

DÉTECTION DE RUPTURES MULTIPLES DANS DES SÉRIES TEMPORELLES MULTIVARIÉES

MARC LAVIELLE

Université René Descartes et Université Paris–Sud,
Laboratoire de Mathématiques
Marc.Lavielle@math.u-psud.fr

GILLES TEYSSIÈRE

Statistique Appliquée et MODélisation Stochastique
CES, Université Paris 1 Panthéon–Sorbonne.
stats@gillesteysièrre.net

Lietuvos Matematikos Rinkinyys, **46**, 351-376. 2006

Résumé

Nous considérons le problème de détection de ruptures multiples pour des séries chronologiques multivariées, y compris des processus fortement dépendants, avec un nombre inconnu de ruptures. Nous supposons que la structure de covariance de ces séries chronologiques change de façon abrupte à des dates inconnues. La méthode adaptative proposée ici est capable de détecter des ruptures dans des séries multivariées indépendantes et identiquement distribuées (i.i.d.), faiblement et fortement dépendantes. Cette méthode adaptative surclasse le critère de Schwartz, principalement dans le cas de données faiblement dépendantes. Nous considérons des applications à des séries chronologiques multivariées de rendements d'indices boursiers journaliers et à des séries générées par un marché financier artificiel.

1 Introduction

Détecter des ruptures dans des séries chronologiques multivariées a un intérêt si nous pensons que ces séries sont corrélées, et/ou que les composantes du processus vectoriel multivarié sont générées par le même processus. Cette hypothèse est pertinente pour des marchés financiers où des actifs corrélés sont échangés. Des faits empiriques, reportés par exemple par Teyssièrre [36, 37], mettent en évidence que les volatilités et co-volatilités de plusieurs séries temporelles, des taux de change, ont le même degré de dépendance forte, une propriété qui peut être causée par la non-stationarité commune à ces séries. La présence de dépendance dans la volatilité des actifs financiers est encore un sujet de débats, bien que de nombreux travaux, par exemple Mikosch et Stărică [33], Kokoszka et Teyssièrre [26], Lavielle et Teyssièrre [30], ont montré que la propriété de dépendance forte des volatilités est vraisemblablement un artefact statistique, c-à-d principalement la conséquence de la concaténation de processus dont les variances inconditionnelles sont différentes ; voir aussi Giraitis *et al.* [14] pour une présentation de la littérature sur les modèles de volatilité.

Du point de vue du praticien, les procédures de détection de ruptures sont d'un grand intérêt car on ne connaît pas le processus qui génère les données étudiées. De plus, les données économiques sont habituellement non stationnaires, et il peut être intéressant d'approximer un processus inconnu et peut-être non stationnaire par des processus locaux stationnaires ; voir par exemple Dalhaus [12].

La littérature sur la détection de ruptures est vaste : on peut citer les ouvrages de Basseville et Nikiforov [1], Brodsky et Darkhovsky [7], Csörgö et Horváth [11], Chen et Gupta [8]. Les articles de Giraitis et Leipus [16, 15], Hawkins [19, 20], Chen et Gupta [9], Mia et Zhao [32], et Sen et Srivastava [35] entre autres, présentent aussi un intérêt.

La théorie statistique pour des processus de volatilité faiblement dépendants avec une seule rupture a été développée plus récemment ; voir par exemple Chu [10], Kokoszka et Leipus [24, 25], Horváth, Kokoszka et Teyssièrre [21], Kokoszka et Teyssièrre [26], Berkes *et al.* [2]. Les processus considérés dans ces travaux ne sont plus i.i.d., mais faiblement dépendants. Pour le cas de processus fortement dépendants, le lecteur est renvoyé aux articles de Giraitis, Leipus et Surgailis [17], Lavielle [27], et au chapitre de Kokoszka et Leipus [23] dans le livre sur la dépendance de longue-portée édité par Doukhan *et al.* (2003), qui passe en revue les travaux récents sur la détection de ruptures dans des séries dépendantes univariées.

La réalisation d'une seule rupture sur des données réelles est plutôt rare, car les données économiques, financières, hydrologiques, biologiques, électrotechniques, etc., présentent des ruptures multiples : voir par exemple Schechtman et Wolfe [34], Braun *et al.* [6], Lavielle et Moulines [29], Lavielle et Teyssièrre [30]. Une procédure statistique capable de détecter des ruptures multiples est donc d'un intérêt pratique. Il a souvent été affirmé que les procédures de test pour des ruptures uniques peuvent être étendues au cas de ruptures multiples en utilisant la procédure de segmentation binaire de Vostrikova [39], qui consiste à appliquer la procédure de détection de rupture unique sur l'échantillon complet, ensuite diviser la série en deux au point de rupture détecté, et appliquer itérativement la procédure de détection de rupture sur les deux segments jusqu'à ce que plus aucune rupture ne soit trouvée.

Dans Lavielle et Teyssièrre [30], nous avons traité du problème de comparaison des procédures de détection globale et locale, et nous avons trouvé que l'utilisation des procédures de détection locales pour détecter des ruptures multiples à l'aide de la méthode de segmentation binaire était trompeuse et conduisait à une surestimation du nombre de ruptures.

Une approche globale signifie que toutes les ruptures sont détectées simultanément. Ces ruptures sont estimées en minimisant une fonction de contraste pénalisée $J(\boldsymbol{\tau}, \mathbf{y}) + \beta \text{pen}(\boldsymbol{\tau})$ (voir [3, 27, 40]). Ici, $J(\boldsymbol{\tau}, \mathbf{y})$ mesure l'ajustement du modèle dont la séquence des ruptures est $\boldsymbol{\tau}$ avec les séries observées \mathbf{y} . Son rôle est de localiser les ruptures avec autant de précision que possible. Pour détecter des ruptures dans la moyenne et/ou la matrice de covariance de séries multivariées, nous définissons la fonction de contraste $J(\boldsymbol{\tau}, \mathbf{y})$ à partir du logarithme de la fonction de vraisemblance gaussienne, même si les observations ne sont pas gaussiennes. Le terme de pénalité $\text{pen}(\boldsymbol{\tau})$ dépend seulement de la dimension $K(\boldsymbol{\tau})$ du modèle $\boldsymbol{\tau}$ et augmente avec $K(\boldsymbol{\tau})$. Le paramètre de pénalité β ajuste le compromis entre la minimisation de $J(\boldsymbol{\tau}, \mathbf{y})$ (obtenue avec une grande dimension de $\boldsymbol{\tau}$) et la minimisation de $\text{pen}(\boldsymbol{\tau})$ (obtenue avec petite dimension de $\boldsymbol{\tau}$).

Des résultats asymptotiques ont été obtenus dans des contextes généraux théorique par Lavielle [27], qui étendent les résultats précédents de Yao [40]. Nous verrons que cette approche est aussi très utile pour des applications pratiques, pour détecter des ruptures dans la moyenne et/ou la variance de séries multivariées, avec la restriction que ces séries ont une segmentation commune $\boldsymbol{\tau}$. Une méthode adaptative est proposée pour estimer le nombre de ruptures. Des expériences de simulation numériques montrent que la méthode proposée surclasse le critère de Schwarz et donne de très bons résultats.

Pour des séries chronologiques multivariées, l'algorithme de détection de ruptures est d'ordre $\mathcal{O}(mn^2)$,

où m est la dimension du processus vectoriel, au lieu de l'ordre $\mathcal{O}(n^2)$ dans le cas univarié.

2 Une méthode d'estimation basée sur le contraste pénalisé pour la détection de ruptures en dimension multivariée

2.1 La fonction de contraste

Nous supposons que le processus m -dimensionnel $\{\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{m,t})'\}$ change de façon abrupte et est caractérisé par un paramètre $\theta \in \Theta$ qui reste constant entre deux ruptures. Nous utiliserons fortement cette hypothèse pour définir notre fonction de contraste $J(\boldsymbol{\tau}, \mathbf{Y})$.

Soit K un entier et soit $\boldsymbol{\tau} = \{\tau_1, \tau_2, \dots, \tau_{K-1}\}$ une suite ordonnée d'entiers vérifiant $0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < n$. Pour tout $1 \leq k \leq K$, soit $U(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \theta)$ une fonction de contraste pour estimer la vraie valeur inconnue du paramètre d'intérêt sur le segment k . En d'autres termes, l'estimateur du contraste minimum de $\hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})$, calculé sur le $k^{\text{ème}}$ segment de $\boldsymbol{\tau}$, est défini comme la solution du problème de minimisation suivant :

$$U\left(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})\right) \leq U(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \theta), \quad \forall \theta \in \Theta. \quad (1)$$

Pour tout $1 \leq k \leq K$, définissons G comme suit :

$$G(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}) = U\left(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})\right). \quad (2)$$

Ensuite, définissons la fonction de contraste $J(\boldsymbol{\tau}, \mathbf{Y})$:

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K G(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}), \quad (3)$$

avec $\tau_0 = 0$ et $\tau_K = n$.

Dans cet article, nous considérons des ruptures dans la matrice de covariance de la suite $\{\mathbf{Y}_t\}$. Plus précisément, nous supposons qu'il existe un entier K^* , une suite $\boldsymbol{\tau}^* = \{\tau_1^*, \tau_2^*, \dots, \tau_{K^*}^*\}$ avec $\tau_0^* = 0 < \tau_1^* < \dots < \tau_{K^*-1}^* < \tau_{K^*}^* = n$ et K^* ($m \times m$) matrices de covariance $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_{K^*}$ telles que $\text{Cov}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))' = \boldsymbol{\Sigma}_k$ for $\tau_{k-1}^* + 1 \leq t \leq \tau_k^*$.

Modèle M1: Il existe un vecteur m -dimensionnel μ tel que $\mathbb{E}(\mathbf{Y}_t) = \mu$ pour $t = 1, 2, \dots, n$. De plus, $\boldsymbol{\Sigma}_k \neq \boldsymbol{\Sigma}_{k+1}$ pour $1 \leq k \leq K^* - 1$.

Dans ce cas simple de ruptures dans la matrice de covariance, et sans rupture dans la moyenne, qui présente un intérêt pour des processus de volatilité multivariés processes, nous pouvons utiliser la fonction de contraste suivante, basée sur la fonction de log-vraisemblance gaussienne :

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K n_k \log |\hat{\boldsymbol{\Sigma}}_{\tau_k}|, \quad (4)$$

où $n_k = \tau_k - \tau_{k-1}$ est la longueur du segment k , $\hat{\boldsymbol{\Sigma}}_{\tau_k}$ est la matrice ($m \times m$) de covariance empirique calculée sur le segment k :

$$\hat{\boldsymbol{\Sigma}}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})'. \quad (5)$$

Ici, $\bar{\mathbf{Y}} = n^{-1} \sum_{t=1}^n \mathbf{Y}_t$ est la moyenne empirique de la série m -dimensionnelle \mathbf{Y}_t calculée sur la série complète.

Modèle M2: Il existe K^* vecteurs m -dimensionnels μ_1, \dots, μ_{K^*} tels que $\mathbb{E}(\mathbf{Y}_t) = \mu_k$ pour $\tau_{k-1}^* + 1 \leq t \leq \tau_k^*$. De plus, $(\mu_k, \Sigma_k) \neq (\mu_{k+1}, \Sigma_{k+1})$ pour $1 \leq k \leq K^* - 1$.

