

STATISTIQUE ET ANALYSE DES DONNÉES

FAOUZI LYAZRHI

**Détection de ruptures dans un modèle linéaire gaussien
: optimalité de procédures de décision multiple basées
sur le rapport de vraisemblance**

Statistique et analyse des données, tome 16, n° 3 (1991), p. 107-125

http://www.numdam.org/item?id=SAD_1991__16_3_107_0

© Association pour la statistique et ses utilisations, 1991, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

**DETECTION DE RUPTURES DANS UN MODELE
LINEAIRE GAUSSIEN : OPTIMALITE DE PROCEDURES
DE DECISION MULTIPLE BASEES SUR LE RAPPORT
DE VRAISEMBLANCE**

Faouzi LYAZRHI

Laboratoire de Statistique et Probabilités

Université Paul Sabatier

118, route de Narbonne, 31062 Toulouse cedex

Résumé.

Ce papier présente d'abord le problème de ruptures de modèles (dans le cas de modèles linéaires gaussiens) et rappelle différentes façons dont il a été abordé dans la littérature. Une procédure de décision multiple dans la détection d'une ou plusieurs ruptures dans un modèle linéaire gaussien est ensuite proposée, procédure optimale au sens de Bayes. On montre que cette procédure est équivalente à la procédure de décision multiple déduite du test de rapport de vraisemblance largement utilisé dans la littérature. Cela confère à ces méthodes usuelles une nouvelle propriété d'optimalité. Les cas de la régression linéaire simple et de l'échantillon sont traités en détail à titre d'illustration.

Mots-clés : *Rupture, modèle linéaire gaussien, vraisemblance, régression linéaire simple, procédure bayésienne.*

Abstract.

This paper presents the change-point problem (in the framework of gaussian linear models) as well as the different approaches that can be found in the literature. A procedure of multiple decision for detecting one or several change-point (s) in the linear gaussian model is proposed. This procedure presents a Bayes optimal character. The relation with other procedures derived from the likelihood ratio test are investigated which allows us to show a new optimal property of these usual techniques. The cases of a sample and of the simple linear regression are worked out in more details.

Key-words : *Change-point, linear gaussian model, likelihood, simple linear regression, Bayes procedure.*

INTRODUCTION.

Dans ce papier, nous nous intéressons à l'optimalité d'une procédure de Bayes basée sur le rapport de vraisemblance dans la détection d'un ou plusieurs point(s) de rupture dans un modèle linéaire gaussien. En fait, des techniques de détection ont été proposées dans la littérature, mais sans être discutées du point de vue de leurs propriétés, ou en étant discutées dans d'autres optiques.

Nous montrons l'optimalité de certaines d'entre elles pour une large classe de lois a priori sur les paramètres. Par ailleurs, nous élargissons le cadre des travaux les plus usuels en présentant le problème non comme un choix entre les hypothèses d'absence ou de présence de rupture, mais comme un choix entre l'hypothèse d'absence de rupture et l'une des hypothèses de présence de rupture en un ou des point(s) donné(s).

Notre démarche s'inspire de celle adoptée par Butler (1983) dans le cadre de la sélection de variables, ou Caussinus et Vaillant (1985) avec un point de vue plus orienté vers la détection des valeurs aberrantes. Le problème de rupture de modèles peut être présenté de la façon suivante : on observe dans le temps une certaine variable aléatoire et on suppose que les t_1 premières observations sont régies par un modèle M_1 , les t_2-t_1 observations suivantes par un modèle M_2 , etc, les points t_1, t_2, \dots s'appellent points de rupture.

Schématiquement, il y a quatre façons d'aborder le problème de rupture de modèles :

- (i) On suppose connus a priori le nombre et la position des points t_1, t_2, \dots ; dans ce cas, on peut faire de l'ajustement de modèles, tester la valeur des paramètres du modèle, etc.
- (ii) On suppose connu a priori le nombre des points t_1, t_2, \dots mais pas leurs positions ; dans ce cas, on peut chercher à estimer ces positions ou tester des valeurs données.
- (iii) On suppose qu'a priori chaque instant d'observation est susceptible d'être un point de rupture.
- (iv) On suppose qu'a priori chaque observation peut appartenir à n'importe quel sous modèle d'une liste donnée M_1, M_2 etc.

Dans ce dernier cas, l'ordre dans lequel les diverses observations interviennent n'a pas beaucoup d'importance, ce qui dirige vers l'utilisation de méthodes telles que la classification automatique (Cox and Spjøtvoll, 1982). Il existe une littérature abondante traitant du sujet avec des points de vue proches de (ii) et (iii) : Shaban (1980), Zacks (1982) donnent une bibliographie détaillée des premiers travaux.

Beaucoup d'auteurs intéressés par des questions concrètes ont traité ce problème dans le cas de l'échantillon gaussien (ou éventuellement pour des distributions exponentielles), par exemple Raftery and Akman (1986), Worsley (1986). Dans ce cas, on suppose que la rupture est brutale et la détection de cette rupture est d'autant plus facile que son amplitude est grande. Une fois l'hypothèse nulle écartée, l'emplacement exact du point de rupture est important dans la pratique, par exemple pour l'interprétation des causes de cette rupture. Des auteurs ont traité ce problème dans le cadre de la régression linéaire simple : Kim et Siegmund (1989), Knowles and Siegmund (1989), Freeman (1986), McCabe and Harrison (1980). En particulier ces derniers ont considéré le cas d'une rupture moins brutale.

