

# Sur l'utilisation conjointe de la régression sur composantes principales et des ondelettes\*

Salwa BENAMMOU, Nabihha HAOUAS & Zied KACEM

Computational Mathematics Laboratory  
&  
Faculté de Droit et des Sciences Economiques et Politiques de Sousse, Tunisie

e-mail: saloua.benam mou@fdseps.rnu.tn

e-mail: n.haouas@yahoo.fr

e-mail: zied\_kacem2004@yahoo.fr

## RÉSUMÉ

La régression sur composantes principales (RCP) est une régression sur les facteurs d'une ACP préalablement effectuée sur des variables initialement corrélées. L'utilisation de l'ACP permet de remplacer les variables initiales, par des composantes principales qui conservent la quasi-totalité de l'information, et qui présentent l'avantage d'être non corrélées. Ces composantes, sont prises comme variables explicatives pour une régression linéaire multiple. La qualité de la modélisation par RCP reste affectée par l'existence de bruit dans les variables initiales. Nous proposons dans ce travail un débruitage des données par ondelettes (wavelets) permettant de séparer le signal du bruit sans perte d'information. Nous montrons, sur des données boursières françaises, que l'élimination du bruit sur les composantes principales par un seuillage doux à base d'ondelettes améliore la qualité d'ajustement du modèle de régression ainsi que la qualité des prévisions. Nous confirmons le résultat par simulation.

**Mots clés:** Ondelettes, Seuillage, RCP, ACP, débruitage, simulation

## ABSTRACT

The principal components regression (PCR) is an applied regression on PCA factors of a PCA beforehand carried out on initially strongly correlated variables. The use of the PCA allows replacing the initial variables, by principle components which preserve the quasi-total of information, and which present the advantage to be non-correlated. These components are taken as explanatory variables for a multiple linear regression. The PCR modeling quality remains affected by the existence of noise in the initial variables. In this work, we propose a denoising of the data by wavelets (thresholding) making possible separation of the signal from the noise without losing information. We show using French stock-exchange data, that the elimination of the noise on the main constituents by a soft thresholding based on wavelets improves the quality of adjustment of the regression model as well as forecast quality. We confirm result by simulation.

**Keywords:** Wavelets, thresholding, PCR, PCA, denoising, simulation

(\*)Nous tenons à remercier les Professeurs Pierre CAZES et Samir BENAMMOU, pour leurs aides et suggestions lors de la rédaction de ce travail.

# 1. Introduction

La régression sur composantes principales (RCP) est une méthode d'analyse des données. Elle étudie les relations entre un groupe de variables explicatives, et un autre de variables à expliquer. Elle est utilisée lorsque le nombre de variables est très grand. Elle se déroule en deux étapes : la première consiste à appliquer une ACP sur les variables initiales et trouver les composantes principales associées, la deuxième est une régression sur un nombre réduit de composantes retenues selon un des critères usuel de l'ACP.

Malgré les avantages de la RCP et ses résultats incontestables dans différents domaines d'applications, elle se trouve toujours confronté au problème d'existence de bruit dans les données initiales qui affecte la qualité d'ajustement du modèle.

La technique de débruitage par ondelettes (voir VIGNERON et al. 2003 et AMINGHAFARI et al. 2006) semble remédier à ce problème, et permet de séparer le signal du bruit par seuillage, sans perte d'information.

Nous proposons ici une approche combinant la technique de débruitage des données par ondelettes et la régression sur composantes Principales.

## 2. Rappel sur les ondelettes

### 2.1 Transformée de Fourier

○Ancêtre des ondelettes, la transformée de Fourier (TF) est un outil traditionnel de traitement du signal. Elle repose sur le fait que tout signal peut être décomposé en une suite de sinusoides de différentes fréquences, sous la forme (1).

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \quad (1)$$

Où

$f$  est un signal,

$\omega$  la fréquence du signal

et  $F$  la transformée de Fourier du signal  $f$ .

La transformation de Fourier fait ressortir le contenu fréquentiel de  $f$  mais ne permet pas de **localiser temporellement** les événements (saut, changement de fréquence, etc.).