Pour la détection de ruptures dans le vecteur des moyennes et/ou la matrice de covariance d'une suite multivariée de variables aléatoires, cette fonction de contraste se réduit à :

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K n_k \log |\widehat{\Sigma}_{\tau_k}| \quad (6)$$

où la matrice $\widehat{\Sigma}_{\tau_k}$ de covariance empirique de dimension $(m \times m)$ est calculée sur le segment k :

$$\widehat{\Sigma}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})' \quad (7)$$

où $\bar{\mathbf{Y}}_{\tau_k} = n_k^{-1} \sum_{t=\tau_{k-1}+1}^{\tau_k} \mathbf{Y}_t$ est la moyenne empirique de la série m -dimensionnelle \mathbf{Y}_t calculée sur ce segment.

Des résultats asymptotiques pour l'estimateur du contraste minimum de $\boldsymbol{\tau}^*$ peuvent être obtenus dans la cadre asymptotique suivant :

A1 Pour tout $1 \leq i \leq m$ et tout $1 \leq t \leq n$, définissons $\eta_{t,i} = Y_{t,i} - \mathbb{E}(Y_{t,i})$. Il existe $C > 0$ et $1 \leq h < 2$ tels que pour tout $u \geq 0$ et tout $s \geq 1$,

$$\mathbb{E} \left(\sum_{t=u+1}^{u+s} \eta_{t,i} \right)^2 \leq C(\theta) s^h. \quad (8)$$

(**A1** est vérifiée avec $h = 1$ pour des suites faiblement dépendantes et $1 < h < 2$ pour des suites fortement dépendantes.)

A2 Il existe une suite $0 < a_1 < a_2 < \dots < a_{K^*-1} < a_{K^*} = 1$ telle que pour tout $n \geq 1$ et pour tout $1 \leq k \leq K^* - 1$, $\tau_k^* = [na_k]$.

Quand le vrai nombre K^* de segments est connu, nous avons le résultat suivant sur la vitesse de convergence de l'estimateur de contraste minimum de $\boldsymbol{\tau}^*$:

Theorème 2.1 *Supposons que les conditions **A1-A2** soient vérifiées. Dans le cas du modèle **M1** (resp. model **M2**), soit $\hat{\boldsymbol{\tau}}_n$ les dates qui minimisent le contraste empirique $J(\boldsymbol{\tau}, \mathbf{Y})$ défini par l'équation (4) (resp. (6)). Alors, la suite $\{n \|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}^*\|_\infty\}$ est uniformément tendue en probabilité :*

$$\lim_{n \rightarrow \infty} \lim_{\delta \rightarrow \infty} \mathbb{P} \left(\max_{1 \leq k \leq K^*-1} |\hat{\tau}_{n,k} - \tau_k^*| > \delta \right) = 0. \quad (9)$$

(Ici, $J(\boldsymbol{\tau}, \mathbf{Y})$ est minimisé sur toutes les séquences possibles $\boldsymbol{\tau}$ de longueur K^*)

Preuve: La preuve est une application directe du Theorème 2.4 de Lavielle [27]. Nous pouvons aisément vérifier que les hypothèses **H1-H2** de [27] sont vérifiées par les modèles **M1** et **M2** et sous les hypothèses **A1-A2**. ■

Ce résultat signifie que la vitesse de convergence de $\hat{\boldsymbol{\tau}}_n$ ne dépend pas de la structure de covariance de la série $\{\mathbf{Y}_t\}$. Pour des séries fortement mélangées, ainsi que pour des séries fortement dépendantes, la vitesse optimale est atteinte puisque $\|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}^*\|_\infty = \mathcal{O}_P(1)$.

2.2 Fonction de pénalité pour le problème de détection de ruptures

Quand le nombre de ruptures est inconnu, nous l'estimons en minimisant une version pénalisée de la fonction $J(\boldsymbol{\tau}, \mathbf{Y})$. Pour toute séquence de dates de ruptures $\boldsymbol{\tau}$, définissons par $\text{pen}(\boldsymbol{\tau})$ la fonction de $\boldsymbol{\tau}$ qui augmente avec le nombre $K(\boldsymbol{\tau})$ de segments de $\boldsymbol{\tau}$. Enfin, soit $\{\hat{\boldsymbol{\tau}}_n\}$ la suite de dates de ruptures qui minimise :

$$U(\boldsymbol{\tau}) = J(\boldsymbol{\tau}, \mathbf{Y}) + \beta \text{pen}(\boldsymbol{\tau}). \quad (10)$$

Cette procédure est intuitivement simple : le critère d'ajustement doit être compensé de telle façon qu'une trop grande segmentation soit pénalisée. Cependant, cette compensation ne doit pas être trop forte car une fonction de pénalisation trop forte conduit à une sous-estimation du nombre de segments.

Si β est une fonction de n qui tend vers 0 à une vitesse appropriée quand n tend vers l'infini, le théorème suivant établit que le nombre de segments estimés converge en probabilité vers K^* et que l'équation (9) est toujours vérifiée.

Theorème 2.2 *Soit $\{\beta_n\}$ une suite positive de nombres réels telle que*

$$\beta_n \xrightarrow[n \rightarrow \infty]{} 0 \text{ and } n^{2-h} \beta_n \xrightarrow[n \rightarrow \infty]{} \infty. \quad (11)$$

*Alors, sous les hypothèses **A1-A2**, le nombre estimé de segments $K(\hat{\boldsymbol{\tau}}_n)$, où $\hat{\boldsymbol{\tau}}_n$ est l'estimateur du minimum de contraste pénalisé de $\boldsymbol{\tau}^*$, obtenu en minimisant $J(\boldsymbol{\tau}, \mathbf{Y}) + \beta_n \text{pen}(\boldsymbol{\tau})$, converge en probabilité vers K^* .*

(Ici, la fonction de contraste $J(\boldsymbol{\tau}, \mathbf{Y})$ est minimisée sur l'ensemble de toutes les suites possibles $\boldsymbol{\tau}$ et pour tous les K possibles, $1 \leq K \leq K_{max}$, K_{max} étant une borne supérieure finie de K^)*

Preuve: la preuve est une application directe du Theorème 3.1 de Lavielle [27]. ■

En pratique, les résultats asymptotiques ne sont pas très utiles pour sélectionner le terme de pénalité $\beta \text{pen}(\boldsymbol{\tau})$. En effet, étant donné un signal observé de taille fixe et finie n , le paramètre β doit être fixé à une valeur arbitraire. Quand le paramètre β est très grand, seuls les ruptures abruptes les plus significatives sont détectées. A l'inverse, une faible valeur pour β produit un nombre élevé de ruptures estimées. Donc un choix doit être fait, nous devons sélectionner une valeur de β donnant un niveau raisonnable de résolution de la segmentation.

Les auteurs suggèrent des fonctions de pénalité différentes en fonction du modèle considéré. Par exemple, le critère de Schwarz est utilisé par Braun *et al.* [6] pour détecter des ruptures dans des séquences ADN.

Considérons d'abord la fonction de pénalité $\text{pen}(\boldsymbol{\tau})$. Par définition, $\text{pen}(\boldsymbol{\tau})$ doit augmenter avec le nombre de segments $K(\boldsymbol{\tau})$. Selon les critères d'information les plus populaires tels que le critère d'Akaike (AIC) et le critère de Schwarz (BIC), la fonction de pénalité la plus simple $\text{pen}(\boldsymbol{\tau}) = K(\boldsymbol{\tau})$ peut être utilisée. De plus, Yao [40] a démontré la consistance du critère de Schwarz pour quelques modèles.

Remarque 2.3 *Dans le cas multivarié i.i.d., le paramètre de pénalité pour le critère de Schwarz est :*

$$\beta = \frac{m(m+1)}{2} \frac{\log n}{n}. \quad (12)$$

Pour réduire le coût en temps calcul de l'algorithme tout en ayant le niveau précision acceptable requis par l'estimation, les ruptures peuvent être détectées sur une sous-grille $d, 2d, 3d, \dots$ de $1, 2, \dots, n$ (nous

utilisons $d = 10$ pour les expériences de simulations numériques). Dans le cas multivarié i.i.d., où la fonction $J(\boldsymbol{\tau}, \mathbf{Y})$ est estimée sur une grille de longueur d , le paramètre de pénalité pour le critère de Schwarz est :

$$\beta = \frac{m(m+1)}{2} \frac{\log(n/d)}{n}. \quad (13)$$

Dans le cadre des modèles GARCH, les séries sont faiblement dépendantes. Dans les cas univarié fortement dépendant, Lavielle and Moulines [29] ont suggéré d'utiliser :

$$\beta = \frac{4 \log n}{n^{1-\vartheta}}, \quad (14)$$

où $\vartheta \in (0, 1)$ est le paramètre d'échelle du processus. Pour des données faiblement dépendantes, $\vartheta = 0$. Cependant, un praticien étudiant des données réelles ne connaît pas ϑ , et doit l'estimer à partir des données. Si le processus n'est pas stationnaire, les méthodes d'estimation dans le domaine spectral donnent des résultats incorrects qui surrestiment ϑ et donc augmentent de façon artificielle β . Une possibilité est d'utiliser des méthodes d'estimation dans le domaine des ondelettes, mais dans ce cas une grande taille d'échantillon est requise ; voir Teyssière and Abry [38].

Remarque 2.4 Des résultats précis ont été obtenus récemment par Birgé et Massart [3] pour le modèle suivant :

$$Y_i = s^*(i) + \sigma \varepsilon_i, \quad 1 \leq i \leq n, \quad (15)$$

où $s^*(i) = \sum_{k=1}^{K^*} m_k \mathbf{1}_{\{\tau_{k-1}^*+1 \leq i \leq \tau_k^*\}}$ est une fonction constante par morceaux. La suite $\{\varepsilon_i\}$ est une suite de bruits blancs gaussiens, de variance égale à 1. \mathbf{y} est alors une suite gaussienne de variance constante σ^2 et de moyenne m_k pour le segment k . L'estimateur de la méthode des moindres carrés pénalisés est obtenu en minimisant

$$U(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^{K(\boldsymbol{\tau})} \sum_{i=\tau_{k-1}+1}^{\tau_k} (Y_i - \bar{Y}_k)^2 + \beta \text{pen}(\boldsymbol{\tau}). \quad (16)$$

Dans un contexte non asymptotique, Birgé and Massart [3] ont montré qu'une fonction de pénalité de la forme

$$\text{pen}(\boldsymbol{\tau}) = K(\boldsymbol{\tau}) \left(1 + c \log \frac{n}{K(\boldsymbol{\tau})} \right), \quad \beta = \frac{2\sigma^2}{n}, \quad (17)$$

est optimale pour minimiser $\mathbb{E}(\|\hat{s}_{\boldsymbol{\tau}} - s^*\|^2)$, où la suite estimée de moyennes $\{\hat{s}_{\boldsymbol{\tau}}(i)\}$ est définie par $\hat{s}_{\boldsymbol{\tau}}(i) = \sum_{k=1}^{K(\boldsymbol{\tau})} \bar{Y}_k \mathbf{1}_{\{\tau_{k-1}+1 \leq i \leq \tau_k\}}$. Sur la base de quelques expériences de simulation numériques, les auteurs suggèrent d'utiliser $c = 2.5$ pour ce modèle. Notons que quand le nombre K^* de segments est faible en comparaison avec la taille n de la série, cette fonction de pénalité optimale est une fonction presque linéaire de K .

2.3 Un choix adaptatif du paramètre de pénalité

Pour une fonction de contraste J donnée et une fonction de pénalité $\text{pen}(\boldsymbol{\tau})$ donnée, le problème se réduit au choix du paramètre β .