En général, la variable explicative est le temps, et l'on suppose que la rupture est moins brutale. Pour d'autres applications de la régression linéaire dans lesquelles la variable explicative n'est pas forcément le temps, voir Stasinopoulos and Nicholls 1989, Maronna and Yohai 1978. Le plus souvent, les procédures considérées supposent que le point de rupture correspond à l'un des instants d'observation et tentent de répondre à la question suivante : y a-t-il une rupture dans le modèle considéré ou non ? Ces procédures sont en général basées sur le test du rapport de vraisemblance ; si l'hypothèse nulle est rejetée, certains auteurs se sont intéressés à l'estimation ponctuelle du point de rupture : calcul de la distribution d'un estimateur dans le cas gaussien (Broemeling and Choy, 1981), (Holbert and Broemeling, 1977), calcul d'un intervalle de confiance (Kim and Siegmund (1989), (Ferreira, 1975) dans le cas gaussien ou Worsley (1986) dans le cas de lois exponentielles).

D'autres auteurs se sont consacrés au calcul des puissances des différents tests proposés dans la littérature, comme James, James and Siegmund (1987). Le cas où le point de rupture ne ferait pas partie des instants d'observation a fait aussi l'objet de quelques travaux : Hinkley (1969, 1971), Smith and Cook (1980). Cette formulation a permis récemment une approche géométrique pour calculer la valeur critique d'une procédure, l'un des problèmes les plus difficiles liés aux questions de ruptures de modèles (Knowles and Siegmund, 1989). Très peu d'articles se placent dans le cas de la régression linéaire multiple mais il y a des exceptions notables : Brown, Durbin and Evans (1975), Broemeling and Choy (1981), Smith (1977), Booth and Smith (1982). En ce qui concerne les propriétés d'optimalité de ces procédures, notons que Farley and Hinich (1970) (dans le cadre de la régression linéaire simple) et Henderson (1986) (dans le cadre d'un échantillon de variables gaussiennes) ont montré une optimalité du test du rapport de vraisemblance.

Deshayes et Picard (1982) se sont intéressés à la question dans ses propriétés asymptotiques, un point de vue tout à fait différent de celui qui nous préoccupe ici.

Enfin, signalons que la plupart des auteurs se sont intéressés à la détection d'au plus une rupture dans le cas gaussien et ce au niveau de la moyenne. Quelques auteurs ont étudié le cas d'une rupture au niveau de la variance (Brown, Durbin and Evans, 1975; Quandt, 1972).

En ce qui nous concerne, nous essayons de répondre à la question qui nous semble la plus pertinente. "Y a-t-il une ou plusieurs rupture(s) dans le modèle et où ?", puisque la question pratique est de savoir où se trouvent les ruptures une fois l'hypothèse écartée. Ce faisant, nous nous plaçons donc dans le cadre d'hypothèses multiples. Pour cette situation, nous proposons une procédure de décision multiple dont nous montrons l'optimalité au sens de Bayes, pour des lois a priori sur les paramètres plus générales que celles proposées antérieurement (Zacks, 1982 ; Booth and Smith, 1982 ; Smith, 1975).

Dans la première partie, nous donnons la procédure générale pour détecter un ou plusieurs point(s) de rupture dans un modèle linéaire quelconque. Dans la deuxième partie, nous montrons son équivalence avec la procédure de décision multiple déduite d'un test du rapport de vraisemblance. Dans la troisième partie, nous illustrons ces résultats avec le cas de la régression linéaire simple et de l'échantillon de variables aléatoires gaussiennes. Dans la quatrième partie, nous proposons une méthode pour calculer les valeurs critiques à l'aide de simulations et, pour certains cas, nous dressons une table donnant les valeurs critiques en fonction des niveaux choisis. Enfin, dans la cinquième partie, nous simulons des modèles avec rupture dans le cas de la régression linéaire simple, afin de quantifier les performances des techniques étudiées.

1. NOTATIONS ET FORMULATION DU PROBLEME.

Nous nous plaçons dans le cas de la détection de plusieurs ruptures dans un modèle linéaire gaussien. Ce dernier étant invariant par un groupe de transformations, l'idée est de chercher une procédure optimale au sens de Bayes parmi celles qui sont invariantes par ces transformations.

1.1. Notations.

On note Y une variable aléatoire à valeur dans \mathbf{R}^n de loi de probabilité $N_n(\mu, \sigma^2 I_n)$ avec $\mu \in Q$ (un sous-espace de \mathbf{R}^n de dimension q) et $\sigma > 0$. On note Π_{Q^\perp} , le projecteur orthogonal sur Q^\perp pour le produit scalaire canonique (Q^\perp désigne l'orthogonal de Q). On se donne J sous-espaces vectoriels Q_j ($Q_j \subset Q^\perp$, $\dim(Q_j) = k_j$). Enfin on note Π_j le projecteur orthogonal sur Q_j .

1.2. Hypothèses statistiques.

On considère les (J+1) hypothèses suivantes :

$$H_0 : \mathfrak{L}(Y) = N_n(\mu, \sigma^2 I_n), \mu \in Q$$

$$H_j : \mathfrak{L}(Y) = N_n(v, \sigma^2 I_n), v \in Q \oplus Q_j, j = 1, \dots, J$$

On pose $\delta = H_{Q^\perp}(v)$ et $m = n - q$.

Le groupe des transformations $\{y \rightarrow ay + b, a > 0, b \in Q\}$ laisse invariant le modèle et un invariant maximal est le vecteur des résidus normés $T = \Pi_{Q^\perp}(Y) / \|\Pi_{Q^\perp}(Y)\|$.

Sous H_0 , la loi de probabilité de T est la loi uniforme sur la sphère unité de Q^\perp , loi notée U_{Q^\perp} . Sous H_j , la densité de la loi de probabilité de T par rapport à U_{Q^\perp} est (cf.

