p \times p$

$$\mathbf{V}_n^0 \equiv \frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{V}_{(i)}^0 + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{V}_{(j)}^0, \tag{4.12}$$

où, pour $i = 1, \ldots, n$

$$\mathbf{V}_{(i)}^0 \equiv \mathbb{V}\mathrm{ar}\,(\mathbf{z}_i^0) = \sigma^2 \mathbf{X}_i \mathbf{X}_i^t. \tag{4.13}$$

Pour $i = 1, \ldots, n$, on considère le vecteur aléatoire suivant

$$\mathbf{w}_i^0 \equiv (\mathbf{V}_n^0)^{-1/2} \mathbf{z}_i(\boldsymbol{\beta^0}).$$

Correspondant aux composantes de $\mathbf{w}_i^0 = (w_{i,1}^0, \ldots, w_{i,p}^0)$, pour $i = 1, \ldots, n$, pour $r \in \mathbb{N}^*$, $t_1, \ldots, t_r \in \{1, \ldots, p\}$, on considère le scalaire suivant

$$\alpha^{t_1 t_2 \ldots t_r} \equiv \frac{1}{n\theta^r} \sum_{i=1}^{k} \mathbb{E}\Big[w_{i,t_1}^0 w_{i,t_2}^0 \ldots w_{i,t_r}^0 \Big] + \frac{1}{n(\theta-1)^r} \sum_{j=k+1}^{n} \mathbb{E}\Big[w_{j,t_1}^0 w_{j,t_2}^0 \ldots w_{j,t_r}^0 \Big] \tag{4.14}$$

et la variable aléatoire suivante

$$\omega^{t_1 t_2 \ldots t_r} \equiv \frac{1}{n\theta^r} \sum_{i=1}^{k} w_{i,t_1}^0 w_{i,t_2}^0 \ldots w_{i,t_r}^0 + \frac{1}{n(\theta-1)^r} \sum_{j=k+1}^{n} w_{j,t_1}^0 w_{j,t_2}^0 \ldots w_{j,t_r}^0 - \alpha^{t_1 t_2 \ldots t_r}, \tag{4.15}$$

où w_{i,t_r}^0 est le r-ème composante du vecteur $\mathbf{w}_i^0$. En particulier, pour tout $t_1, t_2 \in \{1, \ldots, p\}$, $\alpha^{t_1} = 0$, et $\alpha^{t_1 t_2} = \delta^{t_1 t_2}$ désigne le symbole de Kronecker, c'est-à-dire $\alpha^{t_1 t_2} = 1$ si $t_1 = t_2$, et 0 sinon.

4.2.2 Suppositions

Dans cette sous-section, nous énonçons les suppositions sur le design, les erreurs, le nombre des variables explicatives p et sur la localisation du point de rupture. Ces suppositions sont nécessaires afin de garder les propriétés obtenues pour la statistique de vraisemblance empirique pour le modèle en grande dimension sans point de rupture.

Pour les suppositions (A3)-(A6), q désigne une constante tel que $q \geq 4$.

(A1) Il existe deux constantes positives C_0 et C_1, tel que $0 < C_0 < \inf_n \gamma_1(\mathbf{V}_n^0) \leq \sup_n \gamma_1(\mathbf{V}_n^0) < C_1 < \infty$.

(A2) $\mathbb{E}(\varepsilon_1^4) < C_2$, pour $C_2 > 0$ et pour tout n.

(A3) $p^{-1} \sum_{s=1}^{p} |X_{i,s}|^q < C_3, 1 \leq i \leq n$, pour $C_3 > 0$, et $q \geq 4$.

(A4) $\mathbb{E}|\epsilon_1|^{2q} < C_4$, pour $C_4 > 0$.

(A5) $p\ k^{(2-q)/(2q)} \to 0$ et $p\ (n-k)^{(2-q)/(2q)} \to 0$, quand $n \to \infty$.

(A6) $p^{2+4/q}\ k^{-1} \to 0$ et $p^{2+4/q}\ (n-k)^{-1} \to 0$, quand $n \to \infty$.

(A7) $\sum_{r,s=1}^{p} \alpha^{rrss} = O(p^2)$.

(A8) $\sum_{r,s,u=1}^{p} \alpha^{rsu}\alpha^{rsu} = O(p^{5/2})$ et $\sum_{r,s,u=1}^{p} \alpha^{rss}\alpha^{suu} = O(p^{5/2})$.

(A9) Pour tout $i = 1, \ldots, n$, pour $l \in \mathbb{N}^*$, $j_1, \ldots, j_l \in \{1, \ldots, p\}$, et pour $\sum_{i=1}^{l} d_i \leq 6$, il existe une constante positive $C_6 < \infty$, telle que $\mathbb{E}(w_{i,j_1}^{d_1} \ldots w_{i,j_l}^{d_l}) \leq C_6$.

Les suppositions (A3) et (A6) garantissent que les valeurs propres de $\mathbf{S}_n^0$ sont proches du celles de $\mathbf{V}_n^0$. La supposition (A1) implique que $\mathbf{V}_n^0$ est uniformément inversible et bornée pour une valeur de n assez grande. Alors, avec une probabilité proche à 1, $\mathbf{S}_n^0$ est inversible et $0 < C_0 < \gamma_p(\mathbf{S}_n^0) \leq \gamma_1(\mathbf{S}_n^0) < C_1 < \infty$. La supposition (A3) est également imposée par [Guo et al. (2013)], [Hjort et al. (2009)] et [Liu et al(2013)] pour un modèle en grande dimension sans point de rupture. La supposition (A4) ainsi que les suppositions (A3) et (A6) impliquent que $\sup_{1\leq i\leq n} |\boldsymbol{\lambda}^t \mathbf{z}_i^0| = o_{\mathbb{P}}(1)$, ce qui conduit à des développements de Taylor pour (4.8) et (4.9) (voir le Lemme 4.5.6). Les suppositions (A1), (A2), (A4) sont également utilisées par [Guo et al. (2013)] pour les modèles linéaires sans point de rupture avec des régresseurs aléatoires. Les suppositions (A5)-(A9) sont également imposées par [Guo et al. (2013)] et [Liu et al(2013)], afin d'avoir la normalité asymptotique de la statistique du rapport de vraisemblance empirique.

4.3 Résultat principal

Dans cette section, on présente le résultat principal de ce chapitre. La distribution asymptotique de la statistique de test du rapport de vraisemblance empirique sous l'hypothèse H_0 nous permet de tester si le modèle change après l'observation k.

Notons que, pour une valeur fixée de p, la loi asymptotique obtenue par [Liu et al. (2008)] et [Zi et al. (2010)] pour les modèles linéaire et dans le Chapitre 3 pour le modèle non-linéaire

était une loi de χ^2 avec p degrés de liberté. Dans notre cas, on obtient une autre statistique de test avec une distribution asymptotique normale.

Afin d'obtenir la distribution asymptotique, on a besoin de quelques résultats intermédiaires pour étudier les comportements asymptotiques de la statistique du rapport de vraisemblance empirique.

Nous soulignons que la présence du point de rupture k complique l'étude et conduit à une approche différente par rapport au modèle sans point de rupture.

Notons que sous l'hypothèse H_0, on a :

$$\mathbf{z}_i(\boldsymbol{\beta^0}) = \mathbf{z}_i^0 = \mathbf{X}_i\varepsilon_i,$$

alors que sous H_1, pour $i = k+1, \ldots, n$, le vecteur $\mathbf{z}_i(\boldsymbol{\beta^0})$, devient

$$\mathbf{z}_i(\boldsymbol{\beta^0}) = \mathbf{X}_i\mathbf{X}_i^t(\boldsymbol{\beta}_2^0 - \boldsymbol{\beta^0}) - \mathbf{z}_i^0. \tag{4.16}$$

Quand H_0 est vraie, on note la matrice $\mathbf{S}_n(\boldsymbol{\beta^0})$ par $\mathbf{S}_n^0$, telle que

$$\mathbf{S}_n^0 \equiv \frac{1}{n\theta^2}\sum_{i=1}^{k}\mathbf{z}_i^0\mathbf{z}_i^{0t} + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n}\mathbf{z}_j^0\mathbf{z}_j^{0t} \tag{4.17}$$

et le vecteur $\boldsymbol{\psi}_n(\boldsymbol{\beta^0})$ par $\boldsymbol{\psi}_n^0$, tel que

$$\boldsymbol{\psi}_n^0 \equiv \frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{z}_i^0 - \frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{z}_j^0. \tag{4.18}$$

Le multiplicateur de Lagrange $\boldsymbol{\lambda}$ est un élément important dans la formulation de la vraisemblance empirique. Le premier résultat concerne la vitesse de convergence de $\boldsymbol{\lambda}$ définie dans la relation (4.9). Quand p est fixé, [Zi et al. (2010)] ont montré que

$$||\boldsymbol{\lambda}|| = O_{\mathbb{P}}(n^{-1/2}). \tag{4.19}$$

Par la proposition suivante, on montre que quand p croit avec n, la vitesse de convergence de $||\boldsymbol{\lambda}||$ donnée par la relation (4.19) n'est plus valable.

Proposition 4.3.1. *On suppose que (A1), (A3)-(A6) sont vérifiées. Alors, sous l'hypothèse* H_0*, le multiplicateur de Lagrange* $\boldsymbol{\lambda}$ *satisfait*

$$||\boldsymbol{\lambda}|| = O_{\mathbb{P}}(p^{1/2}n^{-1/2}).$$

Preuve. On écrit $\boldsymbol{\lambda}$ sous la forme suivante : $\boldsymbol{\lambda} = ||\boldsymbol{\lambda}||\mathbf{u}$, où $\mathbf{u}$ est un $p-$vecteur et ayant la norme égale à 1. En utilisant le Lemme 4.5.3, pour $\theta = k/n$, pour chaque observation $i = 1, \ldots, k$, on a :

$$0 < 1 + \frac{||\boldsymbol{\lambda}||}{\theta}\mathbf{u}^t\mathbf{z}_i^0 \leq 1 + ||\boldsymbol{\lambda}||T_n^0, \tag{4.20}$$

avec une probabilité égale à 1, où,

$$T_n^0 \equiv \max_{\substack{1\leqslant i\leqslant k \\ k+1\leqslant j\leqslant n}} \left\{ \frac{n}{k}||\mathbf{z}_i^0||, \frac{n}{n-k}||\mathbf{z}_j^0|| \right\}.$$

Pour $j = k+1, \ldots, n$, on a

$$0 < 1 - \frac{||\boldsymbol{\lambda}||}{1-\theta}\mathbf{u}^t\mathbf{z}_j^0 \leq 1 + ||\boldsymbol{\lambda}||T_n^0, \tag{4.21}$$

avec une probabilité égale à 1.

En utilisant les relations (4.20) et (4.21), par la relation (4.9), on obtient que

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{k}\frac{\mathbf{u}^t\mathbf{z}_i^0}{\theta + \boldsymbol{\lambda}^t\mathbf{z}_i^0} - \frac{1}{n}\sum_{j=k+1}^{n}\frac{\mathbf{u}^t\mathbf{z}_j^0}{1-\theta-\boldsymbol{\lambda}^t\mathbf{z}_j^0} \\ &= \frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{u}^t\mathbf{z}_i^0 - \frac{1}{n\theta^2}||\boldsymbol{\lambda}||\sum_{i=1}^{k}\frac{\mathbf{u}^t\mathbf{z}_i^0\mathbf{z}_i^{0t}\mathbf{u}}{1+\theta^{-1}||\boldsymbol{\lambda}||\mathbf{u}^t\mathbf{z}_i^0} - \frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{u}^t\mathbf{z}_j^0 \\ &\quad -\frac{1}{n(1-\theta)^2}||\boldsymbol{\lambda}||\sum_{j=k+1}^{n}\frac{\mathbf{u}^t\mathbf{z}_j^0\mathbf{z}_j^{0t}\mathbf{u}}{1-(1-\theta)^{-1}||\boldsymbol{\lambda}||\mathbf{u}^t\mathbf{z}_j^0}. \end{aligned} \tag{4.22}$$

En utilisant la relation (4.22), ainsi que les notations données par les relations (4.17) et (4.18), on obtient que

$$0 \leq \mathbf{u}^t\boldsymbol{\psi}_n^0 - \frac{||\boldsymbol{\lambda}||}{1+||\boldsymbol{\lambda}||T_n^0}\ \mathbf{u}^t\mathbf{S}_n^0\mathbf{u}.$$

Alors, on a que

$$\mathbf{u}^t\boldsymbol{\psi}_n^0(1+||\boldsymbol{\lambda}||T_n^0) \geq ||\boldsymbol{\lambda}||\mathbf{u}^t\mathbf{S}_n^0\mathbf{u},$$

avec une probabilité égale à 1. Donc,

$$||\boldsymbol{\lambda}|| \leq \frac{\mathbf{u}^t\boldsymbol{\psi}_n^0}{\mathbf{u}^t\mathbf{S}_n^0\mathbf{u} - \mathbf{u}^t\boldsymbol{\psi}_n^0 T_n^0}. \tag{4.23}$$

D'autre part, on a

$$|\mathbf{u}^t \boldsymbol{\psi}_n^0| \leq ||\boldsymbol{\psi}_n^0||.$$

En utilisant le Lemme 4.5.4, on obtient que

$$||\boldsymbol{\psi}_n^0|| = O_{\mathbb{P}}(n^{-1/2}p^{1/2}),$$

ce qui donne

$$\mathbf{u}^t \boldsymbol{\psi}_n^0 = O_{\mathbb{P}}(n^{-1/2}p^{1/2}). \tag{4.24}$$

En utilisant le Lemme 4.5.3 et la relation (4.24), on a que

$$\mathbf{u}^t \boldsymbol{\psi}_n^0 T_n^0 = O_{\mathbb{P}}(n^{-1/2}p^{1/2}) o_{\mathbb{P}}(n^{1/q}p^{1/2}) = o_{\mathbb{P}}(n^{(-q+2)/2q}p).$$

Alors, par la supposition (A5), on obtient que

$$\mathbf{u}^t \boldsymbol{\psi}_n^0 T_n^0 = o_{\mathbb{P}}(1).$$

D'autre part, d'après le Lemme 4.5.2, l'inégalité $\mathbf{u}^t \mathbf{S}_n^0 \mathbf{u} \geq \gamma_p(\mathbf{S}_n^0) > C_0 > 0$ est valable avec une probabilité qui tend vers 1 quand $n \to \infty$. Alors, pour la relation (4.23), on obtient que

$$||\boldsymbol{\lambda}|| = O_{\mathbb{P}}(|\mathbf{u}^t \boldsymbol{\psi}_n^0|/C_0) = O_{\mathbb{P}}(||\boldsymbol{\psi}_n^0||) = O_{\mathbb{P}}(p^{1/2}n^{-1/2}).$$

■

D'après la Proposition 4.3.1, par la supposition (A6), on a que

$$||\boldsymbol{\lambda}|| \xrightarrow{\mathbb{P}} 0, \text{ quand } n \to \infty.$$

Plus précisément, le multiplicateur de Lagrange $\boldsymbol{\lambda}$ a la forme approchée donnée par la Proposition 4.3.2.

Nous rappelons que la matrice carré $\mathbf{V}_n^0$ et le vecteur $\boldsymbol{\psi}_n^0$ sont donnés respectivement par les relations (4.12) et (4.18).

Proposition 4.3.2. *Si les suppositions (A1), (A3)-(A6) sont vérifiées, alors, sous l'hypothèse nulle* H_0*, on a*

$$\boldsymbol{\lambda} = (\mathbf{V}_n^0)^{-1} \boldsymbol{\psi}_n^0 (1 + o_{\mathbb{P}}(1)).$$

Preuve. Par le Lemme 4.5.6 on a que

$$\boldsymbol{\lambda} = (\mathbf{S}_n^0)^{-1} \left(\mathbf{R}_n^0 + \boldsymbol{\psi}_n^0\right) (1 + o_{\mathbb{P}}(1)).$$

D'autre part, par le Lemme 4.5.4, on a que

$$||\boldsymbol{\psi}_n^0|| = O_{\mathbb{P}}(p^{1/2} n^{-1/2}).$$

En utilisant ce fait et la relation (4.90), on obtient

$$\frac{||\mathbf{R}_n^0||}{||\boldsymbol{\psi}_n^0||} = o_{\mathbb{P}}(n^{1/q} p^{1/2} n^{-1/2} p^{1/2}) = o_{\mathbb{P}}(p n^{(2-q)/2q}).$$

Donc, par la supposition (A5), on obtient que

$$\mathbf{R}_n^0 = \boldsymbol{\psi}_n^0 o_{\mathbb{P}}(1).$$

Alors

$$\boldsymbol{\lambda} = (\mathbf{S}_n^0)^{-1} \boldsymbol{\psi}_n^0 (1 + o_{\mathbb{P}}(1)).$$

D'autre part, par le Lemme 4.5.7, on a que

$$\left((\mathbf{S}_n^0)^{-1} - (\mathbf{V}_n^0)^{-1}\right) \boldsymbol{\psi}_n^0 = (\mathbf{V}_n^0)^{-1} \boldsymbol{\psi}_n^0 o_{\mathbb{P}}(1).$$

Alors,

$$(\mathbf{S}_n^0)^{-1} \boldsymbol{\psi}_n^0 = (\mathbf{V}_n^0)^{-1} \boldsymbol{\psi}_n^0 (1 + o_{\mathbb{P}}(1)).$$

Donc, pour $\boldsymbol{\lambda} = (\mathbf{S}_n^0)^{-1} \boldsymbol{\psi}_n^0 (1 + o_{\mathbb{P}}(1))$, on obtient que

$$\boldsymbol{\lambda} = (\mathbf{V}_n^0)^{-1} \boldsymbol{\psi}_n^0 (1 + o_{\mathbb{P}}(1)). \tag{4.25}$$

■

Sous l'hypothèse H_0, on donne par les deux propositions suivantes deux approximations pour la statistique du rapport de vraisemblance empirique $\mathrm{EL}_{nk}(\boldsymbol{\beta^0})$, définie par la relation (4.8), ce qui nous permet de trouver la distribution asymptotique de cette statistique.

Proposition 4.3.3. *Si les suppositions (A1)-(A8) sont vérifiées, alors, sous l'hypothèse nulle* H_0*, on a*

$$\mathrm{EL}_{nk}(\boldsymbol{\beta^0}) = n \boldsymbol{\psi}_n^{0t} (\mathbf{S}_n^0)^{-1} \boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(p^{1/2}).$$

Preuve. Par le Lemme 4.5.5 on a que

$$\mathrm{EL}_{nk}(\boldsymbol{\beta^0}) = 2n\boldsymbol{\lambda}^t\boldsymbol{\psi}_n^0 - n\boldsymbol{\lambda}^t\mathbf{S}_n^0\boldsymbol{\lambda} + \mathcal{E}_3 + o_{\mathbb{P}}(1), \tag{4.26}$$

avec

$$\mathcal{E}_3 \equiv \frac{2}{3}\left[\frac{1}{\theta^3}\sum_{i=1}^{k}(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^3 - \frac{1}{(1-\theta)^3}\sum_{j=k+1}^{n}(\mathbf{z}_j^{0t}\boldsymbol{\lambda})^3\right].$$

On considère le $d-$vecteur suivant

$$\mathbf{R}_n^0 \equiv \frac{1}{n\theta^3}\sum_{i=1}^{k}\mathbf{z}_i^0(\boldsymbol{\lambda}^t\mathbf{z}_i^0)^2 - \frac{1}{n(1-\theta)^3}\sum_{j=k+1}^{n}\mathbf{z}_j^0(\boldsymbol{\lambda}^t\mathbf{z}_j^0)^2. \tag{4.27}$$

Par le Lemme 4.5.6, on a que

$$\boldsymbol{\lambda} = (\mathbf{S}_n^0)^{-1}(\mathbf{R}_n^0 + \boldsymbol{\psi}_n^0)(1 + o_{\mathbb{P}}(1)).$$

Alors, pour (4.26) on a que

$$\mathrm{EL}_{nk}(\boldsymbol{\beta^0}) = n\boldsymbol{\psi}_n^{0t}(\mathbf{S}_n^0)^{-1}\boldsymbol{\psi}_n^0 - n\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0 + \mathcal{E}_3 + o_{\mathbb{P}}(1). \tag{4.28}$$

En parallèle, on étudie les deux termes $\mathcal{E}_3$ et $n(\mathbf{R}_n^0)^t(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0$.