Soit K_{MAX} une borne supérieure de la dimension de $\boldsymbol{\tau}$. Pour tout K , $1 \leq K \leq K_{MAX}$, soit \mathcal{T}_K l'ensemble de tous les modèles de dimension K :

$$\mathcal{T}_K = \{\boldsymbol{\tau} = (\tau_0, \dots, \tau_K) \in \mathbb{N}^{K+1}, \tau_0 = 0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < \tau_K = n\}.$$

Par définition le meilleur modèle $\hat{\tau}_K$ de dimension K minimise la fonction de contraste J :

$$\hat{\tau}_K = \arg \min_{\tau \in \mathcal{T}_K} J(\tau, \mathbf{Y}). \quad (18)$$

Notons que la suite $\{\hat{\tau}_K, 1 \leq K \leq K_{MAX}\}$ peut être facilement calculée. En effet, soit \mathcal{G} la matrice triangulaire supérieure de dimension $(n \times n)$ telle que l'élément (i, j) , pour $j \geq i$ est $\mathcal{G}_{i,j} = G(\mathbf{Y}_i, \mathbf{Y}_{i+1}, \dots, \mathbf{Y}_j)$, où $G(\mathbf{Y}_i, \dots, \mathbf{Y}_j)$ est la fonction de contraste calculée avec $(\mathbf{Y}_i, \mathbf{Y}_{i+1}, \dots, \mathbf{Y}_j)$. Donc, pour tout K , $1 \leq K \leq K_{MAX}$, nous devons trouver le chemin $\tau_0 = 0 < \tau_1 < \tau_2 < \dots, < \tau_{K-1} < \tau_K = n$ qui minimise le coût total :

$$J(\tau, \mathbf{Y}) = \frac{1}{n} \sum_{k=1}^K \mathcal{G}_{\tau_{k-1}, \tau_k}. \quad (19)$$

Un algorithme de programmation dynamique peut calculer récursivement les chemins optimaux ($\hat{\tau}_K, 1 \leq K \leq K_{MAX}$), voir Kay [22]. Puisque nous considérons que les composantes du processus vectoriel m -dimensionnel $\{\mathbf{Y}_t\}$ ont la même segmentation τ , cet algorithme requiert $\mathcal{O}(n^2)$ opérations comme dans le cas unidimensionnel.

Supposons que la pénalisation $\text{pen}(\tau)$ dépende seulement de la dimension du modèle, c-à-d de K le nombre de segments, posons

$$J_K = J(\hat{\tau}_K, \mathbf{Y}), \quad (20)$$

$$p_K = \text{pen}(\tau), \quad \forall \tau \in \mathcal{T}_K \quad (21)$$

$$= \text{pen}(\hat{\tau}_K). \quad (22)$$

Donc, pour tout paramètre de pénalité $\beta > 0$, la solution $\hat{\tau}(\beta)$ minimise le contraste pénalisé :

$$\hat{\tau}(\beta) = \arg \min_{\tau} (J(\tau, \mathbf{Y}) + \beta \text{pen}(\tau)) \quad (23)$$

$$= \hat{\tau}_{\hat{K}(\beta)} \quad (24)$$

où

$$\hat{K}(\beta) = \arg \min_{K \geq 1} \{J_K + \beta p_K\}. \quad (25)$$

La solution $\hat{K}(\beta)$ est une fonction constante par morceaux de β . Plus précisément, si $\hat{K}(\beta) = K$,

$$J_K + \beta p_K < \min_{L \neq K} (J_L + \beta p_L). \quad (26)$$

Donc, β vérifie

$$\max_{L > K} \frac{J_K - J_L}{p_L - p_K} < \beta < \min_{L < K} \frac{J_L - J_K}{p_K - p_L}. \quad (27)$$

Donc, il existe une suite $\{K_1 = 1 < K_2 < \dots\}$, et une suite $\{\beta_0 = \infty > \beta_1 > \dots\}$, avec

$$\beta_i = \frac{J_{K_i} - J_{K_{i+1}}}{p_{K_{i+1}} - p_{K_i}}, \quad i \geq 1, \quad (28)$$

telle que $\hat{K}(\beta) = K_i, \forall \beta \in [\beta_i, \beta_{i-1})$. De plus, le sous-ensemble $\{(p_{K_i}, J_{K_i}), i \geq 1\}$ est l'enveloppe convexe de l'ensemble $\{(p_K, J_K), K \geq 1\}$.

La séquence estimée $\hat{\tau}(\beta)$ ne doit pas fortement dépendre du choix du coefficient de pénalisation β . En d'autres termes, une petite variation de β ne doit pas mener à une solution $\hat{\tau}$ radicalement différente. Cette stabilité de la solution par rapport au choix de β sera assurée si nous retenons seulement les plus grands intervalles $[\beta_i, \beta_{i-1}), i \geq 1$.

En résumé, nous proposons la procédure suivante :

1. pour $K = 1, 2, \dots, K_{MAX}$, calculer $\hat{\tau}_K$, $J_K = J(\hat{\tau}_K, \mathbf{Y})$ et $\hat{p}_K = \text{pen}(\hat{\tau}_K)$,
2. calculer les séries $\{K_i\}$ et $\{\beta_i\}$, et les longueurs $\{l_{K_i}\}$ des intervalles $[\beta_i, \beta_{i-1})$,
3. garder le(s) plus grande(s) valeur(s) de K_i telle(s) que $l_{K_i} \gg l_{K_j}$, pour $j > i$.

Remarque 2.5 Choisir le plus grand intervalle sous-estime habituellement le nombre de ruptures. En effet, cet intervalle correspond habituellement à un très petit nombre de ruptures et nous détectons seulement les ruptures les plus brutales avec une telle fonction de pénalité. Ceci explique pourquoi on doit chercher la dimension la plus élevée K_i telle que $l_{K_i} \gg l_{K_j}$, pour tout $j > i$, pour retrouver les détails les plus petits.

Au lieu de calculer seulement une seule configuration de ruptures, cette méthode nous permet de mettre en avant différentes solutions de dimensions différentes. En effet, ce serait une illusion de croire qu'une méthode complètement aveugle peut donner la "meilleure" solution dans chaque situation. Si deux dimensions K_i et K_j vérifient le critère suggéré à l'étape 3, il est plus adéquat de proposer ces deux solutions à l'utilisateur, au lieu d'enlever l'une d'entre elles arbitrairement.

Remarque 2.6 Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- i) examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,
- ii) sélectionner la dimension K pour laquelle J_K cesse de décroître significativement.

En d'autres termes, cette approche heuristique cherche la courbure maximale de la courbe (p_K, J_K) . La dérivée seconde de cette courbe est reliée directement avec la longueur de l'intervalle $([\beta_i, \beta_{i-1}), i \geq 1)$. En effet, si nous représentons les points (p_K, J_K) , pour $1 \leq K \leq K_{MAX}$, β_i est la pente entre les points (p_{K_i}, J_{K_i}) et $(p_{K_{i+1}}, J_{K_{i+1}})$. Donc, chercher quand J_K cesse de décroître équivaut à chercher une rupture dans la pente de cette courbe. Enfin, la variation de la pente au point (p_K, J_K) est précisément la longueur l_{K_i} de cet intervalle $[\beta_i, \beta_{i-1})$.

2.4 Une procédure automatique pour estimer K

La méthode proposée requiert une inspection soigneuse de la séquence des longueurs $\{l_i\}$ mais est difficile à automatiser. Nous proposons une autre approche pour sélectionner le modèle qui fournit de très bons résultats et qui est bien plus facile à automatiser pour des applications pratiques.

L'idée de cette méthode est de modéliser la décroissance de la suite $\{J_K\}$ quand il n'y a pas de rupture dans la série $\{\mathbf{Y}_t\}$ et de chercher pour quelle valeur de K ce modèle ajuste la série des contrastes observés.

Sans rupture dans la variance, la distribution jointe de la séquence $\{J_K\}$ est très difficile à calculer de façon analytique, mais quelques simulations de Monte-Carlo montrent que cette séquence décroît comme $c_1 K + c_2 K \log(K)$.

Un exemple numérique est illustré par la Figure 1. Nous avons simulé dix séries de variables aléatoires gaussiennes i.i.d. et calculé les séries $\{J_K\}$ pour chacune d'entre elles. L'ajustement avec la fonction $c_1 K + c_2 K \log(K)$ est presque toujours parfait ($r^2 > 0.999$). Néanmoins, les coefficients c_1 et c_2 sont différents pour chacune de ces séries.

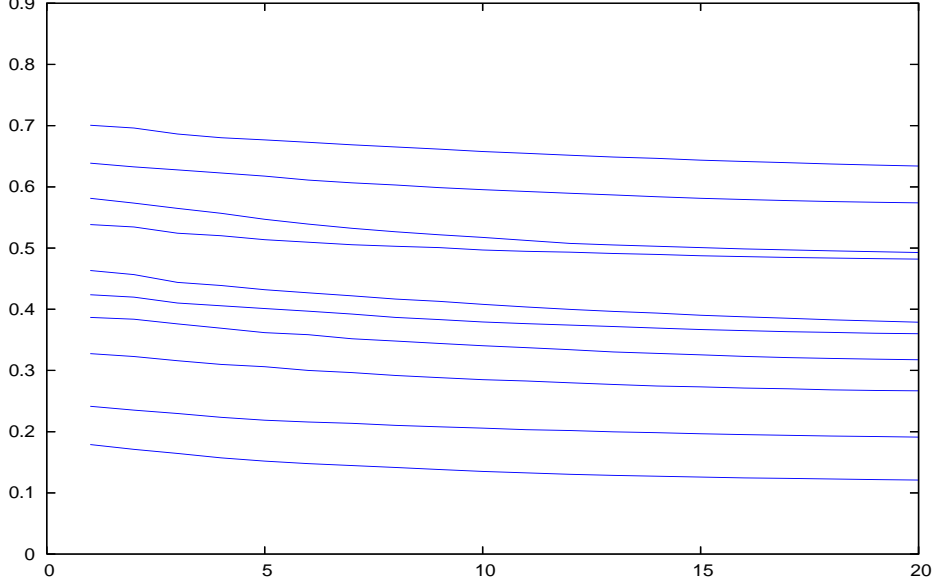


Figure 1: Dix séquences de fonctions de contraste $\{J_K\}$ calculées à partir de dix séquences de variables aléatoires gaussiennes i.i.d., dont le coefficient de corrélation est $\rho = 0.5$

Nous proposons donc l'algorithme suivant :

Algorithme 1 Pour $i = 1, 2, \dots$,

1. ajuster le modèle :

$$J_K = c_1 K + c_2 K \log(K) + e_K,$$

à la série $\{J_K, K \geq K_i\}$, en supposant que $\{e_K\}$ est une suite de variables aléatoires gaussiennes centrées et i.i.d.,

2. évaluer la probabilité que J_{K_i-1} suive aussi ce modèle, c-à-d., estimer la probabilité

$$\mathcal{P}_{K_i} = P(e_{K_i-1} \geq J_{K_i-1} - \hat{c}_1(K_i - 1) + \hat{c}_2(K_i - 1) \log(K_i - 1)), \quad (29)$$

sous le modèle estimé.

Alors, le nombre estimé de segments sera la plus grande valeur de K_i telle que la P -valeur \mathcal{P}_{K_i} soit plus petite qu'un seuil donné α . Nous fixons $\alpha = 10^{-7}$ et $K_{MAX} = 20$ dans les exemples numériques.