Caussinus and Vaillant, 1985) :

$$t \rightarrow g_j(t, \delta, \sigma) = \frac{1}{2^{(m/2)-1} \Gamma(\frac{m}{2})} e^{-\frac{\|\delta\|^2}{2\sigma^2}} h_m\left(\frac{\langle \Pi_j(t), \delta \rangle}{\sigma}\right)$$

avec $h_m(u) = \int_0^\infty e^{uv} e^{-\frac{v^2}{2}} v^{m-1} dv$.

Exemple.

Considérons le cas d'un échantillon $Y = (y_1, \dots, y_n)'$ avec la possibilité d'une rupture au niveau de la moyenne ; on a, en désignant par ε le vecteur $(\varepsilon_1, \dots, \varepsilon_n)'$:

$$H_0 : Y_i = \alpha + \varepsilon_i, \quad 1 \leq i \leq n, \quad \mathfrak{L}(\varepsilon) = N(0, \sigma^2 I_n)$$

$$H_j : \begin{cases} Y_i = \alpha + \varepsilon_i, & 1 \leq i \leq j \\ Y_i = \alpha + \lambda + \varepsilon_i & j < i \leq n \end{cases} \quad \mathfrak{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$$

Le sous-espace Q est ici engendré par $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^n$ et Q_j est engendré par $\Pi_{Q^\perp}(\mathbf{1}_j)$ avec $\mathbf{1}_j = (0, \dots, 0, 1, \dots, 1)'$ où le premier 1 occupe le rang (j+1).

1.3. Lois a priori.

1°) On donne à H_j une probabilité a priori $p_j > 0$ avec $\sum_{j=0}^J p_j = 1$.

2°) On suppose que, sous H_j , $j \neq 0$, la loi a priori de δ/σ est sphérique (il est facile de voir au niveau de la densité de T que δ et σ n'interviennent que par leur rapport).

Remarques

- Pour l'exemple ci-dessus le choix fait au 1°) suppose que chaque instant d'observation est susceptible de coïncider avec le point de rupture avec la probabilité p_j ,

$p_0 = 1 - \sum_{j=1}^{n-1} p_j$ est la probabilité qu'il n'y ait pas rupture.

- La loi sphérique dans le 2°) est une généralisation de la loi normale ou des lois non informatives qu'on prend le plus souvent comme lois a priori dans la littérature (Zacks, 1982 ; Booth and Smith, 1982). En effet, cela revient à considérer une loi non informative pur toute direction de Q_j , et n'importe quelle loi de probabilité pour l'amplitude de la rupture. Par ailleurs nos hypothèses supposent que l'endroit où a lieu la rupture et l'amplitude de cette rupture sont indépendants.

Notons que, pour $\dim(Q_j) = 1$, c'est-à-dire δ à valeurs réelles, l'hypothèse de sphéricité correspond à une hypothèse de symétrie de la distribution de δ/σ autour de l'origine.

1.4. Fonction de perte.

On suppose la fonction de perte suivante pour la décision de H_i alors que H_j est vraie :

$$\mathcal{L}(i,j,\delta) = \begin{cases} 0 & \text{si } i = j \\ \mathcal{L} & \text{si } j \neq i, j = 0 \\ \mathcal{L}(\|\delta\|) & \text{si } j \neq i, j \neq 0 \end{cases}$$

Remarque.

Il paraît naturel que la fonction de perte dépende de l'amplitude de la rupture. Mais comme l'ordre dans lequel on observe les variables (souvent le temps) est en général

important, notre fonction de perte devrait aussi tenir compte de la différence entre l'endroit jugé comme point de rupture et le vrai point de rupture. Nous pensons que cette question méritera d'être développée ultérieurement.

1.5. Procédure optimale au sens de Bayes.

Dans tout ce qui suit on va s'intéresser à la détection d'un nombre donné de ruptures, de même nature ; on suppose que les hypothèses $(H_j)_{j \neq 0}$ ont même dimension $k_j = k$ (voir Remarque 1.6). D'autre part on suppose :

- les p_j sont égaux pour $j \neq 0$.
- sous H_j , la loi de δ/σ est identique pour tout $j \neq 0$.

Dans ce cas, d'après Caussinus and Vaillant (1985) on a :

Lemme.

Une procédure optimale au sens de Bayes est la suivante :

$$\begin{cases} \text{décider } H_0 & \text{si } \max_{i \neq 0} \|\Pi_i(T)\| \leq C \\ \text{décider } H_j & \text{si } \|\Pi_j(T)\| > \max[C, \max_{i \neq j} \|\Pi_i(T)\|], j \neq 0 \end{cases}$$

Nous verrons dans la quatrième partie une méthode pour calculer C à l'aide de simulations.

Remarque.

Dans le cas d'une fonction de perte de la forme $\ell(i,j,\delta) = 0$ si $i = j$ et $\ell(i,j,\delta) = 1$ si $i \neq j$, maximiser $\|\Pi_j(T)\|$ revient à maximiser la probabilité à posteriori $P(H_j/T)$. Ce cas a été considéré par certains auteurs en particulier par Booth and Smith (1982) dans le cadre d'hypothèses multiples.

1.6. Expression de $\|\Pi_i(T)\|$.

H_j est en général défini à partir d'un sous-espace de \mathbb{R}^n engendré par des vecteurs donnés e_{j1}, \dots, e_{jk} . Alors Q_j est engendré par $\Pi_{Q^\perp}(e_{j1}), \dots, \Pi_{Q^\perp}(e_{jk})$, qu'on suppose linéairement indépendants. T est le vecteur des résidus normés qu'on note $\frac{R}{\|R\|}$ avec $R = \Pi_{Q^\perp}(Y)$. Pour alléger l'écriture on note P la matrice de Π_{Q^\perp} ; rappelons que $\sigma^2 P$ est la matrice des variances-covariances des résidus.