Pour résoudre ce problème, GABOR (1964) a introduit la transformée de Fourier à court terme. Cette dernière se base sur le fenêtrage (Windowing). Il s'agit de segmenter en tranche de temps fixes le signal à analyser et d'appliquer par la suite la Transformée de Fourier (TF) à chaque tranche.

La suite logique pour la résolution de ce problème a été l'élaboration d'un puissant outil localisé en temps et en fréquences dénommé la transformée en ondelettes (wavelets transform).

## 2.2. La théorie des ondelettes

### 2.2.1. La transformée en ondelettes discrètes

La Transformée en Ondelette Discrète (TOD) dite encore Discret Wavelet Transform (DWT) est une technique utilisée dans la compression de données (voir par exemple MALLAT 1989). La décomposition sur différents espaces pour les basses et les hautes fréquences est réalisée par approximations successives de l'information initiale. Elle se fait du plus grossier au plus fin, moyennant des algorithmes efficaces de traitement du signal. Les espaces sur lesquels se fait la décomposition sont appelés espaces d'approximations et espaces de détails.

### 2.2.2. Espaces d'approximations

Les espaces d'approximations sont des sous espaces de l'espace de Hilbert  $L^2(\mathbb{R})$  (des fonctions d'une variable réelle de carrés intégrables). Une analyse au niveau de résolution  $j$ , ( $j \in \mathbb{Z}$ ) de la fonction  $f$  est obtenue en appliquant un opérateur linéaire  $A_j$  sur  $f$

$A_j f$  est alors un élément d'un sous espace  $V_j$  de  $L^2(\mathbb{R})$  et  $A_j$  un projecteur de  $L^2(\mathbb{R})$  sur  $V_j$ .

On construit alors une analyse multirésolution comme étant une suite de sous-espaces emboîtés  $V_j$ , de sortes que le passage d'un espace à l'autre, soit le résultat d'un changement d'échelle (zoom).

L'axiomatique correspondante peut s'exprimer sous forme de cinq conditions (voir par exemple TRUCHETET 1998) :

$(V_j)_{j \in \mathbb{Z}}$  une suite de sous espaces fermés de  $L^2(\mathbb{R})$  tels que  $V_{j+1} \subset V_j \forall j \in \mathbb{Z}$ , c-à-d :

$$\dots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \dots \quad (2)$$

$$\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}) \quad (\text{c-à-d l'adhérence de } \bigcup_{j \in \mathbb{Z}} V_j \text{ est } L^2(\mathbb{R})) \quad (3)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \quad (4)$$

$$\forall j \in \mathbb{Z}, f(x) \in V_j \Leftrightarrow f(2^{-1}x) \in V_{j+1} \text{ ce qui équivaut à } f(2^j x) \in V_0 \quad (5)$$

$$\forall k \in \mathbb{Z}, f(x) \in V_0 \Leftrightarrow f(x-k) \in V_0 \text{ (Invariance par translation).} \quad (6)$$

Il existe alors une fonction dite "fonction d'échelle" notée  $\varphi(x)$  qui engendre une base orthonormée de  $V_j$ . Les fonctions de base sont construites par dilatations et translations de la fonction  $\varphi(x)$  selon la relation (7):

$$\varphi_{j,n}(x) = 2^{\frac{-j}{2}} \varphi(2^{-j}x - n), n \in \mathbb{Z} \quad (7)$$

Les conditions (2) à (7) définissent alors une analyse multirésolution de  $L^2(\mathbb{R})$ . L'approximation au niveau de résolution  $j$  de la fonction  $f$  est définie par une suite notée  $(a_j^n)_{n \in \mathbb{Z}}$  de produits de convolutions :

$$a_n^j = \langle f, \Phi_{j,n} \rangle \quad (8)$$

### 2.2.1. Espaces de détails

Les espaces de détails sont des sous-espaces complétant les espaces d'approximations  $V_j$ . En reprenant par exemple les notations de TRUCHETET (1998) :

si  $\forall j \in \mathbb{Z}$ ,  $W_j$  est un supplémentaire orthogonal de  $V_j$  dans  $V_{j-1}$ , on a alors (par définition)

$$V_{j-1} = V_j \oplus W_j \quad (9)$$

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j \quad (10)$$

$W_{j-1}$  étant orthogonal à  $V_{j-1}$ , la relation (10) implique que  $W_{j-1}$  est aussi orthogonal à  $W_j$  ; d'où la relation (11)

$$\forall j, k \in \mathbb{Z} \quad k \neq j \text{ alors } W_j \perp W_k. \quad (11)$$