Remarque 2.7 Une nouvelle méthode a été introduite par Lavielle et Ludeña [28] pour estimer le nombre de coefficients significatifs pour des problèmes de sélection de modèles. Le contraste utilisé pour ce problème de seuillage est basé sur un centrage aléatoire bien choisi des sommes cumulées partielles des observations ordonnées. Ce contraste se comporte aussi comme $c_1 K + c_2 K \log(K)$ mais dans ce cas les constantes c_1 et c_2 peuvent être calculées analytiquement. En effet, nous montrons dans ce papier que ce contraste décroît comme $-K - K \sum_{k=K+1}^n 1/k \approx -K - K \log(n/K)$.

3 Expériences numériques avec des données simulées

3.1 Le cas i.i.d.

Nous évaluons la performance de la procédure de détection de ruptures avec des fonctions de pénalité différentes en considérant le cas d'un processus i.i.d. Normal bivarié ($m = 2$) $\{\mathbf{Y}_t = (Y_{1,t}, Y_{2,t})'\}$. Sous l'hypothèse nulle de matrice de covariance constante, le processus est défini ainsi :

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} \right], \quad (30)$$

où $\Sigma_{\tau_k} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}$ est constante sur tous les intervalles $[\tau_k, \tau_{k+1}]$.

Nous étudions la taille et la puissance de la procédure pour des échantillons de taille $n = 500$ et $n = 1000$, sous l'hypothèse nulle d'absence de rupture et sous deux hypothèses alternatives avec ruptures dans la matrice Σ_t

1. **PGD 0**: aucun changement dans les paramètres de Σ_t

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \quad t = 1, \dots, n, \quad \sigma_1^2 = \sigma_2^2 = 1, \sigma_{1,2} = 0.5, \quad |\Sigma| = 0.75, \quad (31)$$

Nous considérons comme alternatives deux processus générant les données (PGD) avec deux ruptures dans la matrice de covariance aux dates τ_1 et τ_2 ,

2. **PGD 1**: deux ruptures importantes (en amplitude) dans les paramètres de Σ_t

$$\begin{aligned} \Sigma_t &= \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \quad t = 1, \dots, \tau_1, \quad \sigma_1^2 = \sigma_2^2 = 1, \sigma_{1,2} = 0.5, \quad |\Sigma_t| = 0.75, \\ \Sigma_t &= \begin{pmatrix} \bar{\sigma}_1^2 & \bar{\sigma}_{1,2} \\ \bar{\sigma}_{1,2} & \bar{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_1 + 1, \dots, \tau_2, \quad \bar{\sigma}_1^2 = 1, \bar{\sigma}_2^2 = 2, \bar{\sigma}_{1,2} = \frac{1}{\sqrt{2}}, \quad |\Sigma_t| = 1.5, \\ \Sigma_t &= \begin{pmatrix} \tilde{\sigma}_1^2 & \tilde{\sigma}_{1,2} \\ \tilde{\sigma}_{1,2} & \tilde{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_2 + 1, \dots, n, \quad \tilde{\sigma}_1^2 = 2, \tilde{\sigma}_2^2 = \frac{1}{\sqrt{2}}, \tilde{\sigma}_{1,2} = 1, \quad |\Sigma_t| = \sqrt{2} - 1, \end{aligned}$$

3. **PGD 2**: une forte rupture et une rupture de faible ampleur dans les paramètres de Σ_t

$$\begin{aligned} \Sigma_t &= \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \quad t = 1, \dots, \tau_1, \quad \sigma_1^2 = \sigma_2^2 = 1, \sigma_{1,2} = 0.5, \quad |\Sigma_t| = 0.75, \\ \Sigma_t &= \begin{pmatrix} \bar{\sigma}_1^2 & \bar{\sigma}_{1,2} \\ \bar{\sigma}_{1,2} & \bar{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_1 + 1, \dots, \tau_2, \quad \bar{\sigma}_1^2 = 1, \bar{\sigma}_2^2 = 2, \bar{\sigma}_{1,2} = \sqrt{1.3}, \quad |\Sigma_t| = 0.7, \\ \Sigma_t &= \begin{pmatrix} \tilde{\sigma}_1^2 & \tilde{\sigma}_{1,2} \\ \tilde{\sigma}_{1,2} & \tilde{\sigma}_2^2 \end{pmatrix}, \quad t = \tau_2 + 1, \dots, n, \quad \tilde{\sigma}_1^2 = 1.5, \tilde{\sigma}_2^2 = 2.2, \tilde{\sigma}_{1,2} = \sqrt{1.5}, \quad |\Sigma_t| = 1.5. \end{aligned}$$

Remarque 3.1 *Ce qui importe réellement est l'amplitude des ruptures dans les composantes de Σ_{τ_k} et non dans le déterminant $|\Sigma_{\tau_k}|$: $|\Sigma_{\tau_k}|$ peut rester constant mais les ruptures sont détectées pourvu que les composantes de Σ_{τ_k} changent substantiellement.*

3.2 Le cas faiblement dépendant : processus GARCH multivariés non-homogènes

Nous considérons ici le modèle bivarié GARCH avec corrélation conditionnelle constante introduit par Bollerslev [4], et défini de la façon suivante :

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \mathbf{\Sigma}_t^{\frac{1}{2}} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right], \quad (32)$$

où les éléments diagonaux de $\mathbf{\Sigma}_t$ varient dans le temps et sont des processus GARCH(1,1) univariés, c-à-d,

$$\mathbf{\Sigma}_t = \begin{pmatrix} \sigma_{1,t}^2 & \rho\sigma_{1,t}\sigma_{2,t} \\ \rho\sigma_{1,t}\sigma_{2,t} & \sigma_{2,t}^2 \end{pmatrix}, \quad \rho \in (-1, 1), \quad \begin{aligned} \sigma_{1,t}^2 &= \omega_1 + \beta_1\sigma_{1,t-1}^2 + \alpha_1Y_{1,t-1}^2 \\ \sigma_{2,t}^2 &= \omega_2 + \beta_2\sigma_{2,t-1}^2 + \alpha_2Y_{2,t-1}^2 \end{aligned}. \quad (33)$$

Le coefficient de corrélation ρ est constant. Cette spécification simple implique que la matrice de covariance conditionnelle est toujours définie positive. Les lecteurs intéressés par les propriétés statistiques formelles des processus multivariés de type ARCH sont renvoyés à la lecture de Boussama [5] et Doukhan, Teyssière et Winant [13].

Nous étudions la performance de procédures de détection de ruptures dans deux cas :

1. **PGD 3** : l'hypothèse nulle de constance des paramètres des composantes GARCH(1,1) de $\mathbf{\Sigma}_t$. Nous fixons $\omega_1 = 0.1$, $\beta_1 = 0.3$, $\alpha_1 = 0.2$, $\omega_2 = 0.15$, $\beta_2 = 0.2$, $\alpha_2 = 0.2$ et $\rho = 0.5$.
2. **PGD 4** : l'hypothèse alternative de constance par intervalles des paramètres de la forme fonctionnelle des éléments diagonaux de $\mathbf{\Sigma}_t$. Ce processus bivarié localement stationnaire est défini ainsi :

$$\begin{aligned} \sigma_{1,t}^2 &= \omega_1 + \beta_1\sigma_{1,t-1}^2 + \alpha_1Y_{1,t-1}^2, & \rho = 0.5, & \quad t = 1, \dots, \tau_1, \\ \sigma_{2,t}^2 &= \omega_2 + \beta_2\sigma_{2,t-1}^2 + \alpha_2Y_{2,t-1}^2 \end{aligned} \quad (34)$$

$$\begin{aligned} \sigma_{1,t}^2 &= \bar{\omega}_1 + \bar{\beta}_1\sigma_{1,t-1}^2 + \bar{\alpha}_1Y_{1,t-1}^2, & \rho = 0.3, & \quad t = \tau_1 + 1, \dots, \tau_2, \\ \sigma_{2,t}^2 &= \bar{\omega}_2 + \bar{\beta}_2\sigma_{2,t-1}^2 + \bar{\alpha}_2Y_{2,t-1}^2 \end{aligned} \quad (35)$$

$$\begin{aligned} \sigma_{1,t}^2 &= \tilde{\omega}_1 + \tilde{\beta}_1\sigma_{1,t-1}^2 + \tilde{\alpha}_1Y_{1,t-1}^2, & \rho = 0.7, & \quad t = \tau_2 + 1, \dots, n. \\ \sigma_{2,t}^2 &= \tilde{\omega}_2 + \tilde{\beta}_2\sigma_{2,t-1}^2 + \tilde{\alpha}_2Y_{2,t-1}^2 \end{aligned} \quad (36)$$

Nous fixons $\omega_1 = 0.1$, $\beta_1 = 0.3$, $\alpha_1 = 0.2$, $\omega_2 = 0.15$, $\beta_2 = 0.2$, $\alpha_2 = 0.2$, $\tilde{\omega}_1 = \bar{\omega}_1 = 0.2$, $\tilde{\beta}_1 = \bar{\beta}_1 = 0.1$, $\tilde{\alpha}_1 = \bar{\alpha}_1 = 0.1$, $\tilde{\omega}_2 = \bar{\omega}_2 = 0.05$, $\tilde{\beta}_2 = \bar{\beta}_2 = 0.3$, $\tilde{\alpha}_2 = \bar{\alpha}_2 = 0.2$, c-à-d, à la date τ_1 tous les paramètres du processus changent, tandis qu'à la date τ_2 , seul le coefficient de corrélation ρ change.

Pour $n = 500$, nous fixons $\tau_1 = 200$ et $\tau_2 = 350$, tandis que pour $n = 1000$, nous choisissons $\tau_1 = 400$ et $\tau_2 = 700$.

Les tables 1, 3 et 2 ci-dessous reportent les performances de la procédure de détection de ruptures en utilisant le critère de Schwarz, équation (13), pour les **PGD 1** à **PGD 4**.

Remarque 3.2 Dans la cadre du modèle GARCH bivarié avec une seule rupture, nous avons fait une comparaison, sur la base de simulations, entre cette procédure et l'extension au cas multivarié de procédures de détection paramétriques univariées, par exemple le test du rapport de vraisemblance généralisé ; voir Kokoszka et Teyssière [26]. La procédure semiparamétrique gaussienne considérée ici apparaît largement plus rapide car estimer les paramètres d'un modèle GARCH(1,1) bivarié sur une fenêtre mobile est assez coûteux en temps calcul.

Table 1: Nombre moyen de ruptures détectées et leur localisation en utilisant le critère de Schwarz, $n = 500, \tau_1 = 200, \tau_2 = 350$ (sur la base de 5000 simulations). Les écart-types sont entre parenthèses

PGD	Nombre de ruptures	$\hat{\tau}_1$	$\hat{\tau}_2$
PGD 0	0.2590 (0.59)	—	—
PGD 1	2.3148 (0.67)	191.1580 (52.11)	324.4140 (59.33)
PGD 2	2.3310 (0.66)	187.6320 (41.33)	325.5490 (62.23)
PGD 3	2.1626 (1.47)	—	—
PGD 4	3.8324 (1.55)	145.6920 (73.07)	243.7830 (100.99)

Table 2: Nombre moyen de ruptures détectées et leur localisation en utilisant le critère de Schwarz, $n = 1000, \tau_1 = 400, \tau_2 = 700$ (sur la base de 5000 simulations). Les écart-types sont entre parenthèses

PGD	Nombre de ruptures	$\hat{\tau}_1$	$\hat{\tau}_2$
PGD 0	0.1354 (0.43)	—	—
PGD 1	2.2102 (0.51)	385.6060 (70.00)	666.3900 (97.71)
PGD 2	2.2010 (0.50)	385.5100 (63.74)	670.5840 (94.02)
PGD 3	2.4684 (1.68)	—	—
PGD 4	4.2904 (1.83)	287.1220 (143.60)	466.998 (208.85)

Table 3: Nombre moyen de ruptures détectées en utilisant le critère de Schwarz, $n = 100$ (sur la base de 5000 simulations). Les écart-types sont entre parenthèses

PGD	Nombre de ruptures
PGD 0	1.2678 (1.39)
PGD 3	2.1618 (1.68)

Remarque 3.3 Dans le cas faiblement dépendant, c-à-d les **PGD 3** et **PGD 4**, l'utilisation du critère de Schwarz défini par l'équation (13) conduit à une surestimation du nombre de ruptures.