Par la Proposition 4.3.2, on a que

$$\boldsymbol{\lambda} = (\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0(1 + o_{\mathbb{P}}(1)),$$

ce qui implique

$$\mathbf{z}_i^{0t}\boldsymbol{\lambda} = \mathbf{z}_i^{0t}(\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0(1 + o_{\mathbb{P}}(1)). \tag{4.29}$$

Alors, $\mathcal{E}_3$ devient

$$\begin{aligned}\mathcal{E}_3 \;=\; & \frac{2}{3\theta^3}\sum_{i=1}^{k}\left[\mathbf{z}_i^{0t}(\mathbf{V}_n^0)^{-1/2}(\mathbf{V}_n^0)^{-1/2}\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{z}_i^0 + \frac{1}{n(\theta-1)}\sum_{j=k+1}^{n}\mathbf{z}_j^0\Big)\right]^3 \\ & + \frac{2}{3(\theta-1)^3}\sum_{j=k+1}^{n}\left[\mathbf{z}_j^{0t}(\mathbf{V}_n^0)^{-1/2}(\mathbf{V}_n^0)^{-1/2}\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{z}_i^0 + \frac{1}{n(\theta-1)}\sum_{j=k+1}^{n}\mathbf{z}_j^0\Big)\right]^3(1 + o_{\mathbb{P}}(1)).\end{aligned}$$

En utilisant les notations données par les relations (4.14), (4.15) et la loi des grands nombres (Théorème de Markov), on obtient

$$\begin{aligned}
\mathcal{E}_3 &= \frac{2}{\theta^3}\sum_{l=1}^{k}\Big[\mathbf{w}_l^{0t}\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{w}_i^0+\frac{1}{n(\theta-1)}\sum_{j=k+1}^{n}\mathbf{w}_j^0\Big)\Big]^3 \\
&\quad+\frac{2}{(1-\theta)^3}\sum_{l=k+1}^{n}\Big[\mathbf{w}_l^{0t}\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{w}_i^0+\frac{1}{n(\theta-1)}\sum_{j=k+1}^{n}\mathbf{w}_j^0\Big)\Big]^3(1+o_{\mathbb{P}}(1)) \\
&= \frac{2n}{3}\sum_{r,s,u=1}^{p}\omega^r\omega^s\omega^u\Big[\frac{1}{n\theta^3}\sum_{i=1}^{k}w_{i,r}^0w_{i,s}^0w_{i,u}^0+\frac{1}{n(\theta-1)^3}\sum_{j=k+1}^{n}w_{j,r}^0w_{j,s}^0w_{j,u}^0\Big](1+o_{\mathbb{P}}(1)) \\
&= \frac{2n}{3}\sum_{r,s,u=1}^{p}\omega^r\omega^s\omega^u\alpha^{rsu}(1+o_{\mathbb{P}}(1)). \qquad (4.30)
\end{aligned}$$

D'autre part, en remplaçant la valeur obtenue pour $\boldsymbol{\lambda}$ par la Proposition 4.3.2 dans la relation (4.27), on obtient

$$\mathbf{R}_n^0=\Big[\frac{1}{n\theta^3}\sum_{i=1}^{k}\mathbf{z}_i^0\Big(\mathbf{z}_i^{0t}\boldsymbol{\psi}_n^{0t}(\mathbf{V}_n^0)^{-1}\Big)^2-\frac{1}{n(1-\theta)^3}\sum_{j=k+1}^{n}\mathbf{z}_j^0\Big(\mathbf{z}_j^{0t}\boldsymbol{\psi}_n^{0t}(\mathbf{V}_n^0)^{-1}\Big)^2\Big](1+o_{\mathbb{P}}(1)). \quad (4.31)$$

Pour $n(\mathbf{R}_n^0)^t(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0$, on considère la décomposition suivante

$$\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0=\mathbf{R}_n^{0t}\big[(\mathbf{S}_n^0)^{-1}-(\mathbf{V}_n^0)^{-1}\big]\mathbf{R}_n^0+\mathbf{R}_n^{0t}(\mathbf{V}_n^0)^{-1}\mathbf{R}_n^0. \quad (4.32)$$

Par le Lemme 4.5.7*(ii)*, on a que

$$\big[(\mathbf{S}_n^0)^{-1}-(\mathbf{V}_n^0)^{-1}\big]\mathbf{R}_n^0=(\mathbf{V}_n^0)^{-1}\mathbf{R}_n^0o_{\mathbb{P}}(1).$$

Alors, la relation (4.32) devient

$$\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0=\mathbf{R}_n^{0t}(\mathbf{V}_n^0)^{-1}\mathbf{R}_n^0(1+o_{\mathbb{P}}(1)).$$

En utilisant les relations (4.18), (4.27), (4.31) et le fait que pour tout $i=1,\ldots,n$,$\mathbf{w}_i^0=$

$(\mathbf{V}_n^0)^{-1/2}\mathbf{z}_i^0$, alors $n\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0$ peut s'écrire

$$\begin{aligned} n\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0 &= \Bigg\{\frac{1}{\theta^3}\sum_{l=1}^{k}\mathbf{w}_l^0\Big[\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{w}_i^0-\frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{w}_j^0\Big)^t\mathbf{w}_l^0\Big]^2 \\ &\quad -\frac{1}{(1-\theta)^3}\sum_{l=k+1}^{n}\mathbf{w}_l^0\Big[\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{w}_i^0-\frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{w}_j^0\Big)^t\mathbf{w}_l^0\Big]^2 \\ &\quad \cdot\Bigg(\frac{1}{n\theta^3}\sum_{l=1}^{k}\mathbf{w}_i^0\Big[\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{w}_i^0-\frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{w}_j^0\Big)^t\mathbf{w}_l^0\Big]^2 \\ &\quad -\frac{1}{n(1-\theta)^3}\sum_{l=k+1}^{n}\mathbf{w}_l^0\Big[\Big(\frac{1}{n\theta}\sum_{i=1}^{k}\mathbf{w}_i^0-\frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{w}_j^0\Big)^t\mathbf{w}_l^0\Big]^2\Bigg)\Bigg\}(1+o_{\mathbb{P}}(1)). \end{aligned}$$

Avec les notations données par (4.14) et (4.15), on obtient

$$\begin{aligned} n\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0 &= n\sum_{r,s,l,u,v=1}^{p}\omega^r\omega^s\Big[\frac{1}{n\theta^3}\sum_{i=1}^{k}w_{i,r}^0w_{i,s}^0w_{i,l}^0-\frac{1}{n(1-\theta)^3}\sum_{j=k+1}^{n}w_{j,r}^0w_{j,s}^0w_{j,l}^0\Big] \\ &\quad \cdot\omega^u\omega^v\Big[\frac{1}{n\theta^3}\sum_{i=1}^{k}w_{i,u}^0w_{i,v}^0w_{i,l}^0-\frac{1}{n(1-\theta)^3}\sum_{j=k+1}^{n}w_{j,u}^0w_{j,v}^0w_{j,l}^0\Big](1+o_{\mathbb{P}}(1)) \\ &= n\sum_{r,s,l,u,v=1}^{p}\alpha^{rsl}\alpha^{uvl}\omega^r\omega^s\omega^u\omega^v(1+o_{\mathbb{P}}(1)). \end{aligned}$$

En conclusion, pour le terme $\mathcal{E}_3$ de la relation (4.30) et pour $n\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0$, en utilisant les suppositions (A6), (A7) et (A8) ainsi la preuve de la Proposition 1 de [Guo et al. (2013)], on obtient :

$$\mathcal{E}_3 = o_{\mathbb{P}}(p^{1/2})$$

et

$$n\mathbf{R}_n^{0t}(\mathbf{S}_n^0)^{-1}\mathbf{R}_n^0 = o_{\mathbb{P}}(p^{1/2}).$$

En combinant ces deux dernières relations avec la relation (4.28), on obtient que

$$\mathrm{EL}_{nk}(\boldsymbol{\beta}^{\mathbf{0}}) = n\boldsymbol{\psi}_n^0(\mathbf{S}_n^0)^{-1}\boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(p^{1/2}).$$

■

Proposition 4.3.4. *Sous l'hypothèse* H_0*, si les suppositions (A1), (A3), (A4), (A6) et (A7) sont vérifiées, on a*

$$\mathrm{EL}_n(\boldsymbol{\beta^0}) = n\boldsymbol{\psi}_n^{0t}(\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(p^{1/2}).$$

Preuve. Tout d'abord, on prouve que

$$n\boldsymbol{\psi}_n^{0t}\left[(\mathbf{V}_n^0)^{-1} - (\mathbf{S}_n^0)^{-1}\right]\boldsymbol{\psi}_n^0 = o_{\mathbb{P}}(p^{1/2}). \tag{4.33}$$

Pour ceci, on introduit les deux matrices carrées suivantes de dimension $p \times p$

$$\mathbf{B}_n^0 \equiv (\mathbf{V}_n^0)^{-1/2}\,\mathbf{S}_n^0\,(\mathbf{V}_n^0)^{-1/2}$$

et

$$\mathbf{K}_n^0 \equiv \mathbf{I}_p - \mathbf{B}_n^0$$

et le $p-$vecteur suivant

$$\boldsymbol{\eta}_n^0 \equiv (\mathbf{V}_n^0)^{-1/2}\boldsymbol{\psi}_n^0.$$

En utilisant les notations données ci-dessus, la partie droite de la relation (4.33), peut s'écrire

$$n\boldsymbol{\psi}_n^{0t}\left[(\mathbf{V}_n^0)^{-1} - (\mathbf{S}_n^0)^{-1}\right]\boldsymbol{\psi}_n^0 = n\boldsymbol{\eta}_n^{0t}\left[\mathbf{I}_p - (\mathbf{V}_n^0)^{1/2}(\mathbf{S}_n^0)^{-1}(\mathbf{V}_n^0)^{1/2}\right]\boldsymbol{\eta}_n^0 = n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^{-1}\boldsymbol{\eta}_n^0.$$

Pour $n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^{-1}\boldsymbol{\eta}_n^0$, on considère la décomposition suivante

$$\begin{aligned} n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^{-1}\boldsymbol{\eta}_n^0 \;\; = \;\; & \left[n\boldsymbol{\eta}_n^{0t}\mathbf{K}_n^0\boldsymbol{\eta}_n^0 - n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^2\boldsymbol{\eta}_n^0 - \ldots - (-1)^b n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\boldsymbol{\eta}_n^0\right] \\ & + (-1)^b n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b[\mathbf{I}_d - (\mathbf{B}_n^0)^{-1}]\boldsymbol{\eta}_n^0, \end{aligned} \tag{4.34}$$

pour tout $b \in \mathbb{N}^*$.

On étudie la convergence de la décomposition donnée par la relation (4.34). Par le Lemme 6 de [Chen et al. (2009)], si $\mathbf{A}$ est une matrice symétrique, alors on a l'inégalité suivante

$$\boldsymbol{\eta}_n^{0t}\mathbf{A}\boldsymbol{\eta}_n^0 \leq ||\boldsymbol{\eta}_n^0||^2\{\mathrm{tr}(\mathbf{A}^2)\}^{1/2}.$$

Pour le premier terme du membre droit de la relation (4.34), on a que

$$n\boldsymbol{\eta}_n^{0t}\mathbf{K}_n^0\boldsymbol{\eta}_n^0 \leq n\;||\boldsymbol{\eta}_n^0||^2\left[\mathrm{tr}(\mathbf{K}_n^0)^2\right]^{1/2}, \tag{4.35}$$

avec une probabilité égale à 1.

En utilisant la supposition (A1) et le Lemme 4 de [Liu et al(2013)], on obtient que,

$$||\boldsymbol{\eta}_n^0||_2^2 = \boldsymbol{\psi}_n^{0t}(\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0 \leq \frac{1}{\gamma_1(\mathbf{V}_n^0)}||\boldsymbol{\psi}_n^0||_2^2 \leq \frac{1}{C_0}||\,\boldsymbol{\psi}_n^0||_2^2.$$

avec une probabilité égale à 1.

Par le Lemme 4.5.4, on a que

$$||\boldsymbol{\psi}_n^0|| = O_{\mathbb{P}}(p^{1/2}n^{-1/2}),$$

alors

$$||\boldsymbol{\eta}_n^0||_2^2 = O_{\mathbb{P}}(p\,n^{-1}). \tag{4.36}$$

Par le Lemme 4.5.8, pour $1 \leq r \leq p$, on a

$$\gamma_r(\mathbf{K}_n^0) \leq [\mathrm{tr}(\mathbf{K}_n^0)^2]^{1/2} = O_{\mathbb{P}}(p\,n^{-1/2}). \tag{4.37}$$

En utilisant les relations (4.36), (4.37) et la condition $p = o(n^{1/2})$ obtenue par la supposition (A6), pour la relation (4.35), on a que

$$n\boldsymbol{\eta}_n^{0t}\mathbf{K}_n^0\boldsymbol{\eta}_n^0 = n\ O_{\mathbb{P}}(n^{-1}p)\ O_{\mathbb{P}}(p\,n^{-1/2}) = O_{\mathbb{P}}(p^2\ n^{-1/2}) = o_{\mathbb{P}}(p^{1/2}). \tag{4.38}$$

D'autre part, en utilisant les relations (4.36) et (4.37), on obtient

$$|\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\boldsymbol{\eta}_n^0| \leq ||\boldsymbol{\eta}_n^0||_2^2 \max_{1\leq r\leq p}|\gamma_r(\mathbf{K}_n^0)^b| \leq ||\boldsymbol{\eta}_n^0||_2^2[\mathrm{tr}(\mathbf{K}_n^0)^2]^{b/2} = O_{\mathbb{P}}(p\,n^{-1})\ O_{\mathbb{P}}(p^b n^{-b/2}),$$

ce qui donne

$$\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\boldsymbol{\eta}_n^0 = O_{\mathbb{P}}(p^{b+1}n^{-(b+2)/2}).$$

La dernière équation signifie que $n\sum_{b=1}^{\infty}(-1)^{b-1}\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\boldsymbol{\eta}_n^0$ est convergente pour une valeur fixée de n, quand $p = o(n^{1/2})$.

Alors, tenant compte de la relation (4.38), on peut conclure que

$$n\sum_{b=1}^{\infty}(-1)^{b-1}\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\boldsymbol{\eta}_n^0 = o_{\mathbb{P}}(p^{1/2}). \tag{4.39}$$

Dans la suite, on prouve que le terme $n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b[\mathbf{I}_p - (\mathbf{B}_n^0)^{-1}]\,\boldsymbol{\eta}_n^0$ de la relation (4.34) est négligeable quand $b \to \infty$.

Pour le dernier terme de la relation (4.34), on a que

$$|n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b[\mathbf{I}_d - (\mathbf{B}_n^0)^{-1}]\,\boldsymbol{\eta}_n^0| \leq |n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\boldsymbol{\eta}_n^0| + |n(\boldsymbol{\eta}_n^0)^t(\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}\boldsymbol{\eta}_n^0|. \tag{4.40}$$

Pour le premier terme du membre droit de la relation (4.40), par la relation (4.3), on a

$$\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\,\boldsymbol{\eta}_n^0 = O_{\mathbb{P}}(p^{b+1}n^{(-b-2)/2}),$$

avec une probabilité égale à 1. Alors

$$n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b\boldsymbol{\eta}_n^0 = O_{\mathbb{P}}(p^{b+1}n^{-b/2}). \tag{4.41}$$

Pour le deuxième terme du membre droit de la relation (4.40), on a

$$|n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}\boldsymbol{\eta}_n^0| \leq n||\boldsymbol{\eta}_n^0||_2^2\mathrm{M}\mathrm{K}((\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}), \tag{4.42}$$

avec une probabilité égale à 1. Donc, d'après le Lemme 4 de [Liu et al(2013)], si $\mathbf{A} = (a_{ij})$ est matrice symétrique de dimension $p \times p$, alors

$$\mathrm{M}\mathrm{K}(\mathbf{A}) \leq \max_{1\leq i\leq p}|\gamma_i(\mathbf{A})|.$$

Donc

$$\begin{aligned}\mathrm{M}\mathrm{K}((\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}) &\leq d\cdot\mathrm{M}\mathrm{K}(\mathbf{K}_n^0)^b\cdot\mathrm{M}\mathrm{K}(\mathbf{B}_n^0)^{-1}\\ &\leq p\cdot\max_{1\leq r\leq d}|\gamma_r(\mathbf{K}_n^0)^b|\cdot\max_{1\leq r\leq d}|\gamma_r(\mathbf{B}_n^0)^{-1}|.\end{aligned}$$

Par le Lemme 4.5.8, on sait que

$$|\gamma_r(\mathbf{K}_n^0)| \leq [\mathrm{tr}(\mathbf{K}_n^0)^2]^{1/2} = O_{\mathbb{P}}(p\,n^{-1/2})$$

alors,

$$|\gamma_r(\mathbf{K}_n^0)^b| \leq [\mathrm{tr}(\mathbf{K}_n^0)^2]^{b/2} = O_{\mathbb{P}}(p^b n^{-b/2}).$$

D'autre part, par la Proposition 4.5.2, il est clair que

$$\gamma_1(\mathbf{B}_n^0)^{-1} = \gamma_1((\mathbf{S}_n^0)^{-1}\mathbf{V}_n^0) \leq \frac{\gamma_1(\mathbf{V}_n^0)}{C_0},$$

avec une probabilité égale à 1. Ce qui implique que

$$\mathrm{M}\mathrm{K}((\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}) \leq p\;O_{\mathbb{P}}(d^b n^{-b/2}).$$

Alors, on obtient

$$\mathrm{M}\mathrm{K}((\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}) = O_{\mathbb{P}}(p^{b+1}n^{-b/2}). \tag{4.43}$$

En combinant les relations (4.36), (4.42) et (4.43), pour le deuxième terme du membre droit de la relation (4.40), on obtient que

$$|n\boldsymbol{\eta}_n^{0t}(\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}\boldsymbol{\eta}_n^0| \leq n\,p||\boldsymbol{\eta}_n^0||_2^2 \mathbb{M}\mathbb{X}((\mathbf{K}_n^0)^b(\mathbf{B}_n^0)^{-1}) = O_{\mathbb{P}}(p^{b+3}n^{-b/2}). \tag{4.44}$$

Les relations (4.40), (4.41), (4.44) et la supposition (A6), impliquent que

$$\begin{aligned} n\boldsymbol{\eta}_n^{0t}[\mathbf{I}_p - \mathbf{V}_n^0]\boldsymbol{\eta}_n^0 &= O_{\mathbb{P}}(p^{b+1}n^{-b/2}) + O_{\mathbb{P}}(p^{b+3}n^{-b/2}) \\ &= O_{\mathbb{P}}((p^{2+4/b}n^{-1})^b) \\ &= o_{\mathbb{P}}(1). \end{aligned} \tag{4.45}$$

En combinant les résultats donnés par les relations (4.39) et (4.45), on obtient la relation (4.33).