Q_j est engendré par les vecteurs colonnes de la matrice $P\theta_{jk} = P[e_{j1} \dots e_{jk}]$,

Π_j s'écrit matriciellement $P\theta_{jk}\Delta_j^{-1}\theta_{jk}'P'$, avec $\Delta_j = \theta_{jk}'P\theta_{jk}$,

$$\text{d'où } \|\Pi_j(T)\|^2 = \frac{\|\Pi_j(R)\|^2}{\|R\|^2} = \frac{R'\theta_{jk}\Delta_j^{-1}\theta_{jk}'R}{\|R\|^2}.$$

Ainsi on peut formuler le lemme précédent de la manière suivante :

Proposition.

Pour les lois a priori et la fonction de perte considérés, une procédure de décision optimale au sens de Bayes est :

$$\left\{ \begin{array}{l} \text{décider } H_0 \quad \text{si } \max_{i \neq 0} \frac{R'\theta_{ik}\Delta_i^{-1}\theta_{ik}'R}{\|R\|^2} \leq C \\ \text{décider } H_j \quad \text{si } \frac{R'\theta_{jk}\Delta_j^{-1}\theta_{jk}'R}{\|R\|^2} > \max [C, \max_{i \neq j} \frac{R'\theta_{ik}\Delta_i^{-1}\theta_{ik}'R}{\|R\|^2}], j \neq 0 \end{array} \right.$$

Remarques

- Considérons le cas de la régression linéaire simple, c'est-à-dire le cas où $\dim(Q) = 2$ et Q est engendré par les vecteurs $X = [x_1, \dots, x_n]'$ et $\mathbf{1} = [1, \dots, 1]'$.

Supposons que H_j soit définie par :

$$H_j : Y = \alpha X + \beta \mathbf{1} + \alpha' X_j + \beta' \mathbf{1}_j + \varepsilon, \quad \mathcal{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$$

avec $Y = [y_1, \dots, y_n]'$, $X_j = [0, \dots, x_{j+1}, \dots, x_n]'$.

Dans ce cas $\dim(Q_j) = k = 2$ (Q_j est engendré par $\Pi_{Q^\perp}(X_j)$ et $\Pi_{Q^\perp}(\mathbf{1}_j)$) bien que H_j corresponde à une seule rupture à l'instant j . Si l'on s'intéressait à la détection de deux éventuelles ruptures, Q_j serait engendré par quatre vecteurs et on aurait $k = 4$, etc.

- De façon générale si la rupture concerne p paramètres, la détection de k rupture correspondrait à des hypothèses de dimension kp .

- Dans la partie qui suit nous allons montrer que, si l'on retient seulement de cette procédure le rejet ou l'acceptation de H_0 , elle est équivalente au test du rapport de vraisemblance utilisé dans la littérature pour tester H_0 contre $\bigcup_{j=1}^J H_j$.

2. Procédure de Bayes et test du rapport de vraisemblance.

Considérons toujours les (J+1) hypothèses :

$$H_0 : \mathfrak{L}(Y) = N_n(\mu, \sigma^2 I_n), \mu \in Q$$

$$H_j : \mathfrak{L}(Y) = N_n(v, \sigma^2 I_n), v \in Q \oplus Q_j, j = 1, \dots, J$$

Le test de rapport de vraisemblance de H_0 contre $\bigcup_{j=1}^J H_j$ s'exprime au moyen de :

$$M_1 = \frac{\max_{j \neq 0, v \in Q \oplus Q_j, \sigma > 0} L_j(Y, v, \sigma)}{\max_{\mu \in Q, \sigma > 0} L_0(Y, \mu, \sigma)}$$

où $L_j(Y, v, \sigma)$ désigne la vraisemblance sous H_j .

$$\text{On a } L_j(Y, v, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|Y-v\|^2}{2\sigma^2}},$$

$$\text{d'où } M_1 = \max_{j \neq 0} \left(\frac{\|\Pi_{Q^\perp}(Y)\|^2}{\|Y - \Pi_{Q \oplus Q_j}(Y)\|^2} \right)^{n/2}.$$

M_1 est fonction croissante de $M_2 = \max_{j \neq 0} \|\Pi_j(T)\|^2$.

$$\text{En effet } \|Y - \Pi_{Q \oplus Q_j}(Y)\|^2 = \|\Pi_{Q^\perp}(Y) - \Pi_j(Y)\|^2,$$

$$\text{soit } \|Y - \Pi_{Q \oplus Q_j}(Y)\|^2 = \|\Pi_{Q^\perp}(Y)\|^2 - \|\Pi_j(Y)\|^2,$$

$$\text{d'où } \frac{\|\Pi_{Q^\perp}(Y)\|^2}{\|Y - \Pi_{Q \oplus Q_j}(Y)\|^2} = \frac{1}{1 - \frac{\|\Pi_j(Y)\|^2}{\|\Pi_{Q^\perp}(Y)\|^2}} = \frac{1}{1 - \|\Pi_j(T)\|^2}.$$

Comme la fonction $x \rightarrow \frac{1}{1-x}$ est croissante, on obtient le résultat annoncé.