Remarque 3.4 Dans le cas faiblement dépendant, c-à-d le **PGD 4**, si les paramètres ω_j, β_j et α_j changent de telle façon que les variances inconditionnelle σ_j^2 des processus $GARCH(1,1)$ correspondants, égales à $\omega_j/(1 - \beta_j - \alpha_j)$, $j = 1, 2$, restent constantes ainsi que ρ avant et après la rupture, alors le nombre de ruptures détectées tend vers le nombre de ruptures détectées sous hypothèse nulle du **PGD 3**. Ceci est similaire à ce qui est observé pour des tests de rupture dans le cas univarié ; voir Kokoszka et Teyssière [26]. Cependant, dans le cas multivarié, nous attendons que ce cas de constance de la variance inconditionnelle pour toutes les composantes de Σ_t ne se produise pas, donc la procédure reste utile.

Les tables 4, 5 et 6 ci-dessous reportent la performance de la méthode adaptative. Bien que cette procédure automatique ne soit pas adaptée à un exercice de simulation, car elle requiert une interaction avec l'utilisateur pour sélectionner la segmentation, elle à une meilleure performance que celle du critère de Schwartz, principalement dans le cas de données faiblement dépendantes, ce qui est intéressant pour étudier des données financières. Une comparaison entre les résultats reportés dans les tables 3 et 6 montrent que pour de très petits échantillons ($n = 100$) générés par un processus homogène, l'utilisation du critère de Schwartz conduit à une surestimation du nombre de ruptures, tandis que la méthode

adaptative donne des résultats fiables. Pour les plus grands échantillons, la méthode adaptative donne des estimations précises des dates de rupture : le biais et les écart-types des dates de ruptures estimées sont bien plus petits quand on utilise la méthode adaptative que quand on utilise le critère de Schwartz.

Table 4: Nombre moyen de ruptures détectées et leur localisation en utilisant la méthode adaptative, $n = 500, \tau_1 = 200, \tau_2 = 350$ (sur la base de 5000 simulations). Les écart-types sont entre parenthèses

PGD	Nombre de ruptures	$\hat{\tau}_1$	$\hat{\tau}_2$
PGD 0	0.1248 (0.62)	—	—
PGD 1	1.7974 (0.52)	236.1160 (69.32)	345.6080 (27.48)
PGD 2	1.8290 (0.61)	196.7920 (25.61)	342.9910 (43.26)
PGD 3	0.2962 (0.90)	—	—
PGD 4	1.5650 (0.83)	217.1770 (64.31)	330.1390 (61.25)

Table 5: Nombre moyen de ruptures détectées et leur localisation en utilisant la méthode adaptative, $n = 1000, \tau_1 = 400, \tau_2 = 700$ (sur la base de 5000 simulations). Les écart-types sont entre parenthèses

PGD	Nombre de ruptures	$\hat{\tau}_1$	$\hat{\tau}_2$
PGD 0	0.1312 (0.62)	—	—
PGD 1	1.9968 (0.19)	405.4200 (47.37)	697.6540 (29.17)
PGD 2	2.0508 (0.32)	396.2610 (33.38)	693.3310 (57.41)
PGD 3	0.3130 (0.84)	—	—
PGD 4	2.0554 (0.74)	402.9410 (83.95)	671.697 (103.005)

Table 6: Nombre moyen de ruptures détectées en utilisant la méthode adaptative pour $n = 100$ (sur la base de 5000 simulations). Les écart-types sont entre parenthèses

PGD	Nombre de ruptures
PGD 0	0.1442 (0.66)
PGD 3	0.1314 (0.59)

4 Application à des séries financières multivariées

Puisque la méthode fonctionne bien pour détecter des ruptures dans des processus GARCH multivarié qui sont largement utilisés pour modéliser des séries multivariées de rendements d'actifs financiers, nous appliquons cette méthode à des séries multivariées de rendements sur des données de marchés financiers réels et artificiels.

4.1 Le série bivariée de rendements sur les indices FTSE 100 et S&P500

Nous considérons le processus bivarié de rendements journaliers $(r_{1,t}, r_{2,t})$, avec $r_{i,t} = 100 \times \log(Y_{i,t}/Y_{i,t-1})$, les $Y_{i,t}$ étant observés sur une échelle des temps commune entre janvier 1986 et novembre 2002, c-à-d, $n = 4225$.

Les résultats obtenus avec la méthode adaptative et $d = 1$ sont reportés dans la table 7 et par la Figure 2. Les nombres possibles de segments obtenus par pénalisation sont 1, 4, 6, 11, 19 et 24. Le nombre de segments estimés est clairement $\hat{K} = K_3 = 6$. En effet, cette solution est obtenue pour $\beta \in [85.57, 203.22)$: la longueur de cet intervalle est $l_3 = 117.65$. De plus, la P -valeur pour $K_3 = 6$ est $p_3 = 8.81e-10$ qui est bien plus petite que les P -valeurs obtenues avec des valeurs plus grandes de K . Nous voyons que l'ajustement obtenu pour la série $(J_K, 6 \leq K \leq 30)$ est excellent, mais que ce modèle produit une prédiction sous-estimée de J_5 .

K_i	β_i	β_{i-1}	l_i	p_i
1	692.97	Inf	Inf	-
4	203.22	598.20	394.98	1.76 e-011
6	85.57	203.22	117.65	8.81 e-010
11	41.46	65.88	24.42	0.0134
19	29.68	37.03	7.35	0.2808
24	22.18	27.60	5.42	0.0068

Table 7: Indices FTSE 100 et S&P 500 : intervalles du paramètre de pénalité et les P -valeurs

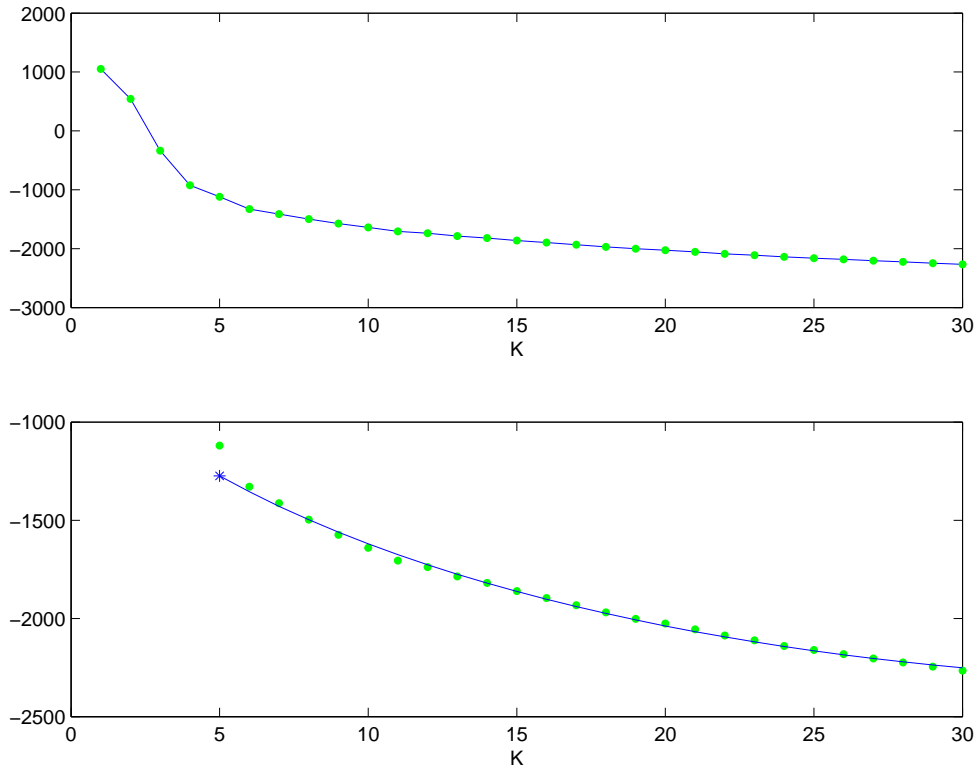


Figure 2: Indices FTSE 100 et S&P500 : détection adaptative avec $\alpha = 10^{-7}$. Haut : la série de contrastes $\{J_K, 1 \leq K \leq 30\}$; Bas : \bullet : la série de contrastes $\{J_K, 6 \leq K \leq 30\}$; - : le modèle ajusté pour $\{J_K, 6 \leq K \leq 30\}$; * : la valeur prédite de J_5 avec ce modèle ajusté

En utilisant la méthode adaptative, nous obtenons 6 segments et $\hat{\tau} = \{448, 508, 1715, 2826, 4119\}$ tandis qu'avec le critère de Schwartz, nous obtenons $\hat{\tau} = \{447, 457, 508, 672, 951, 962, 1152, 1224, 1580, 1715, 2466, 2750, 2967, 2978, 3161, 3215, 3995, 4116\}$. Des résultats similaires sont obtenus avec une sous-grille ($d = 5$ ou $d = 10$).

La segmentation sélectionnée par la méthode adaptative correspond aux évènements suivants :

- $\hat{\tau}_1 = 448$, $\hat{\tau}_2 = 508$: l'accroissement et la baisse de la volatilité après le krach boursier d'octobre 1987,
- $\hat{\tau}_4 = 2826$: la crise Asiatique de l'été 1997,
- $\hat{\tau}_5 = 4119$: mai 2002, augmentation de la volatilité due à l'éclatement de la bulle Internet, et aux banqueroutes frauduleuses d'Enron et WorldCom, qui ont jeté un doute sur la position comptable réelle de plusieurs firmes, et ont accru le niveau d'incertitude.