La Proposition est obtenue en combinant la relation (4.33) et la Proposition 4.3.3. ■

Le théorème suivant établit la normalité asymptotique de la statistique de test de vraisemblance empirique, quand le nombre des variables explicatives p converge vers l'infini quand $n \to \infty$. La preuve de ce théorème est très technique, de plus la présence du point de rupture dans le modèle intervient d'une manière essentielle. La Proposition 4.3.3 et la Proposition 4.3.4 doivent être utilisées dans la preuve de ce théorème.

Notons que la variance de standardisation Δ_n^2 dépend de la localisation du changement dans l'intervalle $[1:n]$.

Théorème 4.3.1. *Sous l'hypothèse nulle* H_0*, si les suppositions (A1)-(A9) sont vérifiées et* $p = o(n^{1/3})$*, alors*

(i)

$$\frac{n\boldsymbol{\psi}_n^{0t}(\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0 - p}{\Delta_n/n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1), \tag{4.46}$$

(ii)

$$\frac{\mathrm{EL}_{nk}(\boldsymbol{\beta}^{\mathbf{0}}) - p}{\Delta_n/n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1),$$

où $\Delta_n^2 = \sum_{i=1}^{n} \sigma_i^2$*, avec*

$$\sigma_1^2 = \theta^{-4}\left\{(\mathbf{X}_1^t(\mathbf{V}_n^0)^{-1}\mathbf{X}_1)^2\mathbb{E}[\varepsilon_1^4] - [tr((\mathbf{V}_n^0)^{-1}\mathbf{V}_{(1)}^0)]^2\right\}$$

et :

- *pour* $i = 2, \ldots, k+1$,

$$\sigma_i^2 = \frac{4}{\theta^4} \sum_{l=1}^{i-1} tr\left((\mathbf{V}_n^0)^{-1}\mathbf{V}_{(i)}^0(\mathbf{V}_n^0)^{-1}\mathbf{V}_{(l)}^0\right) + \frac{(\mathbf{X}_i^t(\mathbf{V}_n^0)^{-1}\mathbf{X}_i)^2\mathbb{E}[\varepsilon_1^4] - [tr((\mathbf{V}_n^0)^{-1}\mathbf{V}_{(i)}^0)]^2}{\theta^4},$$

- *pour* $i = k+2, \ldots, n$,

$$\begin{aligned}\sigma_i^2 \;=\; & \frac{4}{\theta^2(1-\theta)^2} \sum_{l=1}^{k} tr\left((\mathbf{V}_n^0)^{-1}\mathbf{V}_{(i)}^0(\mathbf{V}_n^0)^{-1}\mathbf{V}_{(l)}^0\right) + \frac{4}{(1-\theta)^4} \sum_{l=k+1}^{i-1} tr\left((\mathbf{V}_n^0)^{-1}\mathbf{V}_{(i)}^0(\mathbf{V}_n^0)^{-1}\mathbf{V}_{(l)}^0\right) \\ & + \frac{(\mathbf{X}_i^t(\mathbf{V}_n^0)^{-1}\mathbf{X}_i)^2\mathbb{E}[\varepsilon_1^4] - [tr((\mathbf{V}_n^0)^{-1}\mathbf{V}_{(i)}^0)]^2}{(1-\theta)^4}.\end{aligned}$$

Preuve. Puisque $\theta = k/n \to \theta^0 \in (0,1)$ et le point k où on réalise le test est connu, alors on suppose que θ est θ^0, et donc il est fixé.

(i) Afin de prouver la relation (4.46), on construit une martingale puis on utilise le Théorème Central Limite pour Martingales (voir [Chow and Teicher (1997)]). On prouve cette relation en quatre étapes. Dans la première étape, on construit la martingale, dans les étapes 2 et 3, on propose deux conditions suffisantes pour utiliser le Théorème Central Limite et finalement on prouve la relation (4.46) dans l'étape 4.

Étape 1. Dans cette étape, on construit une martingale.

Pour $i = 1, \ldots, n$, on définit la suite de vecteurs aléatoires suivants :

$$\mathbf{g}_i^0 \equiv \begin{cases} \dfrac{1}{\theta}\displaystyle\sum_{j=1}^{i} \mathbf{w}_j^0, & \text{si} \quad i \leq k, \\ \dfrac{1}{\theta}\displaystyle\sum_{j=1}^{k} \mathbf{w}_j^0 - \dfrac{1}{1-\theta}\displaystyle\sum_{j=k+1}^{i} \mathbf{w}_j^0, & \text{si} \quad i > k, \end{cases}$$

et également la variable aléatoire

$$H_i^0 \equiv ||\mathbf{g}_i^0||_2^2 - ip.$$

Alors, le membre droit de la relation (4.46) peut s'écrire

$$\frac{n\boldsymbol{\psi}_n^{0t}(\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0 - p}{\Delta_n/n} = \frac{H_n^0}{\Delta_n}.$$

La relation entre $\mathbf{g}_i^0$ et $\mathbf{g}_{i-1}^0$ est

$$\mathbf{g}_i^0 = \begin{cases} \mathbf{g}_{i-1}^0 + \dfrac{\mathbf{w}_i^0}{\theta}, & \text{si} \quad i \leq k, \\ \\ \mathbf{g}_{i-1}^0 - \dfrac{\mathbf{w}_i^0}{1-\theta}, & \text{si} \quad i > k, \end{cases}$$

avec $\mathbf{g}_0^0 \equiv \mathbf{0}$.

On considère maintenant la filtration suivante

$$\mathcal{F}_i = \sigma(\mathbf{w}_1^0, \ldots, \mathbf{w}_i^0) = \sigma(\mathbf{g}_1^0, \ldots, \mathbf{g}_i^0)$$

pour $i = 1 \ldots, n$ est le σ-algèbre généré par $\mathbf{w}_1^0, \ldots, \mathbf{w}_i^0$ ou par $\mathbf{g}_1^0, \ldots, \mathbf{g}_i^0$. Tout d'abord, on étudie si $\{H_i^0, \mathcal{F}_i\}_{i\geq 1}$ est une martingale. Pour ceci, on considère par exemple i tel que $i > k$. Alors

$$\mathbb{E}[H_i^0|\mathcal{F}_{i-1}] = \|\mathbf{g}_{i-1}^0\|^2 + \frac{\mathbb{E}[\|\mathbf{w}_i^0\|^2]}{(1-\theta)^2} - id \neq H_{i-1}^0.$$

En conséquence, $\{H_i^0, \mathcal{F}_i\}_{i\geq 1}$ n'est pas une martingale.

Maintenant, on construit une martingale basée sur $\mathbf{g}_n^0$ par rapport à la filtration $\{\mathcal{F}_i\}_{i\geq 1}$. Pour ceci, on définit la suite de variables aléatoires $U_i \equiv H_i^0 - H_{i-1}^0$, pour $i = 1, \ldots, n$, avec $\mathrm{H}_0^0 = 0$. Alors

$$U_i^0 = \begin{cases} \dfrac{2}{\theta}\mathbf{g}_{i-1}^{0t}\mathbf{w}_i^0 + \dfrac{\|\mathbf{w}_i^0\|^2}{\theta^2} - d, & \text{si} \quad i \leq k, \\ \\ -\dfrac{2}{1-\theta}\mathbf{g}_{i-1}^{0t}\mathbf{w}_i^0 + \dfrac{\|\mathbf{w}_i^0\|^2}{(1-\theta)^2} - d, & \text{si} \quad i > k. \end{cases}$$

On considère les deux suites de variables aléatoires suivantes :

$$\tau_i^0 \equiv U_i^0 - \mathbb{E}[U_i^0]$$

et

$$\varphi_i^0 \equiv \sum_{j=1}^{i} \tau_j^0 = H_i^0 - \sum_{j=1}^{i} \mathbb{E}[U_j^0].$$

Pour tout $i \leq k$, on a

$$\begin{aligned}\mathbb{E}[\varphi_i^0|\mathcal{F}_{i-1}] &= \|\mathbf{g}_{i-1}^0\|^2 + \frac{\mathbb{E}[\|\mathbf{w}_i^0\|^2]}{\theta^2} - id - \sum_{j=1}^{i}\mathbb{E}[U_j^0] \\ &= \|\mathbf{g}_{i-1}^0\|^2 - (i-1)d - \sum_{j=1}^{i-1}\mathbb{E}[U_j^0] \\ &= \varphi_{i-1}^0.\end{aligned}$$

Pour tout $i > k$:

$$\begin{aligned}\mathbb{E}[\varphi_i^0|\mathcal{F}_{i-1}] &= \|\mathbf{g}_{i-1}^0\|^2 + \frac{\mathbb{E}[\|\mathbf{w}_i^0\|^2]}{(1-\theta)^2} - id - \sum_{j=1}^{i}\mathbb{E}[U_j^0] \\ &= \varphi_{i-1}^0.\end{aligned}$$

Alors, $\{\varphi_i^0\}_{i\geq 1}$ est une martingale par rapport à $\{\mathcal{F}_i\}_{i\geq 1}$.

Pour appliquer le Théorème Central Limite pour Martingales de [Chow and Teicher (1997)] pour $\{\varphi_i^0, \mathcal{F}_i, i \geq 1\}$, il suffit de montrer que

$$\sum_{i=1}^{n}\mathbb{E}\big[|\tau_i^0|^3\big] = o(\Delta_n^3) \tag{4.47}$$

et

$$\sum_{i=1}^{n}\mathbb{E}\Big[|\mathbb{E}\big[(\tau_i^0)^2|\mathcal{F}_{i-1}\big] - \sigma_i^2|\Big] = o(\Delta_n^2). \tag{4.48}$$

Étape 2. Dans cette étape, on prouve la relation (4.47).

Afin de faciliter l'écriture, pour $i = 1, \ldots, n$, on note

$$\mathbf{N}_i^0 \equiv \frac{\mathbf{w}_i^0}{\theta}\mathbb{1}_{i\leq k} - \frac{\mathbf{w}_i^0}{1-\theta}\mathbb{1}_{i>k}. \tag{4.49}$$

Alors, la variable aléatoire τ_i^0 peut s'écrire

$$\tau_i^0 = 2\mathbf{g}_{i-1}^{0t}\mathbf{N}_i^0 + ||\mathbf{N}_i^0||_2^2 - \mathbb{E}[||\mathbf{N}_i^0||_2^2]. \tag{4.50}$$

Par la supposition (A9), pour une constante positive $C_7 < \infty$, pour tout $i = 1, \ldots, n$ et tout $j_1, \ldots, j_l = 1, \ldots, p$, $l \in \mathbb{N}$, pour $\sum_{i=1}^{l} d_i \leq 6$, on a

$$\mathbb{E}[N_{i,j_1}^{0d_1} \ldots N_{i,j_l}^{0d_l}] \leq C_7, \tag{4.51}$$

avec N_{i,j_1}^0 la j_1-ème composante du vecteur $\mathbf{N}_i^0$ définie dans la relation (4.49).

Par l'inégalité de Hölder, chaque $b \leq 3$, on a

$$\mathbb{E}[||\mathbf{N}_i^0||^{2b}] \leq C\mathbb{E}[\big(\sum_{j=1}^{p} w_{ij}^2\big)^b] \leq Cp^{b-1}\sum_{j=1}^{p}\mathbb{E}[w_{ij}^{2b}] \leq Cp^b. \tag{4.52}$$

Par l'inégalité de Cauchy Schwartz, on a que

$$\mathbb{E}\big[\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0\big]^3 \leq \big[\mathbb{E}[\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0]^6\big]^{1/2}. \tag{4.53}$$

D'autre part, par le Lemme 7 de [Guo et al. (2013)], on a

$$\mathbb{E}[\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0]^6 \leq C\,p^6(i^3 + i^2 + i). \tag{4.54}$$

Alors, par les relations (4.53) et (4.54), on obtient que

$$\mathbb{E}\big[\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0\big]^3 \leq [C\,p^6(i^3 + i^2 + i)]^{1/2},$$

ce qui implique que

$$\mathbb{E}\big[\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0\big]^3 \leq C\,p^3(i^{3/2} + i + i^{1/2}). \tag{4.55}$$

D'autre part, on a

$$\mathbb{E}\big[||\mathbf{N}_i^0||_2^2 - \mathbb{E}[||\mathbf{N}_i^0||_2^2]\big]^3 = \mathbb{E}\left[||\mathbf{N}_i^0||_2^6 - 3||\mathbf{N}_i^0||_2^4\mathbb{E}[||\mathbf{N}_i^0||_2^2] + 3||\mathbf{N}_i^0||_2^2(\mathbb{E}[||\mathbf{N}_i^0||_2^2])^2 - (\mathbb{E}[||\mathbf{N}_i^0||_2^2])^3\right].$$

En utilisant la relation (4.52), on obtient que

$$\Big|\mathbb{E}\big[||\mathbf{N}_i^0||_2^2 - \mathbb{E}[||\mathbf{N}_i^0||_2^2]\big]^3\Big| \leq Cp^3. \tag{4.56}$$

En utilisant les relations (4.55) et (4.56), on peut écrire

$$\mathbb{E}[|\tau_i^0|^3] \leq Cp^3(1 + i^{3/2} + i + i^{1/2}).$$

Alors

$$\sum_{i=1}^{n} \mathbb{E}[|\tau_i^0|^3] \leq Cnp^3(1 + n^{3/2} + n + n^{1/2}),$$

ce qui donne

$$\sum_{i=1}^{n} \mathbb{E}[|\tau_i^0|^3] \leq Cp^3(n^{5/2} + n^2 + n^{3/2} + n). \tag{4.57}$$

D'autre part, en utilisant la supposition (A1) et par des arguments similaires à ceux de la relation (16) de [Guo et al. (2013)], on obtient que

$$\Delta_n^2 \geq Cn^2p + O_{\mathbb{P}}(p^2n).$$

Alors

$$\Delta_n^3 \geq Cn^3p^{3/2} + O_{\mathbb{P}}(p^3n^{3/2}). \tag{4.58}$$

Par les relations (4.57) et (4.58), on obtient

$$\begin{aligned} \frac{\sum_{i=1}^n \mathbb{E}[|\tau_i^0|^3]}{\Delta_n^3} &\leq \frac{Cp^3(n^{5/2} + n^2 + n^{3/2} + n)}{Cn^3p^{3/2} + O_{\mathbb{P}}(p^3n^{3/2})} \\ &\leq C\,\frac{p^3n^{5/2}}{n^3p^{3/2}} \\ &= O_{\mathbb{P}}\Big((p^3/n)^{1/2}\Big). \end{aligned}$$

Puisque $p = o(n^{1/3})$, la relation (4.47) résulte.

Étape 3. Dans cette étape on prouve la relation (4.48).

Par des calculs élémentaires, et en utilisant la relation (4.52), on obtient

$$(\tau_i^0)^2 = 4\mathbf{g}_{i-1}^{0t}\mathbf{N}_i^0\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0 + 4\mathbf{g}_{i-1}^{0t}\mathbf{N}_i^0(||\mathbf{N}_i^0||_2^2 - \mathbb{E}[||\mathbf{N}_i^0||^2]) + O_{\mathbb{P}}(p^2).$$

On observe également que

$$\sigma_i^2 = \mathbb{E}\big[(\tau_i^0)^2\big].$$

Alors

$$\mathbb{E}\big[(\tau_i^0)^2|\mathcal{F}_{i-1}\big] = 4\mathbf{g}_{i-1}^{0t}\mathbb{E}\big[\mathbf{N}_i^0\mathbf{N}_i^{0t}\big]\mathbf{g}_{i-1}^0 + 4\mathbf{g}_{i-1}^{0t}\mathbb{E}\big[\mathbf{N}_i^0\big(||\mathbf{N}_i^0||^2 - \mathbb{E}[||\mathbf{N}_i^0||^2]\big)\big] + O_{\mathbb{P}}(p^2) \tag{4.59}$$

et

$$\sigma_i^2 = 4\mathbb{E}\big[\mathbf{g}_{i-1}^{0t}\mathbb{E}[\mathbf{N}_i^0\mathbf{N}_i^{0t}]\mathbf{g}_{i-1}^0\big] + \mathbb{E}\left[\left(||\mathbf{N}_i^0||^2 - \mathbb{E}[||\mathbf{N}_i^0||^2]\right)^2\right] + O(p^2).$$

Par l'inégalité de (4.52) on a

$$\mathbb{E}\left[\left(||\mathbf{N}_i^0||^2 - \mathbb{E}[||\mathbf{N}_i^0||^2]\right)^2\right] = O(p^2).$$

Alors

$$\sigma_i^2 = 4\mathbb{E}\big[\mathbf{g}_{i-1}^{0t}\mathbb{E}[\mathbf{N}_i^0\mathbf{N}_i^{0t}]\mathbf{g}_{i-1}^0\big] + O(p^2). \tag{4.60}$$

En utilisant les relations (4.59) et (4.60), pour chaque $i = 1, \ldots, n$, on obtient que

$$\begin{aligned}
\mathbb{E}\big[\big|\mathbb{E}[\tau_i^2|\mathcal{F}_{i-1}] - \sigma_i^2\big|^2\big] &\leq 16\Big(\mathbb{E}\Big[\mathbf{g}_{i-1}^{0t}\mathbb{E}[\mathbf{N}_i^0\mathbf{N}_i^{0t}]\mathbf{g}_{i-1}^0 - \mathbb{E}[\mathbf{g}_{i-1}^{0t}\mathbf{N}_i^0\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0]\Big]^2 \\
&\quad + \mathbb{E}\Big[\mathbf{N}_i^{0t}\big(||\mathbf{N}_i^0||^2 - \mathbb{E}[||\mathbf{N}_i^0||^2]\big)\Big]\mathbb{E}\Big[\mathbf{g}_{i-1}^0\mathbf{g}_{i-1}^{0t}\Big]\mathbb{E}\Big[\mathbf{N}_i^0\big(||\mathbf{N}_i^0||^2 - \mathbb{E}[||\mathbf{N}_i^0||^2]\big)\Big] \\
&\quad + O(p^4)\Big) \\
&\equiv 16\Big(A + B + O(p^4)\Big) \qquad (4.61)
\end{aligned}$$

Pour le terme A de la relation (4.61), on a la décomposition suivante

$$A = \mathbb{E}\Big[\mathbf{g}_{i-1}^{0t}\mathbb{E}[\mathbf{N}_i^0\mathbf{N}_i^{0t}]\mathbf{g}_{i-1}^0\Big]^2 - \Big[\mathbb{E}[\mathbf{g}_{i-1}^{0t}\mathbf{N}_i^0\mathbf{N}_i^{0t}\mathbf{g}_{i-1}^0]\Big]^2 \equiv A_1 - A_2. \tag{4.62}$$

Avant d'analyser les termes A_1 et A_2, on observe que

$$\mathbb{E}(\mathbf{N}_i^0\mathbf{N}_i^{0t}) = (\mathbf{V}_n^0)^{-1/2}\left[\frac{1}{\theta^2}\mathbb{1}_{i\leq k} + \frac{1}{(1-\theta)^2}\mathbb{1}_{i>k}\right]\mathbf{V}_{(i)}^0(\mathbf{V}_n^0)^{-1/2}. \tag{4.63}$$