Le test du rapport de vraisemblance de H_0 contre $\bigcup_{j=1}^J H_j$ peut donc s'écrire :

$$\begin{cases} \text{décider } H_0 & \text{si } \max_{i \neq 0} \|\Pi_i(T)\| \leq C \\ \text{décider } \bigcup_{j=1}^J H_j & \text{si } \max_{i \neq 0} \|\Pi_i(T)\| > C \end{cases}$$

La procédure de décision multiple associée de façon naturelle à ce test est :

$$\begin{cases} \text{décider } H_0 & \text{si } \max_{i \neq 0} \|\Pi_i(T)\| \leq C \\ \text{décider } H_j & \text{si } \|\Pi_j(T)\| > \max[C, \max_{i \neq j} \|\Pi_i(T)\|], j \neq 0 \end{cases}$$

D'après la première partie (Lemme), on sait que cette procédure basée sur le rapport de vraisemblance est optimale au sens de Bayes.

Dans la partie qui suit nous allons appliquer cette procédure au cas de la détection d'une seule rupture successivement dans le cadre de la régression linéaire simple et dans celui d'un échantillon en nous efforçant de faire le lien avec les procédures proposées dans la littérature et montrer ainsi une (nouvelle) propriété d'optimalité de celles-ci.

3. Application à la régression linéaire simple et au cas d'un échantillon.

3.1. Cas de la régression linéaire simple.

On se place dans le cas d'une seule rupture, et on suppose que J est le nombre d'instants d'observation susceptibles de correspondre avec un point de rupture, vérifiant $1 \leq J < n$. L'espace Q dans ce cas est engendré par les vecteurs X et $\mathbf{1}$.

3.1.1. Rupture par saut.

Ce modèle a été étudié par Kim and Siegmund (1989), Maronna and Yohai (1978). Dans ce cas, $k = 1$, Q_j est engendré par $\Pi_{Q^\perp}(\mathbf{1}_j)$,

$$H_0: Y = \alpha X + \beta \mathbf{1} + \varepsilon \quad \mathcal{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$$

$$H_j: Y = \alpha X + \beta \mathbf{1} + \beta' \mathbf{1}_j + \varepsilon \quad \mathcal{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$$

La procédure de Bayes devient :

$$\begin{cases} \text{décider } H_0 & \text{si } \max_{i \neq 0} \|\Gamma_n(i)\| \leq C \\ \text{décider } H_j & \text{si } \|\Gamma_n(j)\| > \max[C, \max_{i \neq j} \|\Gamma_n(i)\|], j \neq 0 \end{cases}$$

$$\text{avec } \|\Gamma_n(i)\|^2 = \frac{\langle R, \mathbf{1}_i \rangle^2}{\Delta_i \|\mathbf{R}\|^2} \text{ et } \Delta_i = \mathbf{1}_i' P \mathbf{1}_i.$$

Expression de $T_n(i)$.

On pose

- $R = [r_1, \dots, r_n]'$ avec $r_i = y_i - (\hat{\alpha}x_i + \hat{\beta})$
- $\hat{\alpha}, \hat{\beta}$, les estimateurs de maximum de vraisemblance de α et β sous l'hypothèse H_0
- $\bar{X}_i = \frac{1}{i} \sum_{k=1}^i x_k$, $\bar{X}_{n-i} = \frac{1}{n-i} \sum_{k=i+1}^n x_k$

Dans ce cas on a
$$\Delta_i = \frac{i(n-i)}{n} \left[1 - \frac{n(n-i)}{i} \frac{(\bar{X}_{n-i} - \bar{X}_i)^2}{\|X - \bar{X}_n\|^2} \right]$$

et
$$\|T_n(i)\|^2 = i^2 \frac{[\bar{Y}_i - \bar{Y}_n - \hat{\alpha}(\bar{X}_i - \bar{X}_n)]^2}{\Delta_i \|R\|^2}$$

Dans ce cas notre procédure est équivalente à celle de Kim and Siegmund (1989) et Maronna and Yohai (1978) basée sur le test du rapport de vraisemblance.

3.1.2. Rupture de pente avec continuité.

On suppose ici qu'il y a continuité de la régression au point x_i où a lieu la rupture de pente. Dans ce cas $k = 1$ et Q_j est engendré par $\Pi_{Q^\perp}(e_j)$ avec $e_j = [0, \dots, 0, x_{j+1} - x_j, \dots, x_n - x_j]'$,

$H_0 : Y = \alpha X + \beta 1 + \varepsilon \quad \mathfrak{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$

$H_j : Y = \alpha X + \beta 1 + \alpha' e_j + \varepsilon \quad \mathfrak{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$

La procédure de Bayes devient :

$$\begin{cases} \text{décider } H_0 \text{ si } \max_{i \neq 0} \|T_n(i)\| \leq C \\ \text{décider } H_j \text{ si } \|T_n(j)\| > \max_{i \neq j} [C, \max \|T_n(i)\|], j \neq 0 \end{cases}$$

avec $\|T_n(i)\|^2 = \frac{\langle R, e_i \rangle^2}{\Delta_i \|R\|^2}$ et $\Delta_i = e_i' P e_i$.

Seuls à notre connaissance, Farley and Hinich (1970) ont étudié ce cas et ont montré que si l'amplitude de la rupture est faible et les poids p_i égaux, alors le test du rapport de vraisemblance est localement optimal pour tester H_0 contre $\bigcup_{j=1}^{n-1} H_j$.