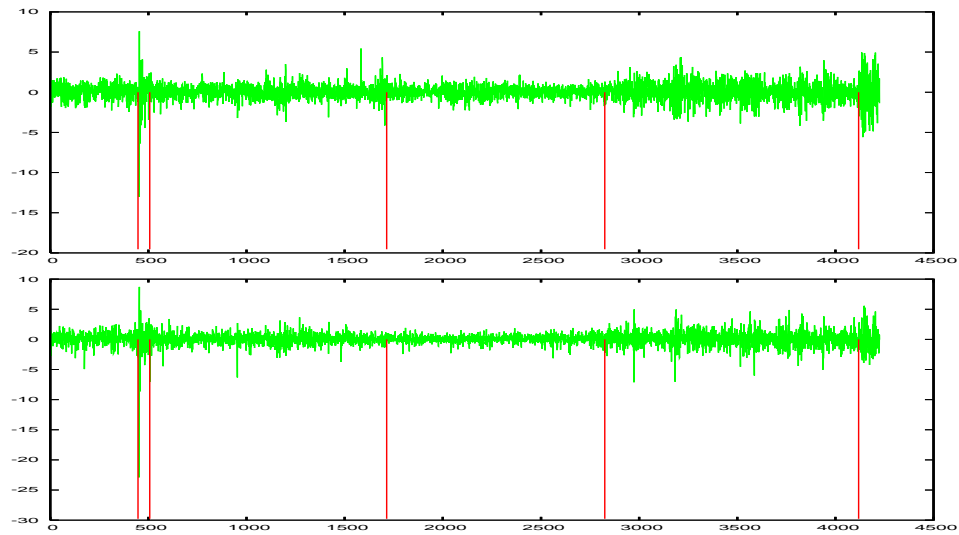


Figure 3: Détection adaptative. Haut : La série des rendements sur le FTSE 100 avec les ruptures estimées représentées par des lignes verticales ; Bas : La série des rendements sur le S&P 500 avec les ruptures estimées représentées par des lignes verticales

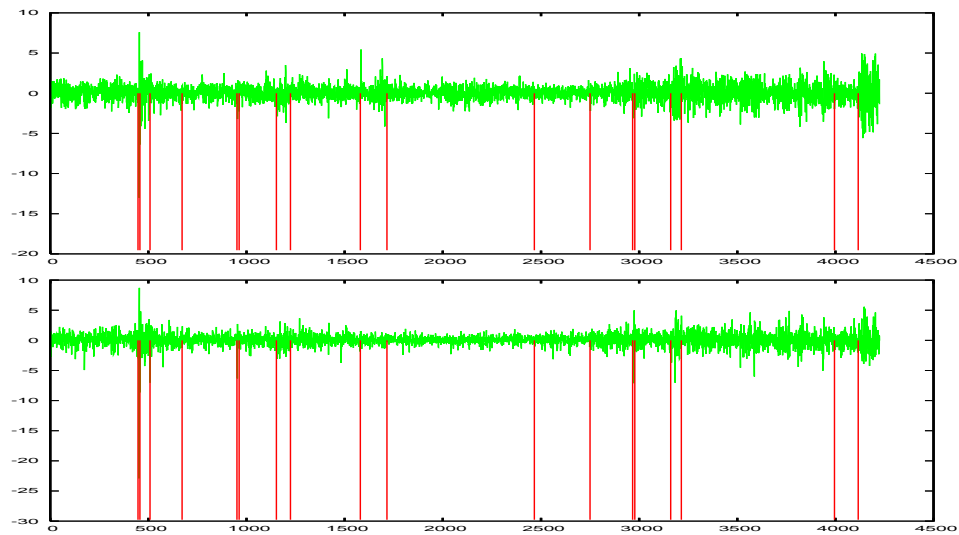


Figure 4: Critère de Schwarz. Haut : La série des rendements sur le FTSE 100 avec les ruptures estimées représentées par des lignes verticales ; Bas : La série des rendements sur le S&P 500 avec les ruptures estimées représentées par des lignes verticales

Il apparaît que la procédure adaptative détecte seulement les ruptures majeures, c-à-d les krachs boursiers.

Pour illustrer simplement la segmentation, la figure 5 ci-dessous montre la fonction d'autocorrélation empirique (ACF) des séries de rendements en valeur absolue sur les indices S&P 500 ($|r_S|$), FTSE 100 ($|r_F|$), et la co-volatilité des deux séries $\sqrt{|r_S r_F|}$, pour l'échantillon complet et pour le segment [508 : 1715]. L'apparente dépendance forte de ces trois séries, telle qu'observée sur l'échantillon complet, apparaît comme un artefact statistique : il y a en effet une forme de dépendance forte, car la fonction d'autocorrélation empirique n'est pas nulle pour les ordres les plus élevés, mais le comportement de l'autocorrélation empirique sur l'intervalle [508 : 1715] est différent de celui observé sur l'échantillon complet.

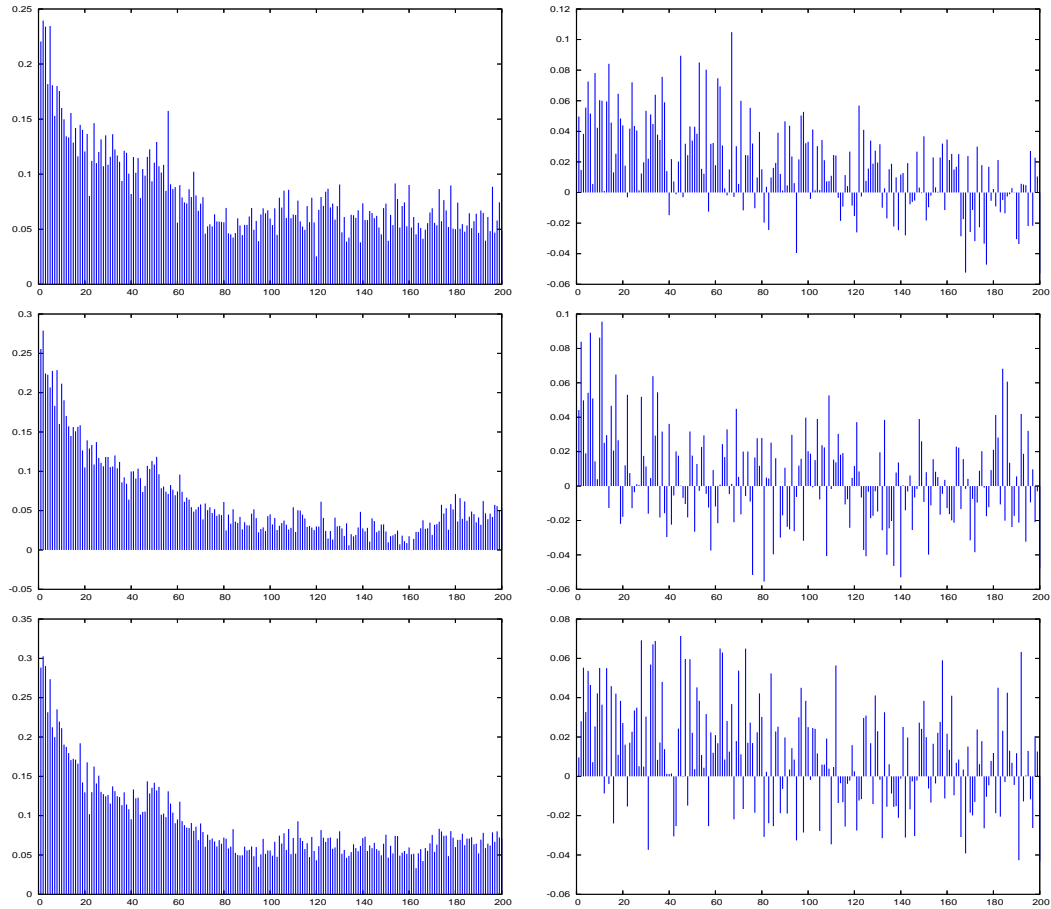


Figure 5: Colonne de gauche : De haut en bas les autocorrélations des rendements en valeur absolue sur le S&P 500 ($|r_S|$), le FTSE 100 ($|r_F|$), et la série de leur co-volatilité $\sqrt{|r_S r_F|}$ sur l'échantillon complet. Colonne de droite : La fonction d'autocorrélation empirique de ces trois séries sur l'intervalle [508 : 1715]

4.2 Une série trivariée de rendements d'actifs dans le secteur financier

Nous considérons ici trois séries de logarithmes de rendements d'actifs sur trois banques, Citybank (City), Bank of America (BoA) et Banque Nationale de Paris (BNP), observés journalièrement entre le 8 avril 1999 et avril 2002, soit 752 observations. Ces séries ont été étudiées dans un cadre univarié par Kokoszka et Teyssière [26]. Avec la méthode adaptative nous obtenons $\hat{\tau} = \{512, 612, 634\}$, les deux dernières ruptures correspondant aux fortes variations causées par les événements du 11 septembre 2001.

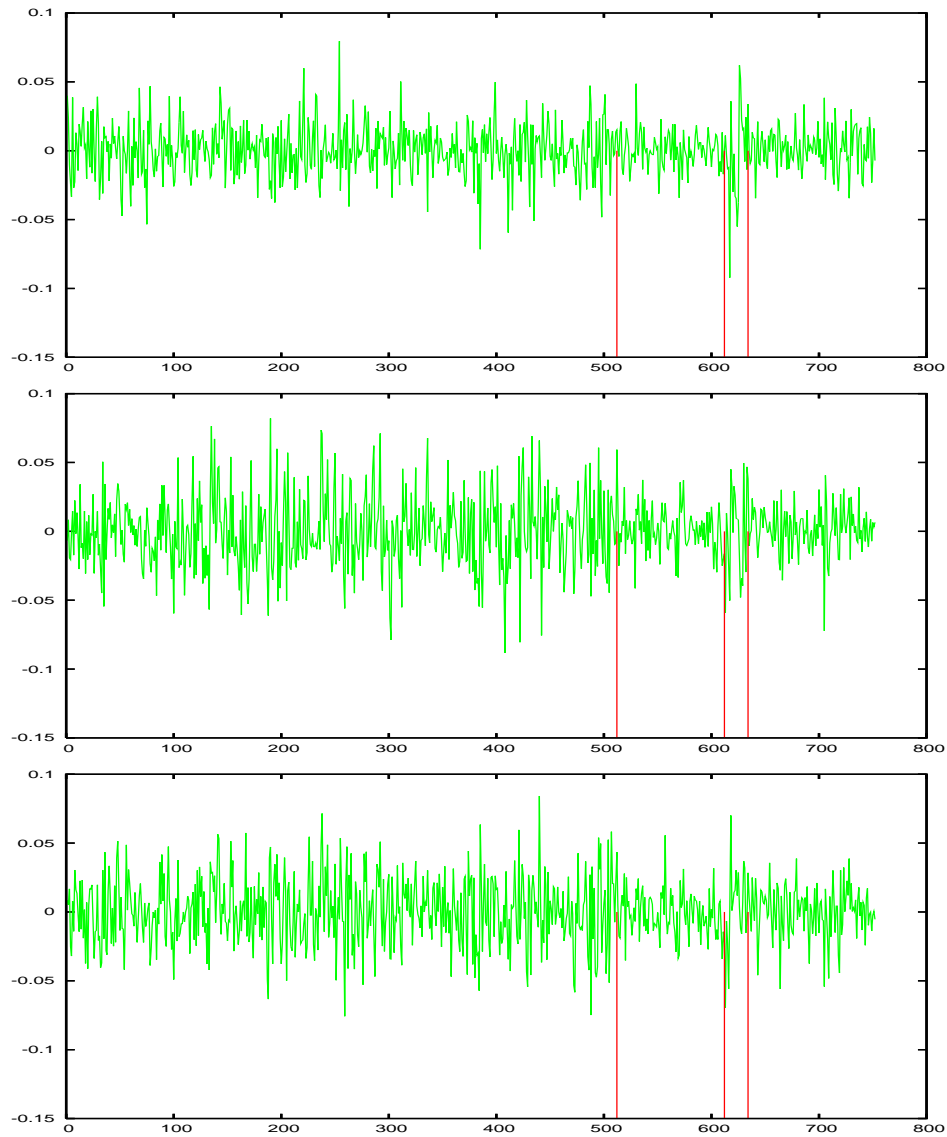


Figure 6: Méthode adaptative:

Haut : La série des rendements sur l'actif BNP avec les ruptures estimées représentées par des lignes verticales ; Milieu : La série des rendements sur l'actif BoA avec les ruptures estimées représentées par des lignes verticales ; Bas : La série des rendements sur l'actif City avec les ruptures estimées représentées par des lignes verticales.