Afin de faciliter l'écriture, on considère la matrice suivante

$$\mathbf{M}_i^0 \equiv \left(\frac{1}{\theta^2}\mathbb{1}_{i\leq k} + \frac{1}{(1-\theta)^2}\mathbb{1}_{i>k}\right)\mathbf{V}_{(i)}^0.$$

Alors, $\mathbb{E}[\mathbf{N}_i^0\mathbf{N}_i^{0t}]$ peut s'écrire sous la forme suivante

$$\mathbb{E}[\mathbf{N}_i^0\mathbf{N}_i^{0t}] = (\mathbf{V}_n^0)^{-1/2}\,\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}. \tag{4.64}$$

Ainsi, le terme A_1 de la relation (4.62), peut s'écrire

$$A_1 = \mathbb{E}\Big[\mathbf{g}_{i-1}^{0t}(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\ (\mathbf{V}_n^0)^{-1/2}[\mathbf{g}_{i-1}^0\Big]^2.$$

Tenant compte du fait que les vecteurs aléatoires $\mathbf{w}_i^0$ sont indépendants, avec une espérance zero, pour tout $i = 1, \ldots, n$, le terme A_1 peut se décomposer sous la forme suivante

$$A_1 \equiv A_{11} + A_{12} + A_{13}, \tag{4.65}$$

avec

- *si* $i - 1 \leq k$:

$$A_{11} \equiv \frac{1}{\theta^4} \sum_{h \neq l = 1}^{i-1} \mathbb{E}\big[\mathbf{w}_h^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_h^0\big] \mathbb{E}\big[\mathbf{w}_l^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_l^0\big],$$

$$A_{12} \equiv \frac{2}{\theta^4} \sum_{h \neq l = 1}^{i-1} \mathbb{E}\Big[\mathbf{w}_h^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_l^0 \mathbf{w}_h^{0t} (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_l^0\Big]$$

et

$$A_{13} \equiv \frac{1}{\theta^4} \sum_{h = 1}^{i-1} \mathbb{E}\Big[\mathbf{w}_h^{0t} \ (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \ \mathbf{w}_h^0\Big]^2.$$

- *si* $i - 1 > k$:

$$\begin{aligned} A_{11} \ \equiv \ & \frac{1}{\theta^4} \sum_{h \neq l = 1}^{k} \mathbb{E}\big[\mathbf{w}_h^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_h^0\big] \mathbb{E}\big[\mathbf{w}_l^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_l^0\big] \\ & + \frac{1}{(1-\theta)^4} \sum_{h' \neq l' = k+1}^{i-1} \mathbb{E}\big[\mathbf{w}_{h'}^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_{h'}^0\big] \mathbb{E}\big[\mathbf{w}_{l'}^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_{l'}^0\big], \end{aligned}$$

$$\begin{aligned} A_{12} \ \equiv \ & \frac{2}{\theta^4} \sum_{h \neq l = 1}^{k} \mathbb{E}\Big[\mathbf{w}_h^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_l^0 \mathbf{w}_h^{0t} (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_l^0\Big] \\ + \ & \frac{2}{(1-\theta)^4} \sum_{h' \neq l' = k+1}^{i-1} \mathbb{E}\Big[\mathbf{w}_{h'}^0 (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_{l'}^0 \mathbf{w}_{h'}^{0t} (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \mathbf{w}_{l'}^0\Big]. \end{aligned}$$

et

$$A_{13} \ \equiv \ \frac{1}{\theta^4} \sum_{h = 1}^{k} \mathbb{E}\Big[\mathbf{w}_h^{0t} \ (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \ \mathbf{w}_h^0\Big]^2 + \frac{1}{(1-\theta)^4} \sum_{h' = k+1}^{i-1} \mathbb{E}\Big[\mathbf{w}_{h'}^{0t} \ (\mathbf{V}_n^0)^{-1/2} \mathbf{M}_i^0 \ (\mathbf{V}_n^0)^{-1/2} \ \mathbf{w}_{h'}^0\Big]^2.$$

Nous voulons étudier les termes A_{11}, A_{12}, A_{13}. Pour ceci, on considère le cas $i-1>k$, l'autre cas étant similaire.

Pour le terme A_{11}, en appliquant l'inégalité de Cauchy-Schwarz, on obtient

$$\begin{aligned} A_{11} &= \frac{1}{\theta^4}\Big(\sum_{h=1}^{k}\mathbb{E}\Big[\mathbf{w}_h^{0t}(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\ (\mathbf{V}_n^0)^{-1/2}\mathbf{w}_h^0\Big]\Big)^2+\frac{1}{(1-\theta)^4}\Big(\sum_{h'=k+1}^{i-1}\mathbb{E}\Big[\mathbf{w}_{h'}^{0t}(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\ (\mathbf{V}_n^0)^{-1/2}\mathbf{w}_{h'}^0\Big]\Big)^2 \\ &= \frac{1}{\theta^4}\Big(\sum_{h=1}^{k}\mathbb{E}\Big[\mathbf{z}_h^{0t}(\mathbf{V}_n^0)^{-1}\mathbf{M}_i^0\ (\mathbf{V}_n^0)^{-1}\mathbf{z}_h^0\Big]\Big)^2+\frac{1}{(1-\theta)^4}\Big(\sum_{h'=k+1}^{i-1}\mathbb{E}\Big[\mathbf{z}_{h'}^{0t}(\mathbf{V}_n^0)^{-1}\mathbf{M}_i^0\ (\mathbf{V}_n^0)^{-1}\mathbf{z}_{h'}^0\Big]\Big)^2. \end{aligned}$$

Puisque les $(\varepsilon_i)_{1\le i\le n}$ sont indépendantes, alors

$$A_{11}=\frac{1}{\theta^4}\Big(\sum_{h=1}^{k}\mathrm{tr}(\mathbf{V}_{(h)}^0(\mathbf{V}_n^0)^{-1}\mathbf{M}_i^0(\mathbf{V}_n^0)^{-1})\Big)^2+\frac{1}{(1-\theta)^4}\Big(\sum_{h'=k+1}^{i-1}\mathrm{tr}\big(\mathbf{V}_{(h')}^0(\mathbf{V}_n^0)^{-1}\mathbf{M}_i^0\ (\mathbf{V}_n^0)^{-1}\big)\Big)^2. \tag{4.66}$$

De même, pour le terme A_{12}, on a :

$$\begin{aligned} A_{12} &= \frac{2}{\theta^4}\mathbb{E}\Big[\sum_{h\neq l=1}^{k}\mathbf{w}_h^0(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbf{w}_l^0\mathbf{w}_l^{0t}(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbf{w}_h^0\Big] \\ &\quad+\frac{2}{(1-\theta)^4}\mathbb{E}\Big[\sum_{h'\neq l'=k+1}^{i-1}\mathbf{w}_{h'}^0(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbf{w}_{l'}^0\mathbf{w}_{l'}^{0t}(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbf{w}_{h'}^0\Big] \\ &= \frac{2}{\theta^4}\mathbb{E}\Big[\sum_{h\neq l=1}^{k}\mathbf{w}_h^0(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbb{E}\big[\mathbf{w}_l^0\mathbf{w}_l^{0t}\big](\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbf{w}_h^0\Big] \\ &\quad+\frac{2}{(1-\theta)^4}\mathbb{E}\Big[\sum_{h'\neq l'=k+1}^{i-1}\mathbf{w}_{h'}^0(\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbb{E}\big[\mathbf{w}_{l'}^0\mathbf{w}_{l'}^{0t}\big](\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\mathbf{w}_{h'}^0\Big] \\ &\le \frac{2}{\theta^4}pk^2\,\gamma_1^4(\mathbf{V}_n^0)^{-1}\sup_{h,l\in\{1,\dots,k\}}\Big(\gamma_1(\mathbf{V}_{(l)}^0)\,\gamma_1(\mathbf{V}_{(h)}^0)\Big)\,\gamma_1^2(\mathbf{M}_i^0) \\ &\quad+\frac{2}{(1-\theta)^4}p\,\big(i-1-k\big)^2\,\gamma_1^4(\mathbf{V}_n^0)^{-1}\sup_{h',l'\in\{k+1,\dots,i-1\}}\Big(\gamma_1(\mathbf{V}_{(l')}^0)\,\gamma_1(\mathbf{V}_{(h')}^0)\Big)\,\gamma_1^2(\mathbf{M}_i^0). \end{aligned}$$

Tenant compte de la supposition (A1), on obtient que

$$A_{12}\le Cp(i-1)^2. \tag{4.67}$$

Pour le terme A_{13} de la relation (4.65), on a

$$\begin{aligned} A_{13} &= \frac{1}{\theta^4}\sum_{h=1}^{k}\sum_{j,l,s,t=1}^{p}\Big((\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\Big)_{jl}\Big((\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\Big)_{st}\mathbb{E}(w_{h,j}^0 w_{h,l}^0 w_{h,s}^0 w_{h,t}^0) \\ &+ \frac{1}{(1-\theta)^4}\sum_{h'=k+1}^{i-1}\sum_{j,l,s,t=1}^{p}\Big((\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\Big)_{jl}\Big((\mathbf{V}_n^0)^{-1/2}\mathbf{M}_i^0\,(\mathbf{V}_n^0)^{-1/2}\Big)_{st}\mathbb{E}(w_{h',j}^0 w_{h',l}^0 w_{h',s}^0 w_{h',t}^0). \end{aligned}$$

Tenant compte de la supposition (A9), on obtient que

$$A_{13} \leq \frac{1}{\theta^4}C\ p^4 k\gamma_1^4(\mathbf{V}_n^0)^{-1}\gamma_1^2(\mathbf{M}_i^0) + \frac{1}{(1-\theta)^4}C\ p^4(i-k)\gamma_1^4(\mathbf{V}_n^0)^{-1}\gamma_1^2(\mathbf{M}_i^0).$$

En utilisant également la supposition (A1), on obtient que

$$A_{13} \leq Cp^4(i-1). \tag{4.68}$$

Pour le terme A_2 de la relation (4.62), par des calculs similaires, on obtient que

- *si* $i-1 \leq k$:

$$A_2 = \frac{1}{\theta^4}\Big[\sum_{h=1}^{i-1}\mathrm{tr}(\mathbf{V}_{(h)}^0(\mathbf{V}_n^0)^{-1}\mathbf{M}_i^0(\mathbf{V}_n^0)^{-1})\Big]^2$$

- *si* $i-1 > k$:

$$A_2 = \frac{1}{\theta^4}\Big[\sum_{h=1}^{k}\mathrm{tr}(\mathbf{V}_{(h)}^0(\mathbf{V}_n^0)^{-1}\mathbf{M}_i^0(\mathbf{V}_n^0)^{-1})\Big]^2 + \frac{1}{(1-\theta)^4}\Big[\sum_{h'=k+1}^{i-1}\mathrm{tr}\big(\mathbf{V}_{(h')}^0(\mathbf{V}_n^0)^{-1}\mathbf{M}_i^0\ (\mathbf{V}_n^0)^{-1}\big)\Big]^2. \tag{4.69}$$

Pour $i-1 > k$, par les relations (4.66) et (4.69), on obtient que

$$A_{11} - A_2 \leq 0. \tag{4.70}$$

L'inégalité (4.70) est également vraie pour $i-1 \leq k$.

Puisque le terme A de la relation (4.62) peut s'écrire

$$A = (A_{11} + A_{12} + A_{13}) - A_2,$$

en combinant les relations (4.67), (4.68) et (4.70), on obtient que

$$A \leq Cp(i-1)^2 + Cp^4(i-1). \tag{4.71}$$

Pour le terme B de la relation (4.61), dans le cas $i-1>k$, tenant compte du fait que pour tout $h=1,\ldots,n$, $\mathbb{E}(\mathbf{N}_h^0)=\mathbf{0}_p$ et $\mathbb{E}(\mathbf{w}_h^0)=\mathbf{0}_p$, ainsi que pour $h\neq h'$, $\mathbf{w}_h^0$ est indépendant de $\mathbf{w}_{h'}^0$, alors

$$\begin{aligned}
B &= \mathbb{E}\big[\mathbf{N}_i^{0t}||\mathbf{N}_i^0||^2\big]\mathbb{E}\big[\big(\frac{1}{\theta}\sum_{h=1}^{k}\mathbf{w}_h^0-\frac{1}{1-\theta}\sum_{h'=k+1}^{i-1}\mathbf{w}_{h'}^0\big)\cdot\big(\frac{1}{\theta}\sum_{h=1}^{k}\mathbf{w}_h-\frac{1}{1-\theta}\sum_{h'=k+1}^{i-1}\mathbf{w}_{h'}\big)^t\big]\mathbb{E}\big[\mathbf{N}_i^0||\mathbf{N}_i^0||^2\big]\\
&= \frac{1}{\theta^2}\mathbb{E}\big[\mathbf{N}_i^{0t}||\mathbf{N}_i^0||^2\big]\sum_{h,l=1}^{k}\mathbb{E}\big[\mathbf{w}_h^0\mathbf{w}_l^{0t}\big]\,\mathbb{E}\big[\mathbf{N}_i^{0t}||\mathbf{N}_i^0||^2\big]\\
&\quad+\frac{1}{(1-\theta)^2}\mathbb{E}\big[\mathbf{N}_i^{0t}||\mathbf{N}_i^0||^2\big]\sum_{h',l'=k+1}^{i-1}\mathbb{E}\big[\mathbf{w}_{h'}^0\mathbf{w}_{l'}^{0t}\big]\,\mathbb{E}\big[\mathbf{N}_i^0||\mathbf{N}_i^0||^2\big]\\
&= \frac{1}{\theta^2}\sum_{u,s=1}^{p}\sum_{l,r=1}^{p}\Big(\sum_{h=1}^{k}\mathbf{V}_n^{-1/2}\mathbf{V}_{(h)}^0\mathbf{V}_n^{-1/2}\Big)_{us}\mathbb{E}\big[N_{i,u}^0(N_{i,l}^0)^2\big]\mathbb{E}\big[N_{i,s}^0(N_{i,r}^0)^2\big]\\
&\quad+\frac{1}{(1-\theta)^2}\sum_{u,s=1}^{p}\sum_{l,r=1}^{p}\Big(\sum_{h'=k+1}^{i-1}(\mathbf{V}_n^0)^{-1/2}\mathbf{V}_{(h')}^0(\mathbf{V}_n^0)^{-\frac{1}{2}}\Big)_{us}\mathbb{E}\big[N_{i,u}^0(N_{i,l}^0)^2\big]\mathbb{E}\big[N_{i,s}^0(N_{i,r}^0)^2\big],
\end{aligned}$$

où, $\Big(\sum_{h=1}^{k}\mathbf{V}_n^{-1/2}\mathbf{V}_{(h)}^0\mathbf{V}_n^{-1/2}\Big)_{us}$ est l'élément (u,s) de la matrice $\sum_{h=1}^{k}\mathbf{V}_n^{-1/2}\mathbf{V}_{(h)}^0\mathbf{V}_n^{-1/2}$.

En utilisant le Lemme 4 du [Liu et al(2013)] et la relation (4.51), on obtient

$$\begin{aligned}
B &\leq \frac{1}{\theta^2}\sum_{u,s=1}^{p}\sum_{l,r=1}^{p}\mathrm{Mx}\Big(\sum_{h=1}^{k}\mathbf{V}_n^{-1/2}\mathbf{V}_{(h)}^0\mathbf{V}_n^{-1/2}\Big)_{us}\mathbb{E}\big[N_{i,u}^0(N_{i,l}^0)^2\big]\mathbb{E}\big[N_{i,s}^0(N_{i,r}^0)^2\big]\\
&\quad+\frac{1}{(1-\theta)^2}\sum_{u,s=1}^{p}\sum_{l,r=1}^{p}\mathrm{Mx}\Big(\sum_{h'=k+1}^{i-1}(\mathbf{V}_n^0)^{-1/2}\mathbf{V}_{(h')}^0(\mathbf{V}_n^0)^{-1/2}\Big)_{us}\mathbb{E}\big[N_{i,u}^0(N_{i,l}^0)^2\big]\mathbb{E}\big[N_{i,s}^0(N_{i,r}^0)^2\big]\\
&\leq \frac{C}{\theta^2}p^4(k\gamma_1^2(\mathbf{V}_n^{-1/2})\sup_{h\in\{1,\ldots,k\}}\Big(\gamma_1(\mathbf{V}_{(h)}^0)\Big)+\frac{C}{(1-\theta)^2}p^4(i-1-k)\gamma_1^2(\mathbf{V}_n^{-1/2})\sup_{h\in\{k+1,\ldots,i-1\}}\Big(\gamma_1(\mathbf{V}_{(h')}^0)\Big).
\end{aligned}$$

Alors, par la supposition (A1), on obtient que

$$B\leq Cp^4(i-1). \tag{4.72}$$

De même, on peut prouver que l'inégalité (4.72) est également vraie pour $i-1\leq k$.

En combinant les relations (4.61), (4.71) et (4.72), on obtient que

$$\begin{aligned}\mathbb{E}\big[|\mathbb{E}[(\tau_i^0)^2|\mathcal{F}_{i-1}] - \sigma_i^2|^2\big] &\leq 16\Big(A + B + O(p^4)\Big)\\ &\leq Cp(i-1)^2 + Cp^4(i-1) + Cp^4(i-1) + O(p^4),\end{aligned}$$

ce qui implique que

$$\mathbb{E}\big[|\mathbb{E}[(\tau_i^0)^2|\mathcal{F}_{i-1}] - \sigma_i^2|^2\big] \leq Cp(i-1)^2 + Cp^4(i-1) + O(p^4).$$

Par l'inégalité de Cauchy-Schwarz, on a

$$\begin{aligned}\frac{\sum_{i=1}^n \mathbb{E}\big[|\mathbb{E}[(\tau_i^0)^2|\mathcal{F}_{i-1}] - \sigma_i^2|\big]}{\Delta_n^2} &\leq \frac{\sum_{i=1}^n \Big(\mathbb{E}|\mathbb{E}[(\tau_i^0)^2|\mathcal{F}_{i-1}] - \sigma_i^2|^2\Big)^{1/2}}{\Delta_n^2}\\ &\leq \frac{Cn(n^{1/2}p^2 + p^{1/2}n + O(p^2))}{C\,n^2p + O_{\mathbb{P}}(p^2n)}\\ &\leq C\frac{n^{3/2}p^2}{n^2p}.\end{aligned}$$

ce qui implique la relation (4.48).