Dans ces conditions, on montre qu'il existe une fonction  $\psi$  de  $L^2(\mathbb{R})$  appelée ondelette qui (par dilatations et translations), engendre une base orthonormée de  $W_j$  et donc de  $L^2(\mathbb{R})$ .

Les fonctions de base sont définies par la relation (12):

$$\psi_{j,n}(x) = 2^{\frac{-j}{2}} \psi(2^{-j}x - n), \text{ avec } n \in \mathbb{Z} \quad (12)$$

### 2.3. Débruitage par ondelettes

Le débruitage est indispensable pour améliorer la qualité des signaux et pouvoir les exploiter au mieux ultérieurement.

Dans la littérature récente beaucoup de méthodes ont été développées dans le but de débruiter des signaux « contaminés ». Dans le domaine des ondelettes, ce terme signifie la rejection du bruit par un seuillage adéquat (voir DONOHO et al., 1994).

Nous donnons ici l'algorithme de base de débruitage proposé par DONOHO (1994), qui consiste à débruiter un signal contaminé par un bruit gaussien, centré et de variance  $\sigma^2$  connue.

L'algorithme est composé de trois étapes :

ETAPE 1 : Décomposition du signal, centré par ondelettes.

ETAPE 2 : Seuillage des coefficients de détails par un seuil de DONOHO.

ETAPE 3 : Reconstruction du signal par l'application inverse de la transformée en ondelettes.

DONOHO suppose que le bruit contaminant un signal se concentre dans les détails. Il présente aussi deux types de seuillage de coefficients d'ondelettes appelés : seuillage doux "SOFT thresholding" et seuillage dur "HARD thresholding" illustrés par la figure n°1.

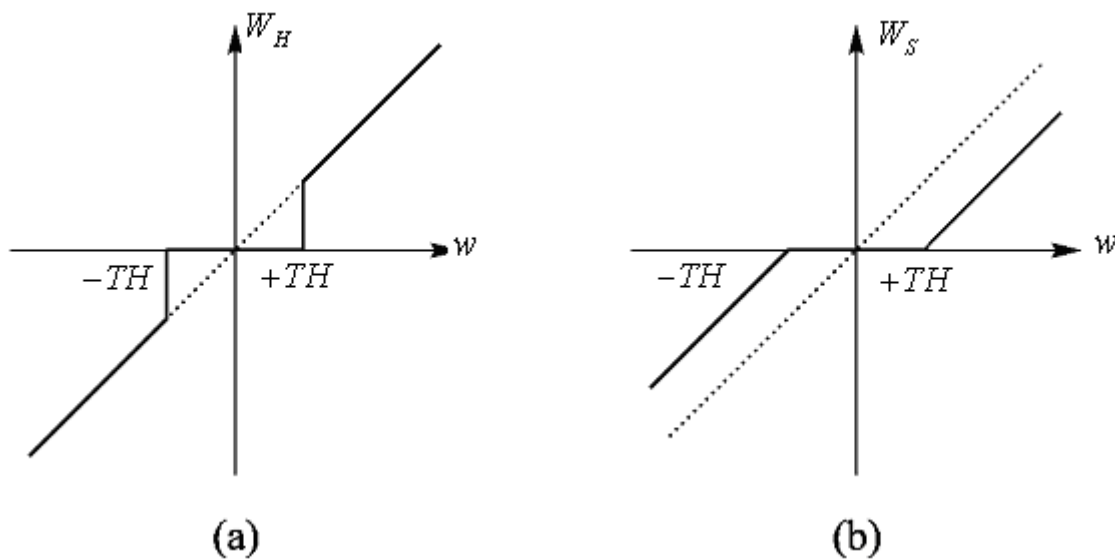


Figure 1  
Seuillage de DONOHO (au seuil TH) : (a) Hard, (b) Soft

Ces deux types de seuillages sont donnés par :

$$W_H = \begin{cases} W & \text{si } |W| > TH \\ 0 & \text{si } |W| \leq TH \end{cases} \quad \text{et} \quad W_S = \begin{cases} W + TH & \text{si } W \leq -TH \\ 0 & \text{si } |W| \leq TH \\ W - TH & \text{si } W \geq TH \end{cases}$$

Où :

TH: Le seuil (threshold).