3.1.3. Rupture de pente et/ou par saut.

Le modèle a été considéré par Kim and Siegmund (1989), Booth and Smith (1982). La rupture éventuelle concerne les deux coefficients de régression de façon quelconque. Dans ce cas $k = 2$, Q_j est engendré par les vecteurs $\Pi_{Q^\perp}(X_j)$ et $\Pi_{Q^\perp}(1_j)$,

$$H_0 : Y = \alpha X + \beta 1 + \varepsilon \quad \mathfrak{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$$

$$H_j : Y = \alpha X + \beta 1 + \alpha' X_j + \beta' 1_j + \varepsilon \quad \mathfrak{L}(\varepsilon) = N_n(0, \sigma^2 I_n)$$

La procédure devient :

$$\left\{ \begin{array}{l} \text{décider } H_0 \text{ si } \max_{i \neq 0} \|T_n(i)\| \leq C \\ \text{décider } H_j \text{ si } \|T_n(j)\| > \max[C, \max_{i \neq j} \|T_n(i)\|], \quad j \neq 0 \end{array} \right.$$

avec $\Theta_{12} = [X_j' 1_j']$, $\Delta_i = \theta_{12}' P \Theta_{12}$ et $\|T_n(i)\|^2 = \frac{R' \Theta_{12} \Delta_i^{-1} \Theta_{12}' R}{\|R\|^2}$.

Booth and Smith (1982) ont proposé une procédure bayésienne basée sur la même statistique avec des lois a priori sur les paramètres moins générales que les nôtres (lois non informatives). Par ailleurs Kim and Siegmund (1989) ont proposé la même statistique basée sur le rapport de vraisemblance.

3.2. Cas d'un échantillon gaussien.

C'est la situation la plus étudiée dans la littérature. Dans ce cas Q est engendré par le vecteur 1 , k vaut 1 et Q_j est engendré par $\Pi_{Q^\perp}(1_j)$.

On a les résultats suivants :

- $J = n-1$
- $P = I - \frac{1}{n} H$
- $H = 1' 1$
- $\Delta_j = (n-j) \frac{1}{n}$
- $\frac{\|\Pi_{Q^\perp}(R)\|^2}{\|R\|^2} = \frac{n}{j(n-j)S^2} \left[\sum_{i=1}^j (y_i - \bar{y}) \right]^2$ avec $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2$.

Hawkins (1977), Worsley (1979) ont proposé le test du rapport de vraisemblance basé sur la statistique : $W = \max_{1 \leq k \leq n-1} \|Z_k\|$

avec $Z_k^2 = (n-2)^2 \frac{T_k^2}{s^2 - T_k^2}$ et $T_k^2 = \frac{n}{k(n-k)} \left[\sum_{i=1}^k (y_i - \bar{y}) \right]^2$,

soit $Z_k^2 = (n-2)^2 \frac{\|\Pi_k(T)\|^2}{1 - \|\Pi_k(T)\|^2}$.

Comme $x \rightarrow \frac{x}{1-x}$ est une fonction croissante, les deux règles de décision sont équivalentes. Booth and Smith (1982) ont proposé une procédure bayésienne basée sur la même statistique en prenant les lois de probabilité a priori moins générales que les nôtres (lois non informatives).

Remarques.

- En pratique le choix des p_j peut être fait pour améliorer la procédure en tenant compte des points qu'on juge plus susceptibles de fournir une rupture : Ferreira (1975).

Dans le cas où les hypothèses H_j n'ont pas la même dimension et pour des p_j non tous égaux, l'optimalité de notre procédure au sens de Bayes semble difficile à montrer.

Notons néanmoins qu'une première généralisation immédiate consiste à supposer que les p_j sont nuls pour certains j et égaux par ailleurs, ce qui revient à prendre $J < n-1$.

- Notre formulation géométrique peut être aussi adoptée pour le problème de la rupture au niveau de la variance (voir Caussinus and Vaillant, 1985).

4. Approche de la valeur critique.

Dans notre règle de détection d'une éventuelle rupture, l'hypothèse nulle est rejetée pour $\max_{i \neq 0} \|\Pi_i(T)\|^2 > w$ ($w = C^2$).

Le niveau de la procédure est donc α tel que $\alpha = P_0(\max_{i \neq 0} \|\Pi_i(T)\|^2 > w)$ où P_0 correspond à la loi de probabilité sous H_0 .

D'autre part l'entier i^* appartenant à $\{1, \dots, n\}$ tel que : $\|\Pi_{i^*}(T)\|^2 = \max_{i \neq 0} \|\Pi_i(T)\|^2$ correspond au point de rupture quand H_0 est rejetée. Pour calculer w en fonction d'un niveau choisi α , une démarche analytique est extrêmement difficile ; seules des méthodes basées sur des majorants ou minorants (Bonferroni), des méthodes de simulation ou des méthodes asymptotiques semblent disponibles facilement (Kim and Siegmund, 1989 ; Knowles and Siegmund, 1989).

Nous proposons dans ce qui suit une étude par simulation pour le cas très fréquent de la régression linéaire simple.

Simulation.

Le point délicat est que w dépend a priori du vecteur $X = (x_1, \dots, x_n)'$ (Kim and Siegmund, 1989 ; Farley and Hinich, 1970).

Nous avons choisi deux situations. Dans la première, X est défini par $x_i = i/n$, $1 \leq i \leq n$; dans la deuxième les x_i ont été générées selon une loi normale centrée réduite (après quoi, la suite ainsi obtenue est supposée fixe). Les constantes a , b et σ sont arbitraires (à cause des propriétés d'invariance). On simule les ε_i indépendants suivant une loi normale centrée réduite et on calcule $y_i = ax_i + b + \sigma\varepsilon_i$, ($1 \leq i \leq n$).