Étape 4. Dans cette étape, on complète la preuve de la relation (4.46) sous la base du Théorème Central Limite pour martingales. En utilisant les relations (4.47) et (4.48) prouvées dans les étapes 2 et 3, on applique le Théorème Central Limite pour Martingales de [Chow and Teicher (1997)] (Théorème 1, page 336), pour $\{\varphi_n^0, \mathcal{F}_n\}_{n\geq 1}$. On obtient :

$$\frac{\varphi_n^0}{\Delta_n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1). \tag{4.73}$$

En utilisant les notations données dans l'*étape 1*, alors

$$n\left(n\boldsymbol{\psi}_n^{0t}(\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0 - p\right) = H_n^0 = \varphi_n^0 + \sum_{i=1}^n \mathbb{E}[U_i^0].$$

D'autre part, puisque $\mathrm{H}_0^0 = 0$, on a

$$\sum_{i=1}^n \mathbb{E}[U_i^0] = \mathbb{E}[H_n^0] = \mathbb{E}[\|\mathbf{g}_n^0\|^2] - np.$$

Tenant compte des relations (4.12) et (4.13), on obtient :

$$\begin{aligned}
\mathbb{E}[\|\mathbf{g}_n^0\|^2] &= \frac{1}{\theta^2}\sum_{i=1}^{k}\mathbb{E}[\|\mathbf{w}_i^0\|^2] + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathbb{E}[\|\mathbf{w}_j^0\|^2] \\
&= \frac{\sigma^2}{\theta^2}\sum_{i=1}^{k}\left(\mathbf{X}_i^t(\mathbf{V}_n^0)^{-1/2}\right)\left((\mathbf{V}_n^0)^{-1/2}\mathbf{X}_i\right) + \frac{\sigma^2}{(1-\theta)^2}\sum_{j=k+1}^{n}\left(\mathbf{X}_j^t(\mathbf{V}_n^0)^{-1/2}\right)\left((\mathbf{V}_n^0)^{-1/2}\mathbf{X}_j\right) \\
&= \frac{\sigma^2}{\theta^2}\sum_{i=1}^{k}\mathrm{tr}\left(\mathbf{X}_i^t(\mathbf{V}_n^0)^{-1/2}\right)\left((\mathbf{V}_n^0)^{-1/2}\mathbf{X}_i\right) + \frac{\sigma^2}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathrm{tr}\left(\mathbf{X}_j^t(\mathbf{V}_n^0)^{-1/2}\right)\left((\mathbf{V}_n^0)^{-1/2}\mathbf{X}_j\right) \\
&= \sigma^2\mathrm{tr}\left((\mathbf{V}_n^0)^{-1/2}\left(\frac{1}{\theta^2}\sum_{i=1}^{k}\mathbf{X}_i\mathbf{X}_i^t + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathbf{X}_j\mathbf{X}_j^t\right)(\mathbf{V}_n^0)^{-1/2}\right) \\
&= n\mathrm{tr}\left((\mathbf{V}_n^0)^{-1/2}\mathbf{V}_n^0(\mathbf{V}_n^0)^{-1/2}\right) \\
&= np.
\end{aligned}$$

Ainsi, $\mathbb{E}[H_n^0] = 0$. Tenant compte de la relation (4.73), on obtient la relation (4.46).

(ii) Cette affirmation est obtenue, en combinant les résultats de l'affirmation *(i)* avec la Proposition 4.3.3 et la Proposition 4.3.4. ■

Le résultat suivant représente un corollaire immédiat du Théorème 4.3.1.

Corollaire 4.3.1. *Tester l'hypothèse nulle* $\mathrm{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_2 = \boldsymbol{\beta^0}$ *contre l'hypothèse alternative* $\mathrm{H}_1 : \boldsymbol{\beta} = \boldsymbol{\beta^0}$, $\boldsymbol{\beta}_2 \neq \boldsymbol{\beta^0}$, *est équivalent à construire la région de confiance pour* $\boldsymbol{\gamma} = \boldsymbol{\beta^0} - \boldsymbol{\beta}_2^0$, *ou tester l'hypothèse nulle* $\mathrm{H}_0' : \boldsymbol{\gamma} = \mathbf{0}_p$.
Alors, en appliquant le Théorème 4.3.1, afin de tester H_0 *contre* H_1, *on considère la statistique de test suivante*

$$\mathcal{Z}(\boldsymbol{\beta^0}) \equiv \frac{n\boldsymbol{\psi}_n^t(\boldsymbol{\beta^0})\left(\mathbf{V}_n^0\right)^{-1}\boldsymbol{\psi}_n(\boldsymbol{\beta^0}) - p}{\Delta_n/n}. \tag{4.74}$$

Le comportement asymptotique de la statistique de test $\mathcal{Z}(\boldsymbol{\beta^0})$ sous l'hypothèse H_1 est donné par le théorème suivant. On montre que sous l'hypothèse alternative H_1, la statistique $\mathcal{Z}(\boldsymbol{\beta^0})$ diverge.

Théorème 4.3.2. *Sous l'hypothèse alternative* H_1, *si les suppositions (A1)-(A9) sont vérifiées et* $p = o(n^{1/3})$, *alors*

$$|\mathcal{Z}(\boldsymbol{\beta^0})| \xrightarrow[n\to\infty]{\mathbb{P}} \infty.$$

Preuve. Par des calculs élémentaires, sous l'hypothèse H_1, on a que

$$\boldsymbol{\psi}_n(\boldsymbol{\beta^0}) = \boldsymbol{\psi}_n^0 - \frac{1}{n(1-\theta)}\Big(\sum_{j=k+1}^{n} \mathbf{X}_j\mathbf{X}_j^t\Big)(\boldsymbol{\beta}_2^0 - \boldsymbol{\beta^0}),$$

avec une probabilité égale à 1.

Également sous H_1, puisque $\mathbb{E}[\mathbf{z}_i^0] = \mathbf{0}_p$, en utilisant la supposition (A1), on obtient

$$\boldsymbol{\psi}_n(\boldsymbol{\beta^0}) = -\frac{1}{n(1-\theta)}\Big(\sum_{j=k+1}^{n} \mathbf{X}_j\mathbf{X}_j^t\Big)(\boldsymbol{\beta}_2^0 - \boldsymbol{\beta^0})(1 + o_{\mathbb{P}}(1)).$$

Alors,

$$n\boldsymbol{\psi}_n^t(\boldsymbol{\beta^0})\big(\mathbf{V}_n^0\big)^{-1}\boldsymbol{\psi}_n(\boldsymbol{\beta^0}) = \Big(n\boldsymbol{\psi}_n^{0t}\big(\mathbf{V}_n^0\big)^{-1}\boldsymbol{\psi}_n^0 + n\Big|(\boldsymbol{\beta}_2^0-\boldsymbol{\beta^0})^t\frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{X}_j\mathbf{X}_j^t\Big|^2\Big)(1+o_{\mathbb{P}}(1)).$$

La statistique de test devient

$$\mathcal{Z}(\boldsymbol{\beta^0}) = \frac{n\boldsymbol{\psi}_n^{0t}\big(\mathbf{V}_n^0\big)^{-1}\boldsymbol{\psi}_n^0 - p}{\Delta_n/n} + n\frac{\Big|(\boldsymbol{\beta}_2^0 - \boldsymbol{\beta^0})^t\frac{1}{n(1-\theta)}\sum_{j=k+1}^{n}\mathbf{X}_j\mathbf{X}_j^t\Big|^2}{\Delta_n/n}(1 + o_{\mathbb{P}}(1)).$$

Puisque $\boldsymbol{\beta}_2^0 \neq \boldsymbol{\beta^0}$, en utilisant le Théorème 4.3.1 et par la supposition (A1), on obtient que

$$\big|\mathcal{Z}(\boldsymbol{\beta^0})\big| \to \infty,$$

en probabilité, quand $n \to \infty$. ■

Remarque 4.3.1. *Pour p fixé, [Liu et al. (2008)] ont proposé une statistique de test pour tester la présence du point de rupture dans le modèle linéaire en maximisant le rapport de vraisemblance empirique par rapport à $\boldsymbol{\lambda}$ et $\boldsymbol{\beta}$. Dans ce chapitre, on a fixé le paramètre de la première phase et on teste si celle de la deuxième phase est le même. Notons que, dans [Liu et al. (2008)] le système d'équations en $\boldsymbol{\lambda}$ et $\boldsymbol{\beta}$ de fonctions scores peut être résolu ce qui n'est pas possible dans notre cas puisque $p \to \infty$. Pour cela on va utiliser les propriétés théoriques du multiplicateur de Lagrange $\boldsymbol{\lambda}$ pour proposer une forme plus simple pour la statistique de vraisemblance empirique et qui sera utilisée dans la pratique. Pour les observations $1, \ldots, k$, si le paramètre $\boldsymbol{\beta^0}$ est inconnu, on peut l'estimer par des calculs simples pour obtenir son estimateur convergent $\widehat{\boldsymbol{\beta}}_k$, c'est-à-dire $\|\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta^0}\| = o_{\mathbb{P}}(1)$.*

4.4 Simulations

Afin d'évaluer la statistique de test spécifiée dans le Théorème 4.3.1*(ii)*, avec $\mathrm{EL}_{nk}(\boldsymbol{\beta^0})$ approximée par la Proposition 4.3.4, on réalise des simulations Monte-Carlo. On étudie le niveau empirique et la puissance empirique. Le niveau considéré est $\alpha = 0.05$.

Les p variables explicatives sont générées comme suit : $X_1 = 1$ et $(X_2, \ldots, X_p) \sim \mathcal{N}_{p-1}(\mathbf{0}_{p-1}, \mathbf{\Sigma})$, avec la matrice de covariance $\mathbf{\Sigma} = (\varsigma_{hl})$, $\varsigma_{hl} = 2^{-|h-l|}$, pour $1 \leq h, l \leq p-1$. La même matrice a été considéré par [Guo et al. (2013)] pour un modèle sans point de rupture. Afin d'avoir des régresseurs déterministes, on considère la même réalisation pour $(\mathbf{X}_i)_{1\leq i\leq n}$ pour chaque réplique Monte-Carlo.

Concernant les coefficients $\boldsymbol{\beta^0}$ du modèle, sous l'hypothèse H_0, on prend $\boldsymbol{\beta^0} = (\beta_1^0, \beta_2^0, \ldots, \beta_p^0) = (1, 2, \ldots, p)$.

On considère des valeurs différentes pour n et k et deux distributions différentes pour l'erreur (ε_i) : la distribution normale standard $\mathcal{N}(0, 1)$ et $\varepsilon \sim \mathcal{E}xp(1) - 1$, où $\mathcal{E}xp(1)$ est la distribution exponentielle de moyenne 1.

4.4.1 L'importance des suppositions (A5) et (A6)

Dans cette sous-section, pour chaque étude on réalise M=2000 répliques Monte-Carlo pour étudier le comportement de la statistique de test sous l'hypothèse nulle et ensuite quand le modèle a un point de rupture. On donne la puissance empirique et le niveau empirique et on les notes respectivement par $\widehat{\pi}$ et $\widehat{\alpha}$.

Analyse du niveaux empiriques.

Pour le modèle sous H_0 (sans point de rupture), afin de calculer le niveau empirique du test, tout d'abord, on calcule la matrice $\mathbf{V}_n^0$ en utilisant la relation (4.12). Une fois que le modèle est généré, on calcule $\mathbf{z}_i^0$ et ensuite $\mathbf{z}_i(\boldsymbol{\beta^0})$ par la relation (4.16). Finalement, on calcule $\boldsymbol{\psi}_n(\boldsymbol{\beta^0})$ par la relation (4.11), Δ_n par le Théorème 4.3.1 et la statistique de test $\mathcal{Z}(\boldsymbol{\beta^0})$ par (4.74). Le niveau empirique $\widehat{\alpha}$ représente le nombre de fois où $|\mathcal{Z}(\boldsymbol{\beta^0})| > c_{1-\alpha/2}$, divisé par M, où $c_{1-\alpha/2}$ est le quantile de la distribution normale standard et M désigne les répliques Monte-Carlo.

Dans le Tableau 4.4.1, pour des valeurs de n et k fixées, les niveaux empiriques augmentent avec la valeur de p, cette augmentation est plus nette dans le cas où les erreurs sont

Tableau 4.4.1 – Niveaux empiriques avec des erreurs exponentielles et gaussiennes.

n	k	p	erreurs exponentielles $\widehat{\alpha}$	erreurs gaussiennes $\widehat{\alpha}$
20	5	2	0.085	0.076
		3	0.110	0.106
		5	0.147	0.136
		7	0.308	0.253
100	25	2	0.092	0.068
		10	0.133	0.106
		20	0.182	0.176
200	75	2	0.074	0.059
		10	0.091	0.082
		20	0.122	0.096
		50	0.180	0.164
400	280	2	0.087	0.094
		10	0.157	0.128
		20	0.162	0.144
		50	0.171	0.168
		100	0.238	0.209
600	350	2	0.083	0.070
		10	0.099	0.090
		20	0.117	0.114
		50	0.148	0.126
		100	0.155	0.178
		200	0.281	0.250
		300	0.310	0.287
1000	350	2	0.071	0.057
		50	0.098	0.092
		100	0.143	0.110
		200	0.181	0.154
		300	0.219	0.207

exponentielles. Nous observons que les suppositions (A5) et (A6) ne sont pas satisfaites, les niveaux empiriques sont supérieurs à la niveau théorique pour les deux types des erreurs

(exponentielle et gaussienne).

Afin de confirmer cette supposition, dans le Tableau 4.4.2, on a varié les valeurs de n et k de sorte que $\theta = k/n = 1/2$ et p vérifie les suppositions (A5), (A6). Alors, on obtient des niveaux empiriques faibles et proches de la niveau théorique (moins de faux modèles).

Analyse de la puissance.

Sous l'hypothèse H_1, on considère $\boldsymbol{\beta}_2^0 = 1 - \boldsymbol{\beta}^{\mathbf{0}}$. Pour tous les cas considérés dans les Tableaux 4.4.1 et 4.4.2 (c'est à dire pour les mêmes valeurs de n, k, p et ε), on obtient que les puissances empiriques sont égales à 1.

Tableau 4.4.2 – Niveaux empiriques avec des erreurs exponentielles et gaussiennes, M= 2000, $\theta = 1/2$.

			erreurs exponentielles	erreurs gaussiennes
n	k	p	$\widehat{\alpha}$	$\widehat{\alpha}$
20	10	2	0.091	0.082
100	50	4	0.085	0.080
200	100	5	0.084	0.078
400	200	10	0.078	0.067
600	300	20	0.081	0.059
800	400	20	0.058	0.071
800	400	30	0.077	0.055
2000	1000	30	0.072	0.060

Amélioration du taux de recouvrement

Pour un niveau α, afin d'obtenir moins des faux point de ruptures ; pour la statistique de test $\mathcal{Z}(\boldsymbol{\beta}^{\mathbf{0}})$, on calcule par M=10000 répliques Monte-Carlo les fractiles d'ordre $(1-\alpha/2)$ et $\alpha/2$, notés respectivement par $\widehat{c}_1$ et $\widehat{c}_2$. On considère la nouvelle valeur critique suivante

$$\widehat{c}_{1-\alpha/2} \equiv \max(\widehat{c}_1, |\widehat{c}_2|).$$

Pour $p = 50$, on considère des différentes valeurs pour n et k, pour des erreurs $\mathcal{N}(0,1)$ et $\mathcal{E}xp(1) - 1$. Les nouvelles valeurs critiques sont données par le Tableau 4.4.3. Pour une

valeur de n fixée. Les nouvelles valeurs critiques ne sont pas influencées par la valeur de k.

Pour chaque n on prend deux valeurs différentes pour k, de sorte que $\theta_1 = k/n = 3/8$ et $\theta_2 = k/n = 5/8$. Ensuite, on calcule $\widehat{c}_{1-\alpha/2}$, les niveaux empiriques et les puissances empiriques et on les note respectivelent par $\widehat{\alpha}$ et $\widehat{\pi}$.

On remarque que les nouvelles valeurs critiques associées aux deux types des erreurs sont supérieures au quantile de la distribution de loi normale standard, alors que leurs valeurs diminuent quand n (et k) augmente et ils se rapprochent du quantile de $\mathcal{N}(0,1)$.

Tableau 4.4.3 – Valeur critique empirique $\hat{c}_{1-\alpha/2}$, niveaux empirique ($\hat{\alpha}$), puissances empirique ($\hat{\pi}$), pour des erreurs exponentielles et gaussiennes, M=2000, $p = 50$.

		erreurs exponentielles			erreurs gaussiennes		
n	k	$\hat{c}_{1-\alpha/2}$	$\hat{\alpha}$	$\hat{\pi}$	$\hat{c}_{1-\alpha/2}$	$\hat{\alpha}$	$\hat{\pi}$
200	75	4.01	0.03	1	3.24	0.03	1
	125		0.03	1		0.07	1
400	150	3.41	0.03	1	2.96	0.03	1
	250		0.03	1		0.02	1
600	225	3.40	0.03	1	2.85	0.03	1
	375		0.03	1		0.02	1
800	300	2.68	0.05	1	2.43	0.03	1
	500		0.05	1		0.05	1
2000	750	2.48	0.03	1	2.30	0.03	1
	1250		0.05	1		0.07	1

Remarque 4.4.1. *Pour les applications sur des données réelles, on teste le modèle (4.1) contre le modèle (4.2). Pour ces modèles, nous connaissons les n valeurs pour la variable réponse Y et pour les $p-1$ variables explicatives $X_2, \ldots, X_p$. Le point k, où l'on teste s'il y a de changement, est connu, tandis que les valeurs de $\boldsymbol{\beta^0}$ de la première phase peuvent être inconnues.*

Pour les observations $i = 1, \ldots, k$, si $\boldsymbol{\beta^0}$ est inconnu, alors on l'estime par un estimateur convergent en utilisant par exemple la méthode des moindres carrés ou la méthode quantile. Une fois $\boldsymbol{\beta^0}$ estimé par $\widehat{\boldsymbol{\beta}}_k$, on cherche un estimateur convergent pour la variance

σ^2 de ε, par exemple

$$\widehat{\sigma}_k^2 = \frac{1}{k-p}\sum_{i=1}^{k}(Y_i - \mathbf{X}_i^t\widehat{\boldsymbol{\beta}}_k)^2.$$

Ainsi, pour chaque $i = 1, \ldots, n$, on calcule $\mathbf{V}_{(i)}^0$, tel que

$$\mathbf{V}_{(i)}^0 = \widehat{\sigma}_k^2 \mathbf{X}_i \mathbf{X}_i^t$$

et ensuite $\mathbf{V}_n^0$ par la relation (4.12).

Pour chaque $i = 1, \ldots, n$, on calcule $\mathbf{z}_i(\widehat{\boldsymbol{\beta}}_k)$, par

$$\mathbf{z}_i(\widehat{\boldsymbol{\beta}}_k) = \mathbf{X}_i(Y_i - \mathbf{X}_i^t\widehat{\boldsymbol{\beta}}_k),$$

ce qui nous permet de calculer le vecteur $\boldsymbol{\psi}_n(\widehat{\boldsymbol{\beta}}_k)$ de la relation (4.11). Avec tous ces éléments, on peut calculer la valeur de la statistique $\mathcal{Z}(\widehat{\boldsymbol{\beta}}_k)$, telle que

$$\mathcal{Z}(\widehat{\boldsymbol{\beta}}_k) = \left(n\boldsymbol{\psi}_n^t(\widehat{\boldsymbol{\beta}}_k)\left(\mathbf{V}_n^0\right)^{-1}\boldsymbol{\psi}_n(\widehat{\boldsymbol{\beta}}_k) - p\right)(\Delta_n/n)^{-1}.$$

Pour calculer Δ_n, on utilise la relation donnée par le Théorème 4.3.1. Pour un niveau donné $\alpha \in (0,1)$, si la valeur de $|\mathcal{Z}(\widehat{\boldsymbol{\beta}}_k)|$ est inférieure à $c_{1-\alpha/2}$, alors l'hypothèse H_0 est acceptée, ce qui signifie que le modèle ne change pas après l'observation k, sinon on accepte l'hypothèse H_1.