W : Le coefficient de détail du signal bruité.

$W_S$  : Le coefficient résultant du seuillage doux (SOFT)

$W_H$  : Le coefficient résultant du seuillage dur (HARD).

Lorsque l'on compare ces deux seuillages, le dur tend à avoir une variance plus importante que le doux, alors que ce dernier présente un biais plus important.

### 3. L'approche RCP-Ondelettes

La Régression sur Composantes Principales (RCP) consiste dans un premier temps à effectuer une Analyse en Composantes Principales (ACP) sur les variables explicatives, généralement bruitées. On effectue ensuite une régression sur les composantes principales retenues.

La RCP permet de faciliter les procédures de calcul et donne de meilleurs résultats que les méthodes de régression classiques. Le résultat reste toujours influencé négativement par l'existence de bruit qui affecte toujours les données réelles.

La méthode que nous proposons ici fait appel à la théorie des ondelettes pour débruiter les données de la RCP. Il s'agit d'une utilisation conjointe de ces deux méthodes qui conduit à ce que nous appelons l'approche "RCP-Ondelettes".

Cette méthode "RCP-Ondelettes" est composée de cinq étapes:

ETAPE 1: S'assurer que la matrice des variables explicatives  $X_i$ , est de longueur dyadique ( $2^n$ ) afin de pouvoir appliquer la transformée en ondelettes discrète (TOD). Dans le cas contraire, compléter la matrice par des zéros (zero padding).

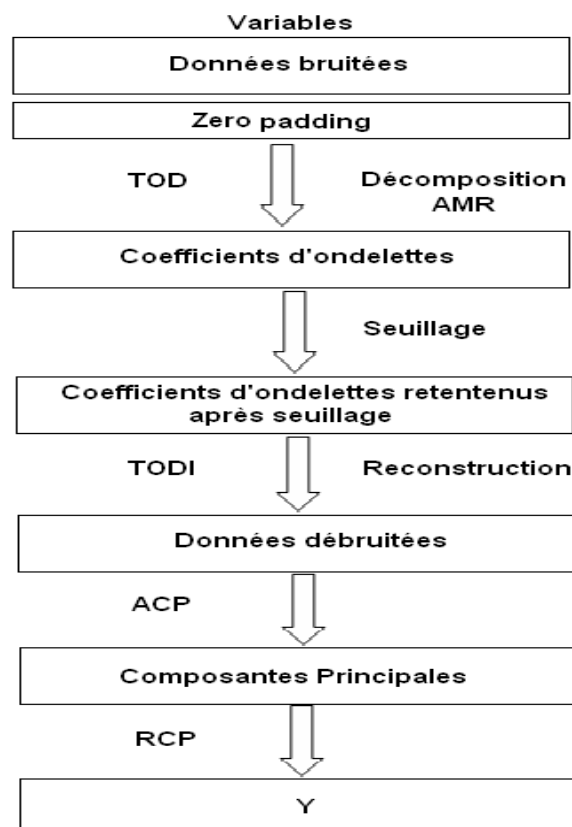
ETAPE 2: Appliquer l'analyse multirésolution (AMR) pour décomposer les  $X_i$  en plusieurs niveaux de résolution. Cette décomposition peut se faire en choisissant une ondelette adéquate selon la nature du signal et sa forme (à titre d'exemple on pourra choisir une ondelette de Haar, de Daubechies, etc.), (voir Daubechies 1988, 1992).

ETAPE 3: Appliquer un seuillage sur les  $X_i$  décomposés afin d'éliminer le bruit. Le choix du type de seuillage et la sélection de la valeur du seuil à utiliser est crucial. Il est intimement lié à la nature du signal à débruiter (régularité, discontinuités,...) et à l'ampleur du bruit en question.