Pour calculer les valeurs critiques, on procède comme suit :

par simulations de 1000 cas sous l'hypothèse nulle on calcule 1000 valeurs de $\max_{2 \leq i \leq n-2} \|\Pi_i(T)\|^2$, puis on cherche les quantiles correspondant aux niveaux choisis ; la valeur critique est estimée par la moyenne des deux valeurs de $\max \|\Pi_i(T)\|^2$ encadrant au mieux la valeur du quantile trouvé.

L'étude a été faite successivement pour $n = 10(10)90$.

Nous donnons dans le tableau suivant les valeurs critiques obtenues dans le cas où $x_i = i/n$ (première ligne) et dans le cas où les valeurs x_i ont été générées selon une loi normale centrée réduite (deuxième ligne).

Tableau 1. Valeurs critiques en fonction des niveaux choisis

n	α	0.01	0.02	0.03	0.04	0.05	0.10	0.25
10		0.8844	0.8447	0.8295	0.8148	0.7981	0.7323	0.6743
		0.8573	0.8336	0.8036	0.7856	0.7744	0.7156	0.6578
20		0.5688	0.5425	0.5265	0.5135	0.4984	0.4351	0.3713
		0.5894	0.5535	0.5273	0.4959	0.4865	0.4271	0.3545
30		0.4545	0.4188	0.3878	0.3772	0.3624	0.3185	0.2596
		0.4289	0.3790	0.3652	0.3478	0.3328	0.2817	0.2287
40		0.3452	0.3246	0.3063	0.2986	0.2767	0.2478	0.2021
		0.3272	0.2939	0.2805	0.2689	0.2583	0.2238	0.1889
50		0.3201	0.2824	0.2680	0.2489	0.2367	0.2007	0.1630
		0.2727	0.2436	0.2290	0.2220	0.2169	0.1852	0.1486
60		0.2458	0.2296	0.2190	0.2091	0.2011	0.1746	0.1409
		0.2340	0.2124	0.2028	0.1956	0.1894	0.1575	0.1252
70		0.2023	0.1920	0.1838	0.1764	0.1710	0.1530	0.1228
		0.2037	0.1856	0.1731	0.1638	0.1597	0.1365	0.1041
80		0.1920	0.1780	0.1697	0.1592	0.1538	0.1352	0.1095
		0.1822	0.1721	0.1573	0.1509	0.1432	0.1265	0.0899
90		0.1707	0.1600	0.1492	0.1397	0.1368	0.1225	0.0983
		0.1570	0.1409	0.1355	0.1302	0.1243	0.1134	0.0876

Les résultats du tableau 1 confirment que w dépend des x_i même si c'est relativement peu pour les niveaux usuels ; cela interdit de dresser des tables "définivites" et suggère que la procédure la plus simple est de réaliser, dans chaque cas rencontré, la simulation sous H_0 avec le vecteur x convenable (cette démarche est tout à fait raisonnable avec l'efficacité actuelle des moyens de calcul). Par ailleurs nos valeurs sont très voisines de celles proposées dans le cas où $x_i = i/n$ par Kim and Siegmund (1989).

5. Etude de performance par simulation.

Nous allons essayer de détecter des points de rupture simulés dans le cas de la régression linéaire simple, la rupture cumulant changement de pente et saut (cf. 3.1.3). Sous une hypothèse alternative le vecteur $Y = (y_1, \dots, y_n)'$ est construit comme suit : on simule $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ à partir du générateur de la loi normale centrée réduite, on se donne le vecteur $X = (x_1, \dots, x_n)'$ défini par $x_i = i/n$, les constantes a, b, a_1, b_1, σ et on calcule Y par $Y = aX + b\mathbf{1} + a_1X_j + b_1\mathbf{1}_j + \sigma\varepsilon$. Par simulation de 1000 cas on calcule la proportion de cas où j est jugé point de rupture (H_j) ou, au contraire, H_0 est retenue.

Dans les tableaux suivants, nous donnons quelques uns des résultats obtenus. Nous avons pris $\sigma = 1$ et rappelons que $\|\delta\|$ est donnée par $\delta = a_1\Pi_{Q^\perp}(X_j) + b_1\Pi_{Q^\perp}(\mathbf{1}_j)$ avec $\Pi_{Q^\perp} = I - X(X'X)^{-1}X'$ et $\|\delta\|^2 = \delta'\delta$.

Tableau 2. $n = 30, j = 0, \|\delta\| = 0, \alpha = 0.05$

H_0	Autre alternative
0.952	0.048

Tableau 3. $n = 30, j = 15, \|\delta\| = 0.30, \alpha = 0.05$

H_0	Autre alternative
0.920	0.08

Tableau 4. $n = 30, j = 15, \|\delta\| = 3.90, \alpha = 0.05$

H_0	H_{14}	H_{15}	H_{16}	Autre alternative
0.390	0.039	0,500	0.034	0.037

Tableau 5. $n = 50, j = 25, \|\delta\| = 0.40, \alpha = 0.05$

H_0	Autre alternative
0.946	0.054

Tableau 6. $n = 50, j = 25, \|\delta\| = 5.00, \alpha = 0.05$

H_0	H_{23}	H_{24}	H_{25}	H_{26}	Autre alternative
0.090	0.011	0.066	0.731	0.066	0.036

Tableau 7. $n = 50, j = 10, \|\delta\| = 4.70, \alpha = 0.05$

H_0	H_9	H_{10}	H_{11}	Autre alternative
0.109	0.080	0.543	0.066	0.202

Tableau 8. $n = 50, j = 35, \|\delta\| = 9.50, \alpha = 0.05$

H_{34}	H_{35}	H_{36}	H_{37}	Autre alternative
0.011	0.979	0.009	0.001	0

Les résultats montrent que le point de rupture est très difficile à détecter si l'amplitude de la rupture est faible même si elle est proche du milieu de l'échantillon. Par contre la procédure est performante pour des amplitudes suffisamment fortes même si la rupture a lieu au début ou vers la fin de l'échantillon. Enfin les résultats obtenus par simulation sont cohérents dans le cas d'absence de rupture.