Si $\boldsymbol{\beta}^{\mathbf{0}}$ est connu, on définit $\widehat{\sigma}_k^2$ comme un estimateur de σ^2, tel que

$$\widehat{\sigma}_k^2 = (k-p)^{-1}\sum_{i=1}^{k}(Y_i - \mathbf{X}_i^t\boldsymbol{\beta}^{\mathbf{0}})^2.$$

Pour chaque $i = 1, \ldots, n$, on calcule $\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}})$, tel que

$$\mathbf{z}_i(\boldsymbol{\beta}^{\mathbf{0}}) = \mathbf{X}_i(Y_i - \mathbf{X}_i^t\boldsymbol{\beta}^{\mathbf{0}}).$$

Ensuite, on calcule $\boldsymbol{\psi}_n(\boldsymbol{\beta}^{\mathbf{0}})$ par la relation (4.11). Finalement, on compare la valeur absolue de $\mathcal{Z}(\boldsymbol{\beta}^{\mathbf{0}})$ avec $c_{1-\alpha/2}$.

4.5 Annexe

Nous rappelons que sous l'hypothèse H_0, le vecteur $\mathbf{z}_i(\boldsymbol{\beta^0})$ est $\mathbf{z}_i^0 = \mathbf{X}_i\varepsilon_i$. Alors, dans toutes les preuves, si l'hypothèse H_0 est vraie, on utilise $\mathbf{z}_i^0$ à la place de $\mathbf{z}_i(\boldsymbol{\beta^0})$.

Les lemmes suivants sont utilisés dans les preuves des Propositions 4.3.1, 4.3.2, 4.3.3, 4.3.4 et des Théorèmes 4.3.1 et 4.3.2.

Les quatre premiers lemmes désignent des résultats équivalents à ceux obtenus par [Guo et al. (2013)] pour le modèle linéaire sans point de rupture.

On considère la matrice suivante $\mathbf{L}_n \equiv \mathbf{S}_n^0 - \mathbf{V}_n^0$, de dimension $p \times p$. Pour $1 \leq u, v \leq p$, $L_{n;u,v}$ désigne l'élément (u, v) de la matrice $\mathbf{L}_n$. On considère également pour $\mathbf{L}_n$: $\max\limits_{1\leq u,v\leq p} |S^0_{n,(u,v)} - V^0_{n,(u,v)}|$, le plus grand élément de la matrice $\mathbf{L}_n$. On note par $S^0_{n,(u,v)}$ et $V^0_{n,(u,v)}$ respectivement l'élément (u, v) de $\mathbf{S}_n^0$ et $\mathbf{V}_n^0$.

Lemme 4.5.1. *Sous l'hypothèse nulle* H_0*, si les suppositions (A3) et (A4) sont vérifiées, alors pour chaque* $\epsilon > 0$*, il existe une constante positive* C_q *qui dépend seulement de* $q \geq 4$*, telle que*

$$\mathbb{P}\big(\max_{1\leq u,v\leq p} |S^0_{n,(u,v)} - V^0_{n,(u,v)}| \geq \epsilon\big) \leq C_q \frac{p^2}{n^{q/2}\epsilon^q}.$$

Preuve. La matrice $\mathbf{L}_n$ peut s'écrire

$$\begin{aligned}
\mathbf{L}_n &= \mathbf{S}_n^0 - \mathbf{V}_n^0 \\
&= \left[\frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{z}_i^0\mathbf{z}_i^{0t} + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{z}_j^0\mathbf{z}_j^{0t}\right] - \left[\frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{V}_{(i)}^0 + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{V}_{(j)}^0\right] \\
&= \left[\frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{X}_i\mathbf{X}_i^t\varepsilon_i^2 + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{X}_j\mathbf{X}_j^t\varepsilon_j^2\right] \\
&\quad -\left[\frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{X}_i\mathbf{X}_i^t\sigma^2 + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{X}_j\mathbf{X}_j^t\sigma^2\right] \\
&= \left[\frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{X}_i\mathbf{X}_i^t(\varepsilon_i^2 - \sigma^2) + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{X}_j\mathbf{X}_j^t(\varepsilon_j^2 - \sigma^2)\right].
\end{aligned}$$

Pour $1 \leq u, v \leq p$, l'élément (u, v) de la matrice $\mathbf{L}_n$ est

$$\begin{aligned} L_{n;u,v} &= \left[\frac{1}{n\theta^2}\sum_{i=1}^{k} X_{i,u}X_{i,v}(\varepsilon_i^2 - \sigma^2) + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} X_{j,u}X_{j,v}(\varepsilon_j^2 - \sigma^2)\right] \\ &\equiv L_{n;u,v}^{(1)} + L_{n;u,v}^{(2)}. \end{aligned} \tag{4.75}$$

Pour le terme $L_{n;u,v}^{(1)}$ de la relation (4.75), puisque $\theta = k/n \to \theta^0 \in (0,1)$ quand $n \to \infty$, par l'inégalité de Markov, pour chaque $\epsilon > 0$ et $q > 0$, on a

$$\mathbb{P}\big(|L_{n;u,v}^{(1)}| \geq \epsilon\big) \leq \frac{\mathbb{E}\big[|\sqrt{n\theta}L_{n;u,v}^{(1)}|^q\big]}{(\sqrt{n\theta}\epsilon)^q} = \frac{n^{-q/2}\theta^{-3q/2}\mathbb{E}\big[|\sum_{i=1}^{k} X_{i,u}X_{i,v}(\varepsilon_i^2 - \sigma^2)|^q\big]}{(n\theta)^{q/2}\epsilon^q}. \tag{4.76}$$

Par le Théorème 2 (page 26) de [Dharmadhikari and Jogdeo (1969)], il existe une constante positive $\tilde{C}_q > 0$, qui dépend seulement de q, où $q \geq 4$, telle que

$$\begin{aligned} \mathbb{E}\big[\big|\sum_{i=1}^{k} X_{i,u}X_{i,v}(\varepsilon_i^2 - \sigma^2)\big|^q\big] &\leq \tilde{C}_q(n\theta)^{(q/2)-1}\sum_{i=1}^{k}\mathbb{E}\big[\big|X_{i,u}X_{i,v}(\varepsilon_i^2 - \sigma^2)\big|^q\big] \\ &\leq \tilde{C}_q\mathbb{E}\big[|\varepsilon_1|^{2q}\big](n)^{(q/2)-1}\sum_{i=1}^{k}\big|X_{i,u}|^q\big|X_{i,v}\big|^q. \end{aligned} \tag{4.77}$$

Tenant compte de l'inégalité (4.76), alors, il existe une constante positive $C > 0$ telle que

$$\mathbb{P}\big(|L_{n;u,v}^{(1)}| \geq \epsilon\big) \leq C\,\tilde{C}_q\mathbb{E}\big[|\varepsilon_1|^{2q}\big]\frac{n^{-1}\sum_{i=1}^{k}\big|X_{i,u}|^q\big|X_{i,v}\big|^q}{n^{q/2}\epsilon^q}.$$

En utilisant les suppositions (A3) et (A4), on obtient que

$$\mathbb{P}\big(|L_{n;u,v}^{(1)}| \geq \epsilon\big) \leq C_q^{(1)}\frac{p^2}{n^{q/2}\epsilon^q}. \tag{4.78}$$

Pour une valeur fixée de $\theta \in (0,1)$ et $j = k+1, \ldots, n$, de la même façon que pour $i = 1, \ldots, k$, on peut démontrer que

$$\mathbb{P}\big(|L_{n;u,v}^{(2)}| \geq \epsilon\big) \leq C\,\tilde{\tilde{C}}_q\mathbb{E}\big[|\varepsilon_1|^{2q}\big]\frac{n^{-1}\sum_{j=k+1}^{n}\big|X_{j,u}|^q\big|X_{j,v}\big|^q}{n^{q/2}\epsilon^q},$$

et

$$\mathbb{P}\big(|L^{(2)}_{n;u,v}| \geq \epsilon\big) \leq C^{(2)}_q \frac{p^2}{n^{q/2}\epsilon^q}. \tag{4.79}$$

On considère

$$C_q = \max\left(C^{(1)}_q, C^{(2)}_q\right). \tag{4.80}$$

En combinant les relations (4.78), (4.79) et (4.80), le lemme résulte. ■

Par le lemme suivant, on prouve que toutes les valeurs propres de $\mathbf{S}^0_n$ convergent uniformément vers celles de $\mathbf{V}^0_n$ avec une vitesse $O_{\mathbb{P}}(p \max_{1\leq u,v\leq p} |S^0_{n,(u,v)} - V^0_{n,(u,v)}|)$, ainsi que toutes les valeurs propres de $\mathbf{S}^0_n$ sont bornées et strictement positives pour une valeur assez grande de n.

Lemme 4.5.2. *Sous les suppositions (A1), (A3), (A4) et (A6), il existe deux constantes positives C_0, et C_1 telles que les inégalités $\gamma_p(\mathbf{S}^0_n) > C_0$ et $\gamma_1(\mathbf{S}^0_n) < C_1$ sont vérifiées avec une probabilité qui tend vers 1 quand $n \to \infty$.*

Preuve. On sait que

$$\gamma^{1/2}_r(\mathbf{V}^0_n)^2 = \gamma_r(\mathbf{V}^0_n)$$

et

$$\gamma^{1/2}_r(\mathbf{S}^0_n)^2 = \gamma_r(\mathbf{S}^0_n).$$

Ce qui donne

$$\begin{aligned}
\max_{1\leq r\leq p} |\gamma_r(\mathbf{V}^0_n) - \gamma_r(\mathbf{S}^0_n)|^2 &\leq \sum_{r=1}^{p} \left|\gamma^{1/2}_r(\mathbf{V}^0_n)^2 - \gamma^{1/2}_r(\mathbf{S}^0_n)^2\right|^2 \\
&= \sum_{r=1}^{p} \gamma_r(\mathbf{V}^0_n)^2 + \sum_{r=1}^{p} \gamma_r(\mathbf{S}^0_n)^2 - 2\sum_{r=1}^{p} \gamma_r(\mathbf{V}^0_n)\gamma_r(\mathbf{S}^0_n) \\
&= tr(\mathbf{V}^0_n)^2 + tr(\mathbf{S}^0_n)^2 - 2\sum_{r=1}^{p} \gamma_r(\mathbf{V}^0_n)\gamma_r(\mathbf{S}^0_n).
\end{aligned}$$

Par l'inégalité de Von Neymann, on a

$$tr(\mathbf{V}^0_n\mathbf{S}^0_n) \leq \sum_{r=1}^{p} \gamma_r(\mathbf{V}^0_n)\gamma_r(\mathbf{S}^0_n).$$

Alors,

$$\begin{aligned}\max_{1\leq r\leq p}\left|\gamma_r(\mathbf{V}_n^0)-\gamma_r(\mathbf{S}_n^0)\right|^2 &\leq tr(\mathbf{V}_n^0)^2+tr(\mathbf{S}_n^0)^2-2tr(\mathbf{V}_n^0\,\mathbf{S}_n^0)\\ &= tr(\mathbf{V}_n^0-\mathbf{S}_n^0)^2,\end{aligned}$$

ce qui donne

$$\max_{1\leq r\leq p}\left|\gamma_r(\mathbf{V}_n^0)-\gamma_r(\mathbf{S}_n^0)\right|\leq[tr(\mathbf{V}_n^0-\mathbf{S}_n^0)^2]^{1/2}\leq p\max_{1\leq u,v\leq p}|V_{n;u,v}^0-S_{n;u,v}^0|. \tag{4.81}$$

La dernière relation est due au fait que la valeur absolue de chaque élément de $(\mathbf{V}_n^0-\mathbf{S}_n^0)^2$ est inférieure à $(\max_{1\leq u,v\leq p}|V_{n;u,v}^0-S_{n;u,v}^0|)^2$. ■

On considère la variable aléatoire suivante

$$T_n^0\equiv\max_{\substack{1\leqslant i\leqslant k\\ k+1\leqslant j\leqslant n}}\left\{\frac{1}{\theta}||\mathbf{z}_i^0||,\frac{1}{1-\theta}||\mathbf{z}_j^0||\right\}.$$

Lemme 4.5.3. *Sous l'hypothèse nulle* H_0*, si les suppositions (A3) et (A4) sont vérifiées, alors, pour* $q\geq 4$*, on a*

$$T_n^0=o_{\mathbb{P}}(p^{1/2}n^{1/q}).$$

Preuve. Puisque $\theta=k/n\to\theta^0\in(0,1)$ quand $n\to\infty$, pour une valeur assez grande de n, on a

$$T_n^0\leq C\max_{1\leq i\leq n}||\mathbf{z}_i^0||,$$

avec une probabilité égale à 1, avec $C>0$. Par le Lemme 3 de [Guo et al. (2013)], on a que $\max_{1\leq i\leq n}||\mathbf{z}_i^0||=o_{\mathbb{P}}(p^{1/2}n^{1/q})$ ce qui termine la preuve. ■

Par le Lemme suivant, on donne l'approximation asymptotique de la norme L_2 du vecteur $\boldsymbol{\psi}_n(\boldsymbol{\beta^0})$, donné par (4.11), sous l'hypothèse H_0.

Lemme 4.5.4. *Sous l'hypothèse nulle* H_0*, si la supposition (A1) est vérifiée, alors*

$$||\boldsymbol{\psi}_n^0||=O_{\mathbb{P}}(p^{1/2}n^{-1/2}).$$

Preuve. Soit $\mathbf{X}^{(1)}$ une matrice de dimension $p \times k$ où les k colonnes sont les variables déterministes $\mathbf{X}_i$, pour $i = 1, \ldots, k$. Soit également $\mathbf{X}^{(2)}$ une matrice de dimension $p \times (n-k)$, où les $(n-k)$ colonnes sont les variables déterministes $\mathbf{X}_j$, pour $j = k+1, \ldots, n$. Puisque les (ε_i) sont indépendantes, alors, on a que

$$\begin{aligned}
\mathbb{E}\big[(\boldsymbol{\psi}_n^0)^t \boldsymbol{\psi}_n^0\big] &= \mathbb{E}\left[\left(\frac{1}{k}\sum_{i=1}^{k} \mathbf{z}_i^{0t} - \frac{1}{n-k}\sum_{j=k+1}^{n} \mathbf{z}_j^{0t}\right)\left(\frac{1}{k}\sum_{i=1}^{k} \mathbf{z}_i^{0} - \frac{1}{n-k}\sum_{j=k+1}^{n} \mathbf{z}_j^{0}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k^2}\sum_{i,i'=1}^{k} \mathbf{z}_i^{0t}\mathbf{z}_{i'}^{0}\right] + \mathbb{E}\left[\frac{1}{(n-k)^2}\sum_{j,j'=k+1}^{n} \mathbf{z}_j^{0t}\mathbf{z}_{j'}^{0}\right] \\
&= \frac{1}{k^2}\sum_{i,i'=1}^{k} \mathbf{X}_i^t\mathbf{X}_{i'}\mathbb{E}(\varepsilon_i\varepsilon_{i'}) + \frac{1}{(n-k)^2}\sum_{j,j'=k+1}^{n} \mathbf{X}_j^t\mathbf{X}_{j'}\mathbb{E}(\varepsilon_j\varepsilon_{j'}) \\
&= \frac{\sigma^2}{n}\left[\frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{X}_i^t\mathbf{X}_i + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{X}_j^t\mathbf{X}_j\right] \\
&= \frac{\sigma^2}{n}\left[tr\big(\frac{1}{n\theta^2}\sum_{i=1}^{k} \mathbf{X}_i^t\mathbf{X}_i\big) + tr\big(\frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n} \mathbf{X}_j^t\mathbf{X}_j\big)\right].
\end{aligned}$$

En utilisant la supposition (A1), on obtient que

$$\mathbb{E}\big[\boldsymbol{\psi}_n^{0t}\boldsymbol{\psi}_n^0\big] \leq \sigma^2\frac{p}{n}\left[\gamma_1\Big(\frac{1}{n\theta}\mathbf{X}^{(1)}(\mathbf{X}^{(1)})^t + \frac{1}{n(1-\theta)}\mathbf{X}^{(2)}(\mathbf{X}^{(2)})^t\Big)\right] = O(p\,n^{-1}).$$

Alors,

$$||\boldsymbol{\psi}_n^0|| = O_{\mathbb{P}}(p^{1/2}n^{-1/2}).$$

■

Le Lemme suivant donne une première approximation pour la statistique de vraisemblance empirique, sous l'hypothèse H_0.

Lemme 4.5.5. *Sous les mêmes suppositions de la Proposition 4.3.3, si l'hypothèse* H_0 *est vraie, alors*

$$\mathbb{E}\mathrm{L}_{nk}(\boldsymbol{\beta^0}) = 2n\boldsymbol{\lambda}^t\boldsymbol{\psi}_n^0 - n\boldsymbol{\lambda}^t\mathbf{S}_n^0\boldsymbol{\lambda} + \frac{2}{3}\left[\frac{1}{\theta^3}\sum_{i=1}^{k}(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^3 - \frac{1}{(1-\theta)^3}\sum_{j=k+1}^{n}(\mathbf{z}_j^{0t}\boldsymbol{\lambda})^3\right] + o_{\mathbb{P}}(1),$$

avec $\theta = k/n$.

Preuve. Le développement limité de la statistique $\mathbb{EL}_{nk}(\boldsymbol{\beta^0})$, spécifiée par la relation (4.8), dans le voisinage de $\boldsymbol{\lambda} = \mathbf{0}_p$, jusqu'à l'ordre 3 peut s'écrire

$$\begin{aligned}
\mathbb{EL}_{nk}(\boldsymbol{\beta^0}) &= \boldsymbol{\lambda}^t\left[\frac{2}{\theta}\sum_{i=1}^{k}\mathbf{z}_i^0 - \frac{2}{1-\theta}\sum_{j=k+1}^{n}\mathbf{z}_j^0\right] - \boldsymbol{\lambda}^t\left[\frac{1}{\theta^2}\sum_{i=1}^{k}\mathbf{z}_i^0\mathbf{z}_i^{0t} + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathbf{z}_j^0\mathbf{z}_j^{0t}\right]\boldsymbol{\lambda} \\
&\quad + \frac{2}{3}\left[\frac{1}{\theta^3}\sum_{i=1}^{k}(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^3 - \frac{1}{(1-\theta)^3}\sum_{j=k+1}^{n}(\mathbf{z}_j^{0t}\boldsymbol{\lambda})^3\right] \\
&\quad + \frac{1}{4!}\left[\sum_{u,v,r,s=1}^{p}\frac{\partial^4\mathbb{EL}_{nk}(\boldsymbol{\beta^0})(\tilde{\boldsymbol{\lambda}}_{uvrs})}{\partial\lambda_u\partial\lambda_v\partial\lambda_r\partial\lambda_s}(\lambda_u)(\lambda_v)(\lambda_r)(\lambda_s)\right] \\
&\equiv \mathcal{E}_1 - \mathcal{E}_2 + \mathcal{E}_3 + \mathcal{E}_4, \qquad (4.82)
\end{aligned}$$

où pour tout $1 \leq u, v, r, s \leq p$, λ_u est la i-ème composante de $\boldsymbol{\lambda}$ et $\tilde{\boldsymbol{\lambda}}_{uvrs} = a_{uvrs}\boldsymbol{\lambda}$, avec $a_{uvrs} \in [0, 1]$.