4.3 Marchés financiers artificiels

Dans les deux exemples précédents, le processus de ruptures était inconnu, et la cause de celles-ci semblait exogène, c-à-d, causée par des évènements extérieurs. Nous considérons ici un processus multivarié avec processus de ruptures connu et endogène. Nous étudions des séries multivariées simulées d'un marché financier artificiel, c-à-d, un système dynamique qui modélise des marchés financiers avec agents interactifs. Ce processus bivarié $(P_{1,t}, P_{2,t})$ de taux de change a été introduit par Teyssière [36], ce cadre pouvant être utilisé pour généraliser les marchés artificiels univariés au cas multivarié.

Ce processus multivarié génère des ruptures communes dans le processus de volatilité, et exhibe le même type de dépendance que celle observée dans les séries financières multivariées ; voir Teyssière [36]. Notons que [18] suggèrent que des ruptures communes peuvent expliquer la persistance commune de la volatilité du prix des actifs financiers.

L'hypothèse principale de ces modèles est que les agents sont hétérogènes, ils diffèrent par leur fonction de prévision du prix des actifs financiers, mais ont un comportement moutonnier, car ils tendent à suivre l'“opinion du marché”, c-à-d, la prévision prédominante du prix des actifs.

Depuis Bachelier, l'hypothèse classique est que les marchés financiers sont “efficients”, c-à-d, conditionnellement à I_t , l'ensemble d'information à la date t , la meilleure prévision du prix des actifs est :

$$E(P_{t+1}|I_t) = P_t. \quad (37)$$

De plus, les prix sont supposés indépendants. Si la première affirmation, equation (37) est acceptable, la seconde affirmation sur l'indépendance est fautive : Mandelbrot [31] donne de nombreux exemples d'opérateurs de marchés financiers utilisant cette dépendance pour élaborer leurs stratégies. Notons que prédire les prix a peu de sens, alors que prédire leur volatilité a un sens ; voir à ce sujet Mandelbrot [31].

Nous considérons un marché avec deux sortes d'opérateurs, chartistes et fondamentalistes, qui diffèrent par leur fonction de prévision sur le taux de change futur. Cette fonction de prévision de l'agent i sur le taux de change futur j , $j = 1, \dots, m$, est une fonction de l'ensemble d'information I_t , et est notée par $E^i(P_{j,t+1}|I_t)$. Les chartistes extrapolent le prix $P_{j,t+1}$ en utilisant une fonction linéaire des prix précédents :

$$E^c(P_{j,t+1}|I_t) = \sum_{l=0}^{M^c} h_{j,l} P_{j,t-l}, \quad j = 1, \dots, m, \quad (38)$$

où $h_{j,l}$, $l = 0, \dots, M^c$ sont des constantes, M^c est la “mémoire” des chartistes, tandis que les fondamentalistes prévoient que le prix suivant diffère peu d'une série de prix “fondamentaux” $\bar{P}_{j,t}$:

$$E^f(P_{j,t+1}|I_t) = \bar{P}_{j,t} + \sum_{l=1}^{M^f} \nu_{j,l} (P_{j,t-l+1} - \bar{P}_{j,t-l}), \quad j = 1, \dots, m, \quad (39)$$

où $\nu_{j,l}$, $l = 1, \dots, M^f$ sont des constantes positives représentant le degré de retour vers les “fondamentaux”, M^f est la “mémoire” des fondamentalistes. Nous supposons que la série des “fondamentaux” $\bar{P}_{j,t}$, qui peuvent être interprétés comme le prix si ils étaient expliqués seulement par un ensemble de variables pertinentes, suit une marche aléatoire :

$$\bar{P}_{j,t} = \bar{P}_{j,t-1} + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \sim N(0, \sigma_{\varepsilon_j}^2), \quad j = 1, \dots, m. \quad (40)$$

Les agents ont la possibilité entre investir sur le marché domestique dans un actif sans risque et investir sur les marchés étrangers dans des actifs risqués. Nous notons par $\rho_{j,t}$ le taux d'intérêt étranger, par $d_{j,t}^i$

la demande de monnaie étrangère j , par l'agent i , à la date t , et par r le taux d'intérêt domestique, avec $\rho_{j,t} > r$. Le taux de change $P_{j,t}$ et le taux d'intérêt étranger $\rho_{j,t}$ sont considérés par les agents comme des variables aléatoires indépendantes, avec $E\rho_{j,t} = \rho_j$ et $\text{Var}(\rho_{j,t}) = \sigma_{\rho_j}^2$. Nous notons par $W_{j,t+1}^i$ la richesse cumulée de l'individu i à la date $t+1$, avec l'actif j :

$$W_{j,t+1}^i = (1 + \rho_{j,t+1})P_{j,t+1}d_{j,t}^i + (W_{j,t}^i - P_{j,t}d_{j,t}^i)(1 + r), \quad j = 1, \dots, m. \quad (41)$$

Les agents i ont une fonction d'utilité standard du type moyenne-variance :

$$U(W_{j,t+1}^i) = E(W_{j,t+1}^i) - \lambda_j \text{Var}(W_{j,t+1}^i), \quad (42)$$

où λ_j est le coefficient d'aversion au risque :

$$E(W_{j,t+1}^i | I_t) = (1 + \rho_j)E^i(P_{j,t+1} | I_t)d_{j,t}^i + (W_{j,t}^i - P_{j,t}d_{j,t}^i)(1 + r), \quad (43)$$

$$\text{Var}(W_{j,t+1}^i | I_t) = (d_{j,t}^i)^2 \zeta_{j,t}, \quad \zeta_{j,t} = \text{Var}(P_{j,t+1}(1 + \rho_{j,t+1})), \quad (44)$$

$E^i(\cdot | I_t)$ étant la prévision d'un agent de type i .

Soit k_t la proportion de fundamentalistes à la date t . Nous supposons que le processus des opinions $\{k_t\}$ est le même pour les deux marchés, car des changements structurels d'opinion n'affectent pas un marché particulier : les mêmes changements d'opinion de chartiste à fondamentaliste, et vice-versa, affectent des marchés corrélés, par exemple les marchés des changes. Les opérateurs des marchés financiers travaillent sur plusieurs taux de change, et un fondamentaliste (resp. chartiste) sur un marché agira comme fondamentaliste (resp. chartiste) sur les autres marchés. Donc, $k_t E^f(P_{j,t+1} | I_t) + (1 - k_t) E^c(P_{j,t+1} | I_t)$ est la prévision pour le marché j .

La demande $d_{j,t}^i$ pour le marché j est obtenue en maximisant la fonction d'utilité espérée. Les conditions du premier ordre pour chaque marché j nous donnent :

$$d_{j,t} = \frac{(1 + \rho_j) (k_t E^f(P_{j,t+1} | I_t) + (1 - k_t) E^c(P_{j,t+1} | I_t)) - (1 + r) P_{j,t}}{2\zeta_{j,t} \lambda_j}. \quad (45)$$

Notons par $X_{j,t}$ l'offre exogène de monnaie étrangère, alors la condition d'équilibre pour chaque marché j requiert que l'offre et la demandent s'équilibrent, c-à-d $X_{j,t} = d_{j,t}$, donc

$$P_{j,t} = \frac{1 + \rho_j}{1 + r} (k_t E^f(P_{j,t+1} | I_t) + (1 - k_t) E^c(P_{j,t+1} | I_t)) - \frac{2\zeta_{j,t} \lambda_j X_{j,t}}{1 + r}. \quad (46)$$

D'après l'équation (46), la dynamique du processus des prix $\{P_t\}$ dépend de l'évolution du processus $\{k_t\}$, c-à-d la proportion de fundamentalistes, qui gouverne la transition entre les deux fonctions de prévision $E^f(P_{j,t+1} | I_t)$ et $E^c(P_{j,t+1} | I_t)$. Plusieurs mécanismes d'évolution du processus d'opinion $\{k_t\}$ ont été proposés dans la littérature : $\{k_t\}$ peut être ou bien le résultat d'une diffusion des opinions de type épidémique, ou bien le résultat d'un processus d'apprentissage fondé sur la comparaison de la précision des fonctions de prévision en utilisant la U -statistique de Theil (voir [36]), ou bien une décision fondée sur la comparaison entre les profits obtenus avec chaque fonction de prévision, etc.

Si nous supposons que $m = 2$, c-à-d, les agents sont actifs sur deux marchés des changes, le processus bivarié des taux de change dépend de deux taux d'intérêt étrangers (ρ_1, ρ_2) . Nous supposons que $2\zeta_{j,t} \lambda_j X_{j,t} / (1 + \rho_j) = \gamma_j \bar{P}_{j,t}$ pour $j = 1, 2$, et $M^f = M^c = 1$, alors le prix d'équilibre pour le modèle bivarié est donné par :

$$\begin{pmatrix} P_{1,t} \\ P_{2,t} \end{pmatrix} = \begin{pmatrix} \frac{k_t - \gamma_1}{A_1} \bar{P}_{1,t} - \frac{k_t \nu_{1,1}}{A_1} \bar{P}_{1,t-1} + \frac{(1 - k_t) h_{1,1}}{A_1} P_{1,t-1} \\ \frac{k_t - \gamma_2}{A_2} \bar{P}_{2,t} - \frac{k_t \nu_{2,1}}{A_2} \bar{P}_{2,t-1} + \frac{(1 - k_t) h_{2,1}}{A_2} P_{2,t-1} \end{pmatrix}, \quad (47)$$

avec

$$A_i = \frac{1+r}{1+\rho_j} - (1-k_t)h_{j,0} - k_t\nu_{j,1}. \quad (48)$$

Nous supposons que le processus bivarié des fondamentaux $(\bar{P}_{1,t}, \bar{P}_{2,t})$ est positivement corrélé comme suit :

$$\begin{pmatrix} \bar{P}_{1,t} \\ \bar{P}_{2,t} \end{pmatrix} = \begin{pmatrix} \bar{P}_{1,t-1} \\ \bar{P}_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} \right], \quad (49)$$

avec $\sigma_{1,2} > 0$. Dans l'exemple considéré ici, nous choisissons $\sigma_{1,2}$ de telle façon que le coefficient de corrélation entre les deux processus $\varepsilon_{1,t}$ et $\varepsilon_{2,t}$ est égal à 0.75, un choix motivé par des résultats empiriques ; voir [37, 36].

Les paramètres des deux processus sont fixés de la façon suivante : les trois taux d'intérêt annuels sont égaux à $\rho_1 = 0.07$, $\rho_2 = 0.08$, $r = 0.04$, $\gamma_j = 0.90 - (1+r)/(1+\rho_j)$, $j = 1, 2$, $h_{1,0} = 0.635$, $h_{2,0} = 0.65$, $h_{j,1} = 1 - h_{j,0}$, $j = 1, 2$, $\nu_{1,1} = 0.60$, $\nu_{2,1} = 0.55$. γ_j est choisi de telle façon que le processus des prix est plus volatile quand l'opinion des chartistes est prédominante, donc les fundamentalistes stabilisent le marché. Avec un autre choix des paramètres, par exemple $\gamma_j = 1 - (1+r)/(1+\rho_j)$, nous obtenons le résultat inverse, c-à-d, les chartistes stabilisent le marché. Nous considérons le choix précédent car il apparaît le plus vraisemblable.