BIBLIOGRAPHIE.

- [1] Booth, N.B and Smith, A.F.M. (1982) *A Bayesian Approach to Retrospective Identification of Change-point*, Journal of Econometrics, 19, 7-22.
- [2] Broemeling, L.F and Choy, J.H. (1981) *Detecting Structural Change in linear models*, Commun. Statist. Theor. Meth., A10 (2u), 2551-2561.

- [3] **Brown, R.L. and Durbin, J. and Evans, J.M.** (1975) *Techniques of Testing the Constancy of Regression Relachips over Time*, J.R.S.S.B. 37, 149-92.
- [4] **Butler, W.** (1983) *Optimal Properties of one-step Variable Selection in Regression*, Ann. Statist., vol. 11, n° 1, 219-24.
- [5] **Caussinus, H. and Vaillant, J.** (1985) *Some Geometric Tools for the Gaussian Linear Model*, in *Linear Statistical Inference, Lecture Notes in Statistics*, n° 35, Springer-Verlag, 1-19.
- [6] **Cox, D.R. and Spjøtvoll, E.** (1982), *On Partitioning Means into Groups*, Scand. J. Statist., 9, 147-152.
- [7] **Deshayes, J. and Picard, D.** (1982) *Test of disorder of regression : Asymptotic Comparison*, Thoery. Prob. Appli., 27, 100-115.
- [8] **Farley, J. and Hinich** (1970) *A test for a Shifting Slope Coefficient in a Linear Model*, J. Amer. Statist. Ass., 65, 1320-99.
- [9] **Ferreira, P.** (1975) *A Bayesian Analysis of switching Regression Model : Known Number of Regimes*, J. Amer. Statist. Ass., 10, 370-75.
- [10] **Freeman, J.M.** (1986) *An Unknown Change-point and Goddness of fit*, The statistician, 35, 335-344.
- [11] **Hawkins, D.M.** (1977) *Testing a Sequence of Observation for a Shift in location*, J. Amer. Statist. Ass., 72, 180-6.
- [12] **Henderson, R.** (1986) *Change-point Problem with correlated Observations with an Application in Material Accountancy*, Technometrics, vol. 28, 381-89.
- [13] **Hinkley, D.J.** (1969) *Inference about the Intersection in Two-phase Regression*, Biometrika, 56, 495-504.
- [14] **Hinkley, D.J.** (1982) *Inference in Two-phase Regression*, J. Amer. Statist. Asso., 66, 736-45.

- [15] **Holbert, D. and Broemeling, L.D.** (1977) *Bayesian inference related to shifting sequences and two-phase regression*, Commun. Statist. Theor. Meth., A6(3), 265-75.
- [16] **James, B. and James, K.L. and Siegmund, D.** (1987) *Test for a change-point*, Biometrika, 74, 1, 71-83.
- [17] **Kim, H. and Siegmund, D.** (1989) *The likelihood ratio test for a change-point in simple linear regression*, Biometrika, 76, 3, 409-23.
- [18] **Knowles, M. and Siegmund, D.** (1989) *On Hotelling's Approach to Testing for a non linear Parameter in Regression*, Intern. Statist., Rev., 57, n° 3, 205-22.
- [19] **McCabe, B.P.M. and Harrison, M.J.** (1980) *Testing the constancy of regression relationships over time using least squares residuals*, App. Statist., 29, 142-48.
- [20] **Maronna, R. and Yohai, U** (1978) *A Bivariate Test for the Detection of a Systematic Change in Means*, J. Amer. Statist. Ass., 73, 640-5.
- [21] **Quandt, R.E.** (1972) *A New Approach to Estimating Switching Regressions*, J. Amer. Statist. Ass., vol. 67, 306-10.
- [22] **Raftery, A.E. and Akman, V.E.** (1986) *Bayesian Analysis of a Poisson Process with a Change-point*, Biometrika, 73, 1, 85-9.
- [23] **Shaban, S.A.** (1980) *Change-point Problem and Two-phase Regression ; an annotated Bibliography*, Intern. Statist. Rev., 48, 83-93.
- [24] **Smith, A.F.M. and Cook, D.** (1980) *Straight lines with a Change-point : a Bayesian Analysis of some Renal Transplant Data*, Appl. Statist., 29, 180-9.
- [25] **Smith, A.F.M. and Cook, D.** (1977) *A Bayesian Analysis of some Time-Varying Models*, in Recent Developments in Statistics, eds. Barra, J.R., Brodeau, F., Romier, G., and Van Cutsem, B., Amsterdam : North-Holland, 257-267.

- [26] Stasinopoulous, M. and Nicholls, A. (1989) *Estimation of Break Point Parameters using GLIM*, GLIM Nrwsletter, 17, 292-300.
- [27] Worsley, K.J. (1979) *On the Likelihood Ratio test for Shift in Location of Normal Population*, J. Amer. Statist. Ass., 74, 36-57.
- [28] Worsley, K.J. (1986) *Confidence Regions and Tests for a Change-point in a Sequence of Exponential Family Variables*, Biometrika, 73, 91-104.
- [29] Zacks, S. (1982) *Classical and Bayesian Approachs to the Change-point Problem : Fixed Samples and Sequential Procedures*, Statistique et Analyse des Données, vol. 7, n° 1, pp. 48.