Tout d'abord, on étudie $\mathcal{E}_4$ qui peut s'écrire

$$\mathcal{E}_4 = -\frac{1}{4!}\left[\frac{12}{\theta^4}\sum_{i=1}^{k}\frac{(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^4}{(1+\theta^{-1}\tilde{\boldsymbol{\lambda}}\mathbf{z}_i^0)^4} + \frac{12}{(1-\theta)^4}\sum_{j=k+1}^{n}\frac{(\mathbf{z}_j^{0t}\boldsymbol{\lambda})^4}{(1-(1-\theta)^{-1}\tilde{\boldsymbol{\lambda}}\mathbf{z}_i^0)^4}\right]. \qquad (4.83)$$

Par la Proposition 2.3.1 du Chapitre 2, pour tout $\epsilon > 0$, il existe deux constantes positives M_1 et M_2 telles que

$$\mathbb{P}\left[\frac{1}{M_1\theta^4}\sum_{i=1}^{k}(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^4 \leq \sum_{i=1}^{k}\frac{(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^4}{(1+\theta^{-1}\tilde{\boldsymbol{\lambda}}\mathbf{z}_i^0)^4} \leq \frac{1}{M_2\theta^4}\sum_{i=1}^{k}(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^4\right] \geq 1-\epsilon.$$

Alors, pour le premier terme du membre droit de la relation (4.83), en appliquant l'inégalité de Cauchy-Schwartz, on obtient que

$$\sum_{i=1}^{k}\frac{(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^4}{(1+\theta^{-1}\tilde{\boldsymbol{\lambda}}\mathbf{z}_i^0)^4} \leq \frac{1}{M_2\theta^4}\sum_{i=1}^{k}(\mathbf{z}_i^{0t}\boldsymbol{\lambda})^4 \leq \frac{1}{M_2\theta^4}\sum_{i=1}^{k}||\mathbf{X}_i||^4\varepsilon_i^4||\boldsymbol{\lambda}||^4.$$

En utilisant la supposition (A2) et le fait que $||\boldsymbol{\lambda}|| = O_{\mathbb{P}}(p^{1/2}n^{-1/2})$ donné par la Proposition 4.3.1 et $p = o(n^{1/2})$, on obtient que le premier terme de (4.83) est $o_{\mathbb{P}}(1)$.

De la même façon, on peut démontrer que chaque terme de (4.83) est $o_{\mathbb{P}}(1)$, ce qui

implique que $\mathcal{E}_4 = o_{\mathbb{P}}(1)$.

En utilisant les notations données par les relations (4.17) et (4.18), on obtient que

$$\mathcal{E}_1 = 2n\boldsymbol{\lambda}^t\boldsymbol{\psi}_n^0$$

et

$$\mathcal{E}_2 = n\boldsymbol{\lambda}^t\mathbf{S}_n^0\boldsymbol{\lambda}.$$

Alors, la relation (4.82) devient

$$\mathrm{EL}_{nk}(\boldsymbol{\beta^0}) = 2n\boldsymbol{\lambda}^t\boldsymbol{\psi}_n^0 - n\boldsymbol{\lambda}^t\mathbf{S}_n^0\boldsymbol{\lambda} + \mathcal{E}_3 + o_{\mathbb{P}}(1)$$

et le Lemme résulte. ■

Sous l'hypothèse nulle H_0, le lemme suivant donne une approximation pour le multiplicateur de Lagrange $\boldsymbol{\lambda}$ ainsi que le comportement asymptotique de $\sup_{1\le i\le n}|\boldsymbol{\lambda}^t\mathbf{z}_i^0| = o_{\mathbb{P}}(1)$.

Lemme 4.5.6. *Sous l'hypothèse nulle* H_0*, si les suppositions (A1), (A3)-(A6) sont vérifiées, alors, pour* $q \ge 4$ *et une valeur fixée de* $\theta \in (0,1)$*, on a*

$$\sup_{1\le i\le n}|\boldsymbol{\lambda}^t\mathbf{z}_i^0| = o_{\mathbb{P}}(1)$$

et

$$\boldsymbol{\lambda} = (\mathbf{S}_n^0)^{-1}(\boldsymbol{\psi}_n^0 + o_{\mathbb{P}}(n^{(1-q)/q}p^{3/2})). \tag{4.84}$$

Preuve. Par la Proposition 4.3.1, on a que

$$||\boldsymbol{\lambda}|| = O_{\mathbb{P}}(n^{-1/2}p^{1/2}).$$

Notons que, par l'inégalité de Cauchy-Schwartz et par le Lemme 4.5.3, on a

$$\frac{1}{\theta}\sup_{1\le i\le k}|\boldsymbol{\lambda}^t\mathbf{z}_i^0| \le ||\boldsymbol{\lambda}||\,T_n^0 = O_{\mathbb{P}}(n^{-1/2}p^{1/2})\;o_{\mathbb{P}}(n^{1/q}p^{1/2}) = o_{\mathbb{P}}(p\,n^{(-q+2)/2q}) \tag{4.85}$$

et

$$\frac{1}{1-\theta}\sup_{k+1\le j\le n}|\boldsymbol{\lambda}^t\mathbf{z}_j^0| \le ||\boldsymbol{\lambda}||\,T_n^0 = O_{\mathbb{P}}(n^{-1/2}p^{1/2})\;o_{\mathbb{P}}(n^{1/q}p^{1/2}) = o_{\mathbb{P}}(p\,n^{(-q+2)/2q}), \tag{4.86}$$

avec $\theta = k/n$. Les deux relations (4.85) et (4.86) ainsi que la supposition (A5) impliquent

$$\sup_{1\leq i\leq n} |\boldsymbol{\lambda}^t \mathbf{z}_i^0| = o_{\mathbb{P}}(1).$$

Maintenant, on prouve la relation (4.84).

Le développement limité de (4.9), dans le voisinage de $\boldsymbol{\lambda} = \mathbf{0}_p$, jusqu'à l'ordre 3 peut s'écrire

$$\begin{aligned}
\mathbf{0}_p \;=\; & \Big[\frac{1}{\theta}\sum_{i=1}^{k}\mathbf{z}_i^0 - \frac{1}{1-\theta}\sum_{j=k+1}^{n}\mathbf{z}_j^0\Big] - \Big[\frac{1}{\theta^2}\sum_{i=1}^{k}\mathbf{z}_i^0\mathbf{z}_i^{0t} + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathbf{z}_j^0\mathbf{z}_j^{0t}\Big]\boldsymbol{\lambda} \\
& + \Big[\frac{1}{\theta^3}\sum_{i=1}^{k}\mathbf{z}_i^0\left(\mathbf{z}_i^{0t}\boldsymbol{\lambda}\right)^2 - \frac{1}{(1-\theta)^3}\sum_{j=k+1}^{n}\mathbf{z}_j^0\left(\mathbf{z}_j^{0t}\boldsymbol{\lambda}\right)^2\Big] \\
& + \Big[\frac{1}{(\theta+\tilde{\boldsymbol{\lambda}}^t\mathbf{z}_i^0)^4}\sum_{i=1}^{k}\mathbf{z}_i^0\left(\mathbf{z}_i^{0t}\boldsymbol{\lambda}\right)^3 - \frac{1}{(1-\theta-\tilde{\boldsymbol{\lambda}}^t\mathbf{z}_j^0)^4}\sum_{j=k+1}^{n}\mathbf{z}_j^0\left(\mathbf{z}_j^{0t}\boldsymbol{\lambda}\right)^3\Big],
\end{aligned} \tag{4.87}$$

où $\tilde{\boldsymbol{\lambda}} = u\boldsymbol{\lambda}$, avec $u \in (0,1)$.

De même façon que pour le terme $\mathcal{E}_4$ de la Proposition 4.3.3, en utilisant la Proposition 2.3.1 du Chapitre 2, on peut démontrer que le dernier terme du membre droit de la relation (4.87) est $o_{\mathbb{P}}(1)$.

Nous rappelons que

$$\mathbf{R}_n^0 \equiv \frac{1}{n\theta^3}\sum_{i=1}^{k}\mathbf{z}_i^0(\boldsymbol{\lambda}^t\mathbf{z}_i^0)^2 - \frac{1}{n(1-\theta)^3}\sum_{j=k+1}^{n}\mathbf{z}_j^0(\boldsymbol{\lambda}^t\mathbf{z}_j^0)^2. \tag{4.88}$$

En utilisant les notations données par les relations (4.17), (4.18) et (4.88), la relation (4.87) devient

$$\boldsymbol{\psi}_n^0 - \mathbf{S}_n^0\boldsymbol{\lambda} + \mathbf{R}_n^0 = \mathbf{0}_p.$$

Ainsi

$$\boldsymbol{\lambda} = (\mathbf{S}_n^0)^{-1}(\mathbf{R}_n^0 + \boldsymbol{\psi}_n^0)(1 + o_{\mathbb{P}}(1)). \tag{4.89}$$

Nous rappelons que

$$T_n^0 \equiv \max_{\substack{1\leqslant i\leqslant k\\ k+1\leqslant j\leqslant n}} \left\{\frac{1}{\theta}||\mathbf{z}_i^0||, \frac{1}{1-\theta}||\mathbf{z}_j^0||\right\}.$$

Alors, pour $\mathbf{R}_n^0$, on a :

$$\begin{aligned}
\big|\big|\mathbf{R}_n^0\big|\big| &\leq \Big[\frac{1}{n\theta^3}\sum_{i=1}^{k}\big|\big|\mathbf{z}_i^0\big|\big|\,|\mathbf{z}_i^{0t}\boldsymbol{\lambda}|^2 + \frac{1}{n(1-\theta)^3}\sum_{j=k+1}^{n}\big|\big|\mathbf{z}_j^0\big|\big|\,|\mathbf{z}_j^{0t}\boldsymbol{\lambda}|^2\Big] \\
&\leq T_n^0\boldsymbol{\lambda}^t\big[\frac{1}{n\theta^2}\sum_{i=1}^{k}\mathbf{z}_i^0\mathbf{z}_i^{0t} + \frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n}\mathbf{z}_j^0\mathbf{z}_j^{0t}\big]\boldsymbol{\lambda} \\
&\leq T_n^0\boldsymbol{\lambda}^t\mathbf{S}_n^0\boldsymbol{\lambda}.
\end{aligned}$$

En utilisant le Lemme 4 de [Liu et al(2013)] et le Lemme 4.5.2, on obtient que

$$\big|\big|\mathbf{R}_n^0\big|\big| \leq T_n^0||\boldsymbol{\lambda}||^2\gamma_1(\mathbf{S}_n^0) = O_{\mathbb{P}}(T_n^0||\boldsymbol{\lambda}||^2).$$

D'autre part, par la Proposition 4.3.1, on a que

$$||\boldsymbol{\lambda}|| = O_{\mathbb{P}}(p^{1/2}n^{-1/2})$$

et par le Lemme 4.5.3 que

$$T_n^0 = o_{\mathbb{P}}(p^{1/2}n^{1/q}).$$

Alors, pour $||\mathbf{R}_n^0||$, on obtient que

$$\begin{aligned}
||\mathbf{R}_n^0|| &= O_{\mathbb{P}}(T_n^0\,||\boldsymbol{\lambda}||^2) \\
&= O_{\mathbb{P}}(n^{-1}p)\,o_{\mathbb{P}}(n^{1/q}p^{1/2}) \\
&= o_{\mathbb{P}}(p^{3/2}n^{(1-q)/q}). \qquad (4.90)
\end{aligned}$$

La relation (4.84) résulte en combinant les relations (4.89) et (4.90). ■

Le résultat suivant donne une approximation asymptotique pour $\big[(\mathbf{S}_n^0)^{-1} - (\mathbf{V}_n^0)^{-1}\big]\boldsymbol{\psi}_n^0$ et $\big[(\mathbf{S}_n^0)^{-1} - (\mathbf{V}_n^0)^{-1}\big]\mathbf{R}_n^0$.

Lemme 4.5.7. *Sous l'hypothèse nulle* (H_0)*, si les suppositions (A3), (A4) et (A6) sont vérifiées, alors*

(i) $\big[(\mathbf{S}_n^0)^{-1} - (\mathbf{V}_n^0)^{-1}\big]\boldsymbol{\psi}_n^0 = \big[(\mathbf{V}_n^0)^{-1}\boldsymbol{\psi}_n^0\big]o_{\mathbb{P}}(1)$.

(ii) $\big[(\mathbf{S}_n^0)^{-1} - (\mathbf{V}_n^0)^{-1}\big]\mathbf{R}_n^0 = \big[(\mathbf{V}_n^0)^{-1}\mathbf{R}_n^0\big]o_{\mathbb{P}}(1)$.

Preuve. (i) En utilisant les suppositions (A3) et (A4) et le Lemme 4.5.1, pour tout $\epsilon > 0$, pour $q \geq 4$, il existe $C_q > 0$, telle que

$$\mathbb{P}\big(p \max_{1\leq u,v\leq p} |S^0_{n,(u,v)} - V^0_{n,(u,v)}| \geq \epsilon\big) \leq C_q \frac{p^{2+q}}{n^{q/2}\epsilon^q}.$$

Donc, sous la supposition (A6), on a

$$\max_{1\leq u,v\leq p} |S^0_{n,(u,v)} - V^0_{n,(u,v)}| = o_{\mathbb{P}}(1). \tag{4.91}$$

Nous rappelons que, pour la matrice $\mathbf{A}$, $||\mathbf{A}||_1$ est la norme subordonnée à la norme vectorielle $||.||_1$.
En utilisant le Lemme 2.6.1 (iii) du Chapitre 2, le Lemme 4.5.2, les relations (4.81), (4.91) et l'identité $\mathbf{V}^0_n[(\mathbf{S}^0_n)^{-1} - (\mathbf{V}^0_n)^{-1}]\boldsymbol{\psi}^0_n = [\mathbf{V}^0_n(\mathbf{S}^0_n)^{-1} - \mathbf{I}_p]\boldsymbol{\psi}^0_n$, on obtient que

$$\begin{aligned}
\big|\big|\mathbf{V}^0_n\big[(\mathbf{S}^0_n)^{-1} - (\mathbf{V}^0_n)^{-1}\big]\boldsymbol{\psi}^0_n\big|\big| &= \big|\big|\big[\mathbf{V}^0_n - \mathbf{S}^0_n\big](\mathbf{S}^0_n)^{-1}\boldsymbol{\psi}^0_n\big|\big| \\
&\leq \big|\big|\mathbf{V}^0_n - \mathbf{S}^0_n\big|\big|_1 \ \big|\big|(\mathbf{S}^0_n)^{-1}\big|\big|_1 \ \big|\big|\boldsymbol{\psi}^0_n\big|\big| \\
&\leq \max_{1\leq r\leq p} |\gamma_r(\mathbf{V}^0_n - \mathbf{S}^0_n)| \ |\gamma_1(\mathbf{S}^0_n)^{-1}| \ \big|\big|\boldsymbol{\psi}^0_n\big|\big| \\
&\leq C\, p \max_{1\leq u,v\leq p} |S^0_{n,(u,v)} - V^0_{n,(u,v)}| \ ||\boldsymbol{\psi}^0_n|| \\
&= \big|\big|\boldsymbol{\psi}^0_n\big|\big|\, o_{\mathbb{P}}(1),
\end{aligned}$$

ce qui implique que

$$\big((\mathbf{S}^0_n)^{-1} - (\mathbf{V}^0_n)^{-1}\big)\,\boldsymbol{\psi}^0_n = (\mathbf{V}^0_n)^{-1}\boldsymbol{\psi}^0_n o_{\mathbb{P}}(1).$$

(ii) La preuve de *(ii)* est similaire à celle de *(i)*. ■

Le lemme suivant est utilisé pour prouver la Proposition 4.3.4.

Nous rappelons que

$$\mathbf{K}^0_n = \mathbf{I}_p - (\mathbf{V}^0_n)^{-1/2}\,\mathbf{S}^0_n\,(\mathbf{V}^0_n)^{-1/2}.$$

Lemme 4.5.8. *Sous l'hypothèse nulle* H_0*, si la supposition (A7) est vérifiée, alors*

$$tr(\mathbf{K}^0_n)^2 = O_{\mathbb{P}}(p^2 n^{-1}).$$

Preuve. On montre que $\mathbb{E}[\mathrm{tr}(\mathbf{K}_n^0)^2] = O(p^2 n^{-1})$. Pour ce faire, on écrit la matrice $\mathbf{K}_n^0$ sous la forme

$$\mathbf{K}_n^0 = \mathbf{I}_p - \frac{1}{n\theta^2}\sum_{i=1}^{k}\mathbf{w}_i^0\mathbf{w}_i^{0t}\frac{1}{n(1-\theta)^2}\sum_{j=k+1}^{n}\mathbf{w}_j^0\mathbf{w}_j^{0t}.$$

Alors

$$\begin{aligned}\mathbb{E}[\mathrm{tr}(\mathbf{K}_n^0)^2] &= \frac{1}{n^2}\sum_{r,s=1}^{p}\mathbb{E}\left[\left(\frac{1}{\theta^2}\sum_{i=1}^{k}w_{i,r}^0 w_{i,s}^0 + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}w_{j,r}^0 w_{j,s}^0\right)^2\right] \\ &\quad -\frac{2}{n}\sum_{r=1}^{p}\left(\frac{1}{\theta^2}\sum_{i=1}^{k}\mathbb{E}[w_{i,r}^0 w_{i,r}^0] + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathbb{E}[w_{j,r}^0 w_{j,r}^0]\right) + p.\end{aligned} \tag{4.92}$$

Pour le premier terme du membre droit de la relation (4.92), en utilisant le fait que pour tout $i = 1, \ldots, n$, les $\mathbf{w}_i^0$ sont indépendants, on a

$$\begin{aligned}&\frac{1}{n^2}\sum_{r,s=1}^{p}\mathbb{E}\left[\left(\frac{1}{\theta^2}\sum_{i=1}^{k}w_{i,r}^0 w_{is}^0 + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}w_{j,r}^0 w_{j,s}^0\right)^2\right] \\ &= \frac{1}{n^2}\sum_{r,s=1}^{p}\left[\frac{1}{\theta^4}\sum_{i=1}^{k}\mathbb{E}[w_{i,r}^0 w_{i,r}^0 w_{i,s}^0 w_{i,s}^0] + \frac{1}{(1-\theta)^4}\sum_{j=k+1}^{n}\mathbb{E}[w_{j,r}^0 w_{j,r}^0 w_{j,s}^0 w_{j,s}^0]\right] \\ &\quad + \frac{1}{n^2}\sum_{r,s=1}^{p}\left[\frac{1}{\theta^2}\sum_{i=1}^{k}\mathbb{E}[w_{i,r}^0 w_{i,s}^0] + \frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathbb{E}[w_{j,r}^0 w_{j,s}^0]\right]^2 \\ &= \frac{1}{n}\sum_{r,s=1}^{p}\alpha^{rrss} + \sum_{r,s=1}^{p}(\alpha^{rs})^2 \\ &= O(p^2 n^{-1}).\end{aligned} \tag{4.93}$$

La dernière équation est due à la supposition (A7) et au fait que $\alpha^{rs} = 0$, pour $r \neq s$.