Nous simulons la série bivariée $(P_{1,t}, P_{2,t})'$ et détectons les ruptures dans la matrice de covariance Σ de la série bivariée de rendements $(r_{1,t}, r_{2,t})'$, avec $r_{j,t} = 100 \times \log(P_{j,t}/P_{j,t-1})$, $j = 1, 2$. Avec la méthode adaptative nous trouvons la segmentation $\hat{\tau} = \{409, 1016, 2115, 2657, 3431\}$, illustrée par la figure 7 ci-dessous, tandis qu'avec le critère de Schwarz pour des observations i.i.d., équation (12), nous obtenons la segmentation suivante $\hat{\tau} = \{112, 169, 463, 1016, 1947, 2115, 2656, 2727, 2830, 3036, 3431\}$. Notons que nous obtenons la segmentation avec la méthode adaptative sans avoir recours à une grille. L'utilisation d'une grille, avec $d = 10$, donne $\hat{\tau} = \{410, 1020, 2120, 2660, 3440\}$, ce qui n'est pas une approximation trop mauvaise.

La figure 7 montre que la procédure de détection adaptative est capable de capturer les changements de volatilité générés par les variations du processus d'opinions $\{k_t\}$ gouvernant l'évolution de la volatilité du processus bivarié des prix.

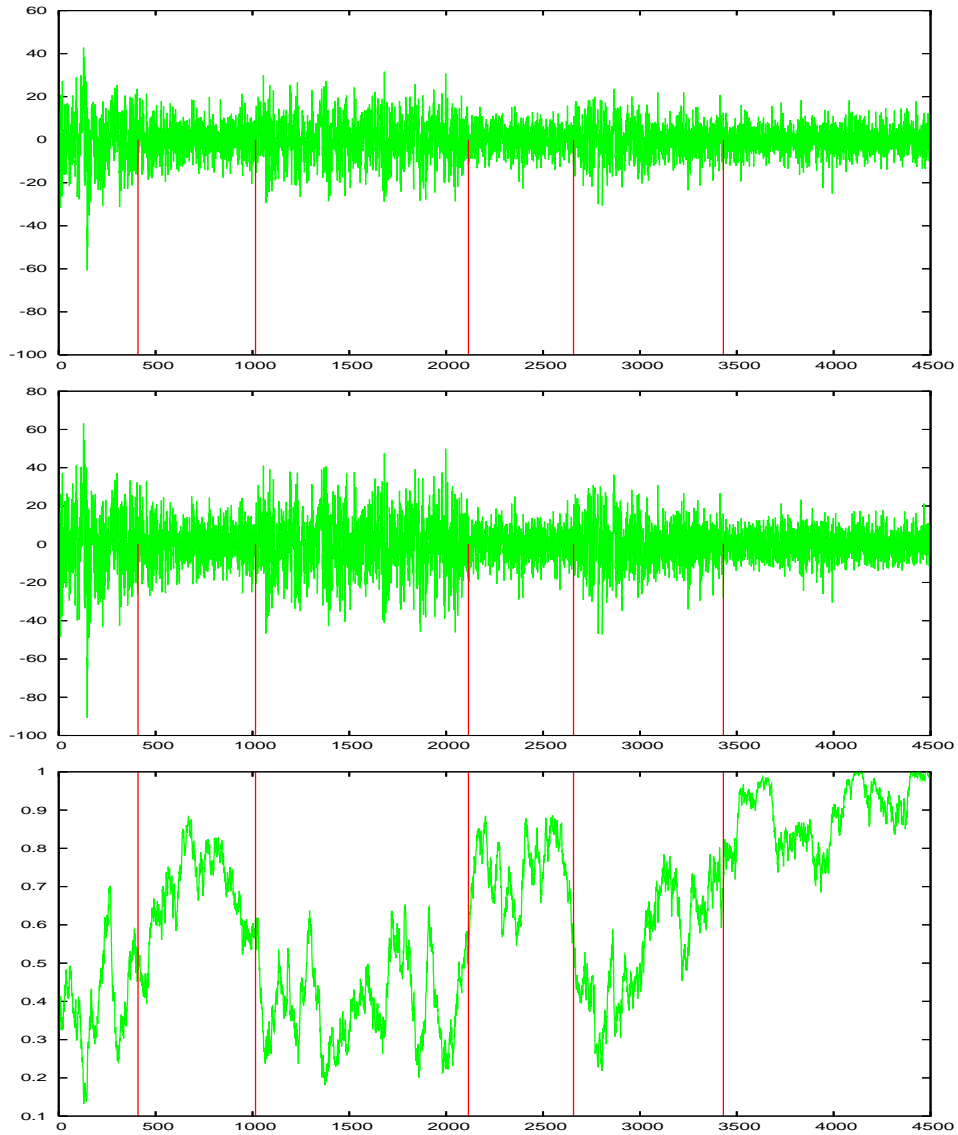


Figure 7: Haut : la série $r_{1,t}$ avec les ruptures estimées ; Milieu : la série $r_{2,t}$ avec les ruptures estimées ; Bas : la série k_t avec les ruptures estimées.

Remerciements : Nous remercions le rapporteur et l'éditeur, Alfredas Račkauskas, pour leur commentaires qui nous ont aidé à améliorer la présentation. Nous sommes reconnaissants à Alfredas Račkauskas de nous avoir proposé de publier cette traduction française.

References

- [1] Basseville, M. et Nikiforov, N. (1993). *The Detection of Abrupt Changes - Theory and Applications*. Prentice-Hall: Information and System Sciences Series.

- [2] Berkes, I., Gombay, E., Horváth, L. et Kokoszka, P.S. (2004). Sequential change–point detection in GARCH(p, q) models. *Econometric Theory*, **20**, 1140–1167.
- [3] Birgé, L. et Massart, P. (2001) Gaussian model selection. *J. Eur. Math. Soc.* **3** 203–268.
- [4] Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics*, **72**, 498–505.
- [5] Boussama, F. (1998). Ergodicité, mélange, et estimation dans les modèles GARCH. Thèse de doctorat, Université Paris 7.
- [6] Braun, J.V., Braun, R.K. et Muller, H.G. (2000). Multiple changepoint fitting via quasilielihood, with application to DNA sequence segmentation. *Biometrika*, **87**, 301–314.
- [7] Brodsky, B. et Darkhovsky, B. (1993). *Nonparametric Methods in Change–Point Problems*. Kluwer Academic Publishers, The Netherlands.
- [8] Chen, J. et Gupta, A.K. (2000). *Parametric Statistical Change Point Analysis*. Birkhäuser Verlag, Basel.
- [9] Chen, J. et Gupta, A.K. (2004). Statistical inference of covariance change points in Gaussian models. *Statistics*, **38**, 17–28.
- [10] Chu, C.-S.J. (1995). Detecting parameter shift in GARCH models. *Econometric Reviews*, **14**, 241–266.
- [11] Csörgö, M. et Horváth, L. (1997). *Limit Theorems in Change–Point Analysis*. Wiley.
- [12] Dalhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25**, 1–37.
- [13] Doukhan, P., Teyssière, G. et Winant, P. (2006). A LARCH(∞) vector valued process, dans : Bertail, P., Doukhan, P. et Soulier, Ph. (Editeurs.), *Dependence in Probability and Statistics*. Lecture Notes in Statistics, Vol. 187. Springer Verlag, pp. 245–258.
- [14] Giraitis, L., Leipus, R. et Surgailis, D. (2005). Recent advances in ARCH modelling, dans : Teyssière, G. et Kirman, A., (Editeurs.), *Long–Memory in Economics*, 3–38. Springer Verlag, Berlin.
- [15] Giraitis, L. et Leipus, R. (1992). Testing and estimating in the change–point problem of the spectral function. *Lithuanian Mathematical Journal*, **32**, 20–38.
- [16] Giraitis, L. et Leipus, R. (1990). Functional CLT for nonparametric estimates of the spectrum and change–point problem for a spectral function. *Lithuanian Mathematical Journal*, **30**, 674–679.
- [17] Giraitis, L., Leipus, R. et Surgailis, D. (1996). The change–point problem for dependent observations. *Journal of Statistical Planning and Inference*, **53**, 297–310.
- [18] Granger, C.W.J. et Hyung, N. (1999). Occasional structural breaks and long–memory. *Prépublication*.
- [19] Hawkins, D.M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, **72**, 180–186.
- [20] Hawkins, D.M. (2001). Fitting multiple change–point models to data. *Computational Statistics and Data Analysis*, **37**, 323–341.
- [21] Horváth, L., Kokoszka, P.S. et Teyssière, G. (2001). Empirical process of the squared residuals of an ARCH sequence. *The Annals of Statistics*, **29**, 445–469.

- [22] Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall: Signal Processing Series.
- [23] Kokoszka, P.S. et Leipus, R. (2003). Detection and estimation of changes in regime, dans : P. Doukhan, G. Oppenheim et M.S. Taqqu (Éditeurs.), *Long-Range Dependence: Theory and Applications*, 325–337. Birkhauser.
- [24] Kokoszka, P.S. et Leipus, R. (2000). Change–point estimation in ARCH models. *Bernoulli*, **6**, 513–539.
- [25] Kokoszka, P.S. et Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal*, **39**, 231–247.
- [26] Kokoszka, P.S. et Teyssière, G. (2002). Change–point detection in GARCH models: Asymptotic and bootstrap tests. *Prépublication*.
- [27] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, **83**, 79–102.
- [28] Lavielle, M. et Ludeña, C. (2005). Random thresholds for linear model selection. *Publication of the INRIA*, **5572**, <http://www.inria.fr/rrrt/rr-5572.html>.
- [29] Lavielle, M. et Moulines, E. (2000). Least–squares estimation of a unknown number of shifts in a time series. *Journal of Time Series Analysis*, **21**, 33–59.
- [30] Lavielle, M. et Teyssière, G. (2005). Adaptive detection of multiple change–points in asset price volatility, dans : G. Teyssière et A. Kirman (Éditeurs.), *Long-Memory in Economics*, 129–156. Springer Verlag, Berlin.
- [31] Mandelbrot, B.B. et Hudson R.L. (2004). *The Misbehavior of Markets: A Fractal View of Risk, Ruin, and Reward*. Basic Books, New-York.
- [32] Mia, B.Q. et Zhao, L.C. (1988). Detection of change points using rank methods. *Communications in Statistics - Theory and Methods*, **17**, 3207–3217.
- [33] Mikosch, T. et Stărică, C. (2003) Long–range dependence effects and ARCH modeling, dans : Doukhan, P., Oppenheim, G. et Taqqu, M.S. (Éditeurs.), *Long-Range Dependence: Theory and Applications*. Birkhauser, Boston, pp. 439–459.
- [34] Schechtman, E. et Wolfe, D.A. (1985). Multiple change points problem - nonparametric procedures for estimation of the points of change. *Communications in Statistics - Simulation and Computation*, **14**, 615–631.
- [35] Sen, A. et Srivastava, M.S. (1975). On tests for detecting change in the mean. *The Annals of Statistics*, **3**, 96–103.
- [36] Teyssière, G. (2003). Interaction models for common long–range dependence in asset price volatility, dans : Rangarajan, G. et Ding, M. (Éditeurs.), *Processes with Long Range Correlations: Theory and Applications*, Lecture Notes in Physics, Vol. 621. Springer Verlag, Berlin, pp. 251–269.
- [37] Teyssière, G. (1997). Modelling exchange rates volatility with multivariate long–memory ARCH processes. *Preprint*.
- [38] Teyssière, G. et Abry, P. (2005). Wavelet analysis of nonlinear long–range dependent processes. Applications to financial time series, dans : G. Teyssière et A. Kirman (Éditeurs.), *Long Memory in Economics*, 173–238. Springer Verlag.

- [39] Vostrikova, L.Ju. (1981). Detection of “disorder” in multidimensional random processes. *Soviet Mathematics Doklady*, **24**, 55–59.
- [40] Yao, Y.C. (1988). Estimating the number of change–points via Schwarz criterion. *Statistics & Probability Letters*, **6**, 181–189.