ETAPE 4: Effectuer une transformée en ondelettes discrète inverse (TODI) pour reconstituer les variables débruitées.

ETAPE 5: Effectuer une ACP sur données débruitées et trouver les composantes principales qui vont servir comme données pour la RCP.

Ces étapes (de l'approche "RCP-Ondelettes") peuvent être schématisées par la figure n°2.



**Figure 2**  
*L'approche RCP-Ondelettes*

## 4. Application sur des données boursières

Pour illustrer notre approche, nous utilisons des données boursières réelles. Il s'agit des rentabilités journalières de 37 titres cotés à la Bourse de Paris (voir tableau n°1) composant l'indice CAC 40 à la date du 31 août 2007<sup>(\*)</sup>, sur la période allant de janvier 2004 à janvier 2008 (soit un nombre dyadique de 1024 observations).

**Tableau 1**  
*Liste des variables*

Variabes	Compagnies
V1	Accor
V2	Air liquide
V3	Alcatel
V4	Arcelor
V5	Axa
V6	Bnp Paribas
V7	GDF
V8	Bouygues
V9	Cap Gemini
V10	Carrefour
V11	Crédit Agricole
V12	Danone
V13	Dexia
V14	EADS
V15	Alstom
V16	Essilor International
V17	France Telecom
V18	EDF
V19	Lafarge
V20	Lagardère
V21	L'oréal
V22	LVMH
V23	Michelin
V24	Pernod Ricard
V25	Peugeot
V26	PPR(Pinault Printemps)
V27	Renault
V28	Saint Cobain
V29	Aventis
V30	Schneider Electric
V31	Société Générale
V32	STMicroelectronics
V33	Suez
V34	Total SA
V35	Veolia environnement
V36	Vinci
V37	Vivendi
V38	Unibail-rodamco
V39	Vallourec
V40	Air France

Nous supposons que les rentabilités boursières sont en partie expliquées par des facteurs communs indéterminés et qui sont de la forme donnée par l'équation (13):

---

(\*) Les trois sociétés Alstom, EDF et GDF du portefeuille CAC 40, sont exclues en raison de l'indisponibilité des données.

$$R_{it} = \sum_{j=1}^K \beta_j^{(i)} * F_{jt} + \xi_{it} \quad (13)$$

où :

$R_{it}$  est la rentabilité du titre  $i$  à la date  $t$ ,

$F_{jt}$  est le  $j^{\text{ème}}$  facteur,

$\beta_j^{(i)}$  est la sensibilité du titre  $i$  au  $j^{\text{ème}}$  facteur,

$\xi_{it}$  est le risque spécifique du titre  $i$  à la date  $t$ ,

et  $K$  est le nombre de facteurs.

**Remarque** Le domaine d'application traité dans cet exemple n'est pas classique, puisque nous n'avons que des variables à expliquer et pas de variables explicatives. L'ACP est faite sur le tableau des variables à expliquer, les premiers facteurs servant, par la suite, de variables explicatives pour la régression. Les variables explicatives sont donc des variables latentes fonctions des variables à expliquer. Nous ne sommes donc pas dans le cadre de la régression classique.

#### 4. 1. Résultats de la RCP

Les six premières valeurs propres de l'ACP sur les rentabilités sont données par le tableau n° 2. Nous retenons cinq facteurs selon le critère de Kaiser (élimination des facteurs associés à des valeurs propres plus petites que 1).

**Tableau 2**  
*Valeurs propres et inerties des six premiers facteurs (cas de la RCP)*

Facteur	Valeur propre	Inertie expliquée	Inertie Cumulée
1	14.143	38.225%	38.225%
2	1.285	3.473%	41.697%
3	1.188	3.212%	44.909%
4	1.100	2.973%	47.883%
5	1.031	2.786%	50.668%
6	0.970	2.620%	53.290%

Nous effectuons une régression linéaire multiple sur les cinq facteurs retenus.