Pour le deuxième terme du membre droit de la relation (4.92), par le fait que $\alpha^{rr} = 1$, on a

que

$$\begin{aligned}
&\frac{1}{n}\sum_{r=1}^{p}\left(\frac{1}{\theta^2}\sum_{i=1}^{k}\mathbb{E}[w_{i,r}^0 w_{i,r}^0]+\frac{1}{(1-\theta)^2}\sum_{j=k+1}^{n}\mathbb{E}[w_{j,r}^0 w_{j,r}^0]\right)\\
&=\sum_{r=1}^{p}\alpha^{rr}\\
&=p. \qquad (4.94)
\end{aligned}$$

En utilisant les relations (4.92), (4.93) et (4.94), on obtient que

$$\mathbb{E}[\mathrm{tr}(\mathbf{K}_n^2)] = O(p^2 n^{-1}).$$

Le Lemme résulte, par l'inégalité de Markov. ■

Bibliographie

[Antoch et al. (2004)] Antoch, J., Gregoire, G., Jaruskova, D. Detection of structural changes in generalized linear models. *Statistics and Probability Letters*, **69**, 315-332, 2004.

[Antoch et al. (2007)] Antoch, J., Grégoire, G., Hušková, M. Tests for continuity of regression functions. *Journal of Statistical Planning and Inference,* **137**(3), 753-777, 2007.

[Asgharian and Wolfson (2001)] Asgharian, M. Wolfson, D. B. Covariates in Multipath Change-Point Problems : Modelling and Consistency of the MLE. *The Canadian Journal of Statistics*, **29**, 515-528, 2001.

[Bacon and Watts (1971)] Watts, Donald G., Bacon, David W. Estimating the transition between two intersecting straight lines. *Biometrika*, **58**,525-534, 1971.

[Bai (1999)] Bai, J. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, **91**, 299-323, 1999.

[Bai and Perron (1998)] Bai, J., Perron, P. Estimating and testing linear models with multiple structural changes. *Econometrica*, **1**, 47-48, 1998.

[Bardet and Kengne (2014)] Bardet, J.M, Kengne, W. Monitoring procedure for parameter change in causal time series. *Journal of Multivariate Analysis*, **251**, 204-221, 2014.

[Bhattacharya (1991)] Bhattacharya, P. K. Weak convergence of the log likelihood process in two phase linear regression problem . *Probability Statistics and Design of Experiments*, **17**,145-156, 1991.

[Bhattacharya (1994)] Bhattacharya, P. K. Some aspects of change-point analysis. *Change-point problems*, **17**,28-56, 1994.

[Boldea and Hall (2013)] Boldea, O., Hall, A.R. Estimation and inference in unstableau nonlinear least squares models. *Journal of Econometrics*, **172**(1), 158-167, 2013.

[Butucea and Gayraud (2015)] Butucea, C., Gayraud, G. Sharp detection of smooth signals in a high-dimensional sparse matrix with indirect observations. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, à paraître, 2015.

[Butucea et al. (2015)] Butucea, C., Ingster, Yu. I., Suslina, I. Sharp Variable Selection of a Sparse Submatrix in a High-Dimensional Noisy Matrix. *ESAIM PS*, **19**, 115-137, 2015.

[Celeux et al. (2006a)] Celeux, G., Marin, J.M, Robert, C.P. Iterated importance sampling in missing data problems. *Computational Statistics & Data Analysis*, **50**(12), 3386–3404, 2006.

[Celeux et al. (2006b)] Celeux, G., Forbes, F., Robert, C. P. Deviance information criteria for missing data models. *Bayesian Analysis*, **1**(4), 651–673, 2006.

[Chen (1993)] Chen, S. On the accuracy of empirical likelihood confidence region for linear regression model. *Annals of the Institute of Statistical Mathematics*, **45**, 621-637, 1993.

[Chen (1994)] Chen, S. Empirical likelihood confidence interval for linear regression coefficients. *Journal of Multivariate Analysis*, **49**(1), 24–40, 1994.

[Chen and Cui (2003)] Chen, S., Cui, H. An extended empirical likelihood for generalized linear models, *Statistica Sinicia* , **13**, 69-81, 2003.

[Chen et al. (2009)] Chen, S. X., Peng, L., Qin, Y. L. Effect of data dimension on empirical likelihood. *Biometrika*, **96**, 1-12, 2009.

[Cheng (1994)] Cheng, P. E. Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association* , **89**, 81-87, 1994.

[Chow and Teicher (1997)] Chow, Y.S., Teicher, H. Probability theory : independence, interchangeability, martingales. *Springer*, *New York*, 1997.

[Ciuperca (2014)] Ciuperca, G. Model selection by LASSO methods in a change-point model. *Statistical Papers*, **55**, 349-374, 2014.

[Ciuperca (2013a)] Ciuperca G. Empirical likelihood for nonlinear model with missing responses. *Journal of Statistical Computation and Simulation*, **83**(4), 737-756, 2013.

[Ciuperca (2013b)] Ciuperca, G. Quantile regression in high-dimension with breaking. *Journal of Statistical Theory and Applications*, **12**, 288-305, 2013.

[Ciuperca (2011a)] Ciuperca, G. A general criterion to determinate the number of change-points. *Statistics and Probability Letters*, **81**(8), 1267-1275, 2011.

[Ciuperca (2011b)] Ciuperca, G. Penalized least absolute deviations estimation for nonlinear model with change-points. *Statistical Papers*, **52**(2), 371-390, 2011.

[Ciuperca and Salloum (2015a)] Ciuperca, G., Salloum, Z. Empirical likelihood test in a posteriori change-point nonlinear model. *Metrika*, **78**(8), 919-952, 2015.

[Ciuperca and Salloum (2015b)] Ciuperca, G., Salloum, Z. Empirical likelihood test for high-dimensional two-sample model. *arXiv :1504.05690*, 2015 .

[Csörgö and Horváth (1997)] Csörgö M., Horváth L. Limit Theorems in Change-point Analysis. *John Wiley and Sons*, 1997.

[Dabo-Niang and Yao (2013)] Dabo-Niang, S., Yao, A.F. Kernel spatial density estimation in infinite dimension space. *Metrika*, **76**(1), 19-52, 2013.

[Dabo-Niang et al. (2015)] Dabo-Niang, S., Kaid, Z., Laksaci, A. Asymptotic properties of the kernel estimate of spatial conditional mode when the regressor is functional. *AStA Advances in Statistical Analysis*, **99**(2), 131-160, 2015.

[Dachian and Yang (2015)] Dachian, S., Yang, L. On a poissonian change-point model with variable jump size. *Statistical inference for Stochastic Processes*, **18**(2), 127-150, 2015.

[Deville and Sarndal (1992)] Deville J. C., Sarndal C. E. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**, 376-382, 1992.

[Dharmadhikari and Jogdeo (1969)] Dharmadhikari, SW., Jogdeo, K. Bounds on moments of certain random variables. *The Annals of statistic*, **40**, 1506-1509, 1969.

[DiCiccio et al. (1991)] DiCiccio, T., Hall, P., Romano, J. Empirical likelihood is bartlett-correctable. *The Annals of Statistics*, **19**(2), 1053-1061, 1991.

[Dicker et al. (2013)] Dicker, L., Huang, B., Lin, X. Variable selection and estimation with the seamless-L_0 penalty. *Statistica Sinica*, **23**, 929-962, 2013.

[Facer and Muller (2003)] Facer, M. R. Muller H. G. Nonparametric estimation of the location of a maximum in a response surface. *Journal of Multivariate Analysis*, **87**(1), 191-217, 2003.

[Fan and Peng (2004)] Fan, J., Peng, H. Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, **32**(3), 928-961, 2004.

[Feder (1975)] Feder, Paul I. The log likelihood ratio in segmented regression. *The Annals of Statistics*, **3**, 84-97, 1975.

[Gallant and Fuller (1973)] Gallant, A. R., Fuller, W. A. Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association*, **68**,144-147, 1973.

[Guan (2004)] Guan, Z. A semiparametric changepoint model. *Biometrika*, **91**(3), 849-862, 2004.

[Guo et al. (2013)] Guo, H., Zou, C., Wang, Z., Chen, B. Empirical likelihood for high-dimensional linear regression models. *Metrika*, **77**(7), 921-945, 2013.

[Hall and Sen (1999)] Hall A., Sen A. Structural stability testing in models estimated by generalized method of moments. *Journal of Business Economic Statistics*, **17**(3), 335-348, 1999.

[Hartley and Rao (1968)] Hartley H. O., Rao J. N. K. A new estimation theory for sample surveys. *Biometrika*, **55**, 547-557, 1968.

[Healy and Westmacott (1956)] Healy, M. J. R., Westmacott, M. Missing values in experiments analysed on automatic computers. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **5**, 203-206, 1956.

[Hinkley (1972)] Hinkley, D. V. Time-Ordered Classification. *Biometrika*, **59**,509-523, 1972.

[Horváth et al. (2004)] Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J. Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, **126**(1), 225-251, 2004.

[Hušková and Kirch (2012)] Hušková, M., Kirch, C. Bootstrapping sequential change-point tests for linear regression. *Metrika* **75**(5), 673-708, 2012.

[Hjort et al. (2009)] Hjort, N.L., Mckeague, I.W., Van Keilegom, I. Extending the scope of empirical likelihood. *The Annals of Statistics*, **37**, 1079-1111, 2009.

[Huang et al.(2012)] Huang, J., Breheny, P., Ma, S. A selective review of group selection in high-dimensional models. *Statistical Science*, **27**(4), 481-499, 2012.

[Huskova et al. (2007)] Hušková, M., Prášková, Z., Steinebach, J. On the detection of changes in autoregressive time series. I. Asymptotics. *Journal of Statistical Planning Inference*, **137**(4), 1243-1259, 2007.

[Horvitz and Thompson (1952)] Horvitz, DG., Thompson, DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat assoc*, **47**, 663-685, 1952.

[Jing (1995)] Jing, B. Two-sample empirical likelihood method. *Statistics and Probability Letters*, **24**, 315-319, 1995.

[Kim and Siegmund (1989)] Kim, H.j., Siegmund, D. The likelihood ratio test for a change-point in simple linear regression model. *Biometrika*, **76**, 409-423, 1989.

[Kühn (2001)] Kühn, C. An estimator of the number of change points based on a weak invariance principle. *Statistics and Probability Letters*, **51**(2), 189-196, 2001.

[Kolaczyk (1994)] Kolaczyk, E. D. Empirical likelihood for generalized linear models. *Statistica Sinicia*, **4**, 199-218, 1994.

[Lai and Xing (2010)] Lai, T. L., Xing, H. Sequential change-point detection when the pre- and post-change parameters are unknown. *Sequential Analysis,* **29**(2), 162–175, 2010.

[Lavielle (1999)] Lavielle, M. Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, **83**, 79-102, 1999.

[Lee et al. (2015)] Lee S., Seo M.H., Shin Y. The LASSO for high-dimensional regression with a possible change-point. *Journal of the Royal Statistical Society : Series B*, DOI : 10.1111/rssb.12108, 2015.

[Liang and Zeger (1986)] Liang, K. Y., Zeger, S. Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22, 1986.

[Little and Rubin (1983)] Little, R.J.A, Rubin, D.B. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *American Statistician*, **37**, 218-220, 1983.

[Little and Rubin (1987)] Little, R.J.A, Rubin, D.B. Statistical analysis with missing data. *John Wiley and sons, Inc., New York*, 1987.

[Little and Rubin (2002)] Little, R.J.A, Rubin, D.B. Statistical analysis with missing data, 2nd edn. *John Wiley and sons, Inc., New York*, 2002.

[Liu et al. (2008)] Liu, Y., Zou, C., Zhang, R. Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics-Theory and Methods*, **37**, 2551-2563, 2008.

[Liu et al(2013)] Liu, Y., Zou, C., Wang, Z. Calibration of the empirical likelihood for high-dimensional data. *Annals of the Institute of Statistical Mathematics*, **65**, 529-550, 2013.

[Lombard (1987)] Lombard, F. Rank tests for change point problems. *Biometrika*, **74**, 615-624, 1987.

[Lung-Yut-Fong et al.(2013)] Lung-Yut-Fong, A., Lévy-Leduc, C., Cappé, O. Distributed detection/localization of change-points in high-dimensional network traffic data. *Statistics and Computing*, **22**, 485-496, 2012.

[Mei (2006)] Mei, Y. Sequential change-point detection when unknown parameters are present in the pre-change distribution. *The Annals of Statistics* **34**(1), 92-122, 2006.

[Mia et al. (1988)] Mia, B., Zhao, L. Detection of change points using rank methods. *Communications in Statistics. Theory and Methods*, **17**, pp. 3207-3217, 1988.

[Muller (2009)] Muller, U.U. Estimating linear functionals in nonlinear regression with responses missing at random. *Annal statistic*, **37**(5A), 2245-2277, 2009.

[Neumeyer and al. (2009)] Neumeyer, N., Van Keilegom, I. Change-point tests for the error distribution in non-parametric regression. *Scandinavian Journal of Statistics* **36**(3), 518-541, 2009.

[Newey and Smith (2004)] Newey W. K., Smith R. J. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* **72**(1), 219-255, 2004.

[Ning and al. (2012)] Ning, W., Pailden, J., Gupta, A. Empirical likelihood ratio test for the epidemic change model. *Journal of Data science*, **10**, 107-127, 2012.

[Nosek (2010)] Nosek, K. Schwarz information criterion based tests for a change-point in regression models. *Statistical Papers*, **51**(4), 915-929, 2010.

[Owen (1988)] Owen, A. B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249, 1988.

[Owen (1990)] Owen, A. B. Empirical likelihood ratio confidence regions. *Annal statistic*, **18**, 90-120, 1990.

[Owen (1991)] Owen, A. B. Empirical likelihood for linear model. *Annal statistic*, **19**, 1725-1747, 1991.

[Owen (2001)] Owen, A. B. Empirical likelihood. *Chapman and Hall/CRC, Boca Raton*, 2001.

[Portnoy (1984)] Portnoy, S. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large.I.Consistency. *The Annals of Statistics*, **12**, 1298-1309, 1984.

[Portnoy (1985)] Portnoy, S. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II. Normal approximation. *The Annals of Statistics*, **13**, 1403-1417, 1985.

[Qin and al. (2009)] Qin, Y., Li, L., Lei, Q. Empirical likelihood for linear regression model with missing responses. *Statistics and Probability Letters*, **79**, 1391-1396, 2009.

[Qin and Lawless (1994)] Qin,J., Lawless, J. Empirical likelihood and general estimating equations. *Annals of Statistics*, **22**(4), 300-325, 1994.

[Qu (2008)] Qu, Z. Testing for structural change in regression quantiles. *Journal of Econometrics*, **146**, 170-184, 2008.

[Ramanayake and Gupta (2003)] Ramanayake, A., Gupta A.K. Tests for an epidemic change in a sequence of exponentially distributed random variables. *Biometrical journal*, **45**, 496-958, 2003.

[Ramanayake (2004)] Ramanayake, A. Tests for a change point in the shape parameter of gamma random variables. *Communications in Statistics -Theory and Methods*, **33**(4), 821-833, 2004.

[Rosenbaum and Rubin (1983)] Rosenbaum, P., Rubin, D. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**(1), 41-55, 1983.

[Samarov et al. (2005)] Samarov, A., Spokoiny, V., Vial, C. Component Identification and Estimation in Nonlinear High-Dimensional Regression Models by Structural Adaptation. *Journal of the American Statistical Association*, **100**(470), 429–445, 2005.

[Saranadasa (1993)] Saranadasa, H. Asymptotic expansion of the misclassification probabilités of D- and A-criteria for discrimination from two high-dimensional populations using the theory of large dimensional random matrices. *Journal of Multivariate Analysis* , **46**, 154-174, 1993.

[Shi (2000)] Shi, J. Empirical likelihood for partialy linear models. *Journal of Multivariate Analysis*, **72**, 132-148, 2000.

[Silverstein (1985)] Silverstein, J.W The smallest eigenvalue of a large dimensional Wishart matrix. *The Annals of Probability* , **13**, 1364-1368, 1985.

[Salloum(2015)] Salloum, Z. Empirical likelihood confidence regions for the parameters of a two phases nonlinear model with and without missing response data. *arXiv :1406.3960*, 2015.

[Seber and Wild (2003)] Seber G., Wild C. Nonlinear regression. *Wiley series in probability and mathematical Statistics, Wiley, Hoboken, NJ, 2003*.

[Serbinowska (1996)] Serbinowska, M. Consistency of an estimator of the number of changes in binomial observations. *Statistics and Probability Letters*, **29**(4), 337-344, 1996.

[Sprent (1961)] Hinkley, D. V. Some Hypotheses Concerning Two Phase Regression Lines. *Biometrics*, **17**,634-645, 1961.

[Sun et al. (2009)] Sun, Z., Wang, Q., Dai, P. model checking for partially linear models with missing responses at random. *Journal of Multivariate Analysis*, **100**, 636-651, 2009.

[Tibshirani (1996)] Tibshirani, R. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society*, Ser. B, **58**, 267-288,1996.

[Van der Vaart and Wellner (1996)] Van der Vaart A., Wellner J. Weak Convergence and Empirical processes with applications to statistics. *Springer Series in Statistics. Springer, New York*, 1996.

[Wei et al. (2013)] Wei, Yu., Cuizhen, Ni., Wangli, Xu. An empirical likelihood inference for the coefficient difference of a two-sample linear model with missing response data. *Metrika*, **81**(8), 1267-1275, 2013.

[Wang and Rao (2002)] Wang, QH., Rao, JNK. Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statistics*, **30**, 896-924, 2002.

[Wang and Chen (2009)] Wang, QH., Rao, JNK. Empirical likelihood for estimating equation with missing values. *The Annals of Statistics*, **37**, 490-517, 2009.

[Wu (2008)] Wu, Y. Simultaneous change point analysis and variable selection in a regression problem. *Journal of Multivariate Analysis*, **99**, 2154-2171, 2008.

[Wu and Liu (2009)] Wu, Y., Liu, Y. Variable selection in quantile regression. *Statistica Sinica*, **19**, 801-817, 2009.

[Xue (2009)] Xue, L. Empirical likelihood for linear models with missing responses. *Journal of Multivariate Analysis*. **100**, 1353-1366, 2009.

[Xu et al. (2012)] Xu, WL., Guo, X., Zhu, LX. Goodness-of-fitting for partial linear model with missing response at random. *Journal of Nonparametric Statistics*, **24**, 103-118, 2012.

[Yao (1993)] Yao, Q.W. Tests for change-points with epidemic alternatives. *Biometrika*, **80**, 179-191, 1993.

[Yao and Au (1989)] Yao, Y.C., Au, S.T. Least-squares estimation of a step function. *Sankhya (Series A)*, **51**(3), 370-381, 1989.

[Yates (1993)] Yates, F. The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture*, **1**, 129-142, 1993.

[Yin et al. (1983)] Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R. Limiting behavior of the eigenvalues of a multivariate F matrix. *Journal mutivariate analysis*, **13**, 508-516.

[Yin et al. (1988)] Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R. On the limit of the large dimensional sample covariance matrix. *Probability theory related fields*, **78**, 509-521.

[Yu et al. (2014)] Yu, W., Niu, C., Xu, W. An empirical likelihood inference for the coefficient difference of a two-sample linear model with missing response data. *Metrika*, **77**, 675-693 , 2014.

[Zou et al. (2007)] Zou C., Liu Y., Qin P., Wang Z. Empirical likelihood ratio test for a change point. *Statistics and Probability Letters*, **77**, 374-382, 2007.

[Zi et al. (2010)] Zi, X., Zou, C., Liu, Y. Two-sample empirical likelihood method for difference between coefficients in linear regression model. *Statistical Papers*, **53**(1), 83-93, 2012.

[Zou and Yuan (2008)] Zou, H., Yuan, M. Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, **36**, 1108-1126, 2008.