La rentabilité du titre BNP PARIBAS par exemple (variable n°6, tableau 1) est donnée par la relation (13)

$$\hat{R}_{6t} = \hat{\beta}_0^{(6)} + \hat{\beta}_1^{(6)} F_{1t} + \hat{\beta}_2^{(6)} F_{2t} + \hat{\beta}_3^{(6)} F_{3t} + \hat{\beta}_4^{(6)} F_{4t} + \hat{\beta}_5^{(6)} F_{5t}. \quad (14)$$

La valeur du  $R^2$  de ce modèle est égale à 0,7193 ce qui reflète une bonne qualité d'ajustement (voir tableau 3). Presque 72% de la variabilité de la variable dépendante est donc expliquée par le modèle.



La valeur de la statistique de Fisher (521,71) nous permet de rejeter l'hypothèse de nullité de tous les paramètres, le modèle est donc globalement significatif.

**Tableau 3**  
*Critères d'évaluation de la régression (cas de la RCP)*

MCE	0.00701
moyenne	0.00042420
Coeff Var	1651.81905
R <sup>2</sup>	0.7193
R <sup>2</sup> Ajusté	0.7179
F	521.71
Pr > F	<.0001

Les p-valus (voir tableau 4) associées aux estimateurs des paramètres  $\beta_1, \beta_2, \beta_4$  et  $\beta_5$  sont inférieures à 0,05. Nous rejetons alors l'hypothèse de nullité de chacun de ces coefficients de la régression.

**Tableau 4**  
*Paramètres de l'estimation (cas de la RCP)*

Variable	Coefficients estimés	Ecart type	t	Pr >  t
constante	0.00042420	0.00021897	1.94	0.0530
F1	0.00285	0.00005825	48.95	<.0001
F2	-0.00060243	0.00019327	-3.12	0.0019
F3	0.00002232	0.00020096	0.11	0.9116
F4	-0.00271	0.00020887	-12.97	<.0001
F5	0.00127	0.00021578	5.90	<.0001

## 4. 2. Résultats de l'approche RCP-ondelettes

Nous effectuons une transformée en ondelettes discrète des variables explicatives. Un seuillage doux appliqué aux coefficients obtenus nous permet d'éliminer le bruit existant au sein de ces variables. Le choix du seuil se fait selon la procédure Minimax (voir par exemple GENCAY et al 2002), qui consiste à choisir un estimateur  $\hat{x}$  de  $x$  qui atteint le risque défini par la relation (15)

$$\tilde{\mathfrak{R}}(\mathcal{F}) = \inf_x \sup_x \mathfrak{R}(\hat{x}, x) \quad (15)$$

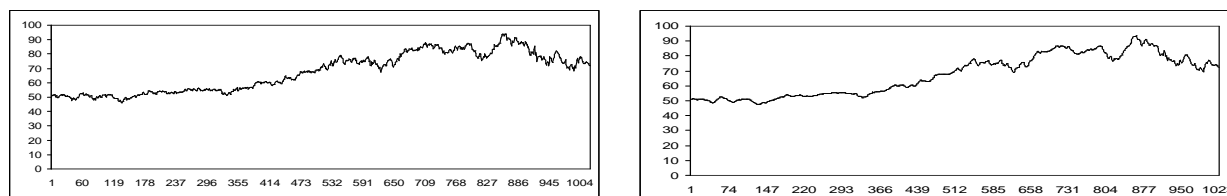
$\mathcal{F}$  étant une classe particulière de fonctions régulières et  $\mathfrak{R}(\hat{x}, x) = E_x l(\hat{x} - x)$  où  $l$  est une fonction de coût paire et convexe ( $\mathfrak{R}(\hat{x}, x)$  pouvant être pris, par exemple, comme l'erreur quadratique moyenne, voir DONOHO 1999)).

Les valeurs des seuils minimax sont calculées par approximation numérique. Les variables brutes présentant des discontinuités et des changements abrupts. La procédure Minimax est bien adaptée. Elle permet de garder les discontinuités et opère un seuillage portant sur la moyenne globale des carrés des erreurs et d'épargner les changements abrupts dans le signal.

L'association seuillage doux et critère Minimax définit ce qu'on appelle « Risk-Shrink » (voir Donoho & Johnstone 1994, 1998).

Nous appliquons, par la suite, une transformée en ondelettes discrète inverse pour reconstituer des variables débruitées.

L'exemple de la variable 'BNP PARIBAS' nous permet de bien visualiser la différence entre le signal avant et après débruitage (figure 3).



**Figure 3**  
*Signal avant débruitage (droite) et après débruitage (gauche). Du titre BNP PARIBAS.*

Le tableau n°5 donne les dix premières valeurs propres de l'ACP sur les données débruitées. En appliquant le critère de Kaiser on garde 9 facteurs.

**Tableau 5**  
*Valeur propre et variance expliquée par les facteurs (cas de l'approche RCP-Ondelettes)*

Facteur	Valeur propre	Inertie expliquée	Inertie Cumulée
1	10.9150017	29.500%	29.50%
2	1.7844851	4.823%	34.32%
3	1.4589165	3.943%	38.27%
4	1.3388564	3.619%	41.88%
5	1.2460272	3.368%	45.25%
6	1.1573831	3.128%	48.38%
7	1.1020222	2.978%	51.36%
8	1.0675986	2.885%	54.24%
9	0.9981457	2.700%	56.94%
10	0.9502233	2.570%	59.51%

Il est à noter que ces neuf facteurs expliquent 59.51% de l'information initiale. L'identification des facteurs (qui n'entre pas dans les objectifs de ce travail), peut se faire aisément à ce niveau.

Pour illustrer le résultat obtenu par la procédure RCP après débruitage, nous reprenons l'exemple de la rentabilité du titre 'BNP PARIBAS', qui est maintenant donnée par la relation (16).

$$\hat{R}_{6t} = \hat{\beta}_0^{(6)} + \hat{\beta}_1^{(6)}F_{1t} + \hat{\beta}_2^{(6)}F_{2t} + \hat{\beta}_3^{(6)}F_{3t} + \hat{\beta}_4^{(6)}F_{4t} + \hat{\beta}_5^{(6)}F_{5t} + \hat{\beta}_6^{(6)}F_{6t} + \hat{\beta}_7^{(6)}F_{7t} + \hat{\beta}_8^{(6)}F_{8t} + \hat{\beta}_9^{(6)}F_{9t}. \quad (16)$$

Le rejet de l'hypothèse nulle  $H_0 : \beta_1 = \beta_2 = \dots = \beta_9 = 0$  nous permet de dire que le modèle est globalement significatif au seuil de 5% (voir tableau n° 6).

**Tableau 6***Critères d'évaluation de la régression (cas de l'approche RCP-Ondelettes)*

MCE	0.00280
moyenne	0.00036157
Coeff Var	774.13695
R <sup>2</sup>	0.7763
R <sup>2</sup> Ajusté	0.7743
F	390.61
Pr > F	<.0001

Les p-valus associées aux estimateurs de tous les paramètres sont inférieures à 5%, donc nous rejetons l'hypothèse de nullité de chacun de ces coefficients de la régression (tableau n°7).

**Tableau 7***Paramètres de l'estimation (cas de l'approche RCP-Ondelettes)*

Variable	Coefficients estimés	Ecart type	t	Pr >  t
Constante	0.00036157	0.00008777	4.12	<.0001
Prin1	0.00132	0.00002658	49.62	<.0001
Prin2	-0.00094108	0.00006573	-14.32	<.0001
Prin3	0.00057449	0.00007270	7.90	<.0001
Prin4	-0.00189	0.00007589	-24.86	<.0001
Prin5	0.00025176	0.00007867	3.20	0.0014
Prin6	-0.00067808	0.00008162	-8.31	<.0001
Prin7	-0.00059313	0.00008365	-7.09	<.0001
Prin8	0.00028708	0.00008499	3.38	0.0008
Prin9	-0.00023033	0.00008764	-2.63	0.0087

### 4. 3. Comparaison de la RCP et de l'approche RCP-ondelettes

Nous remarquons une baisse de la corrélation entre les différentes variables suite au débruitage (voir annexe).

Le tableau n°8 donne les corrélations avant et après débruitage de quelques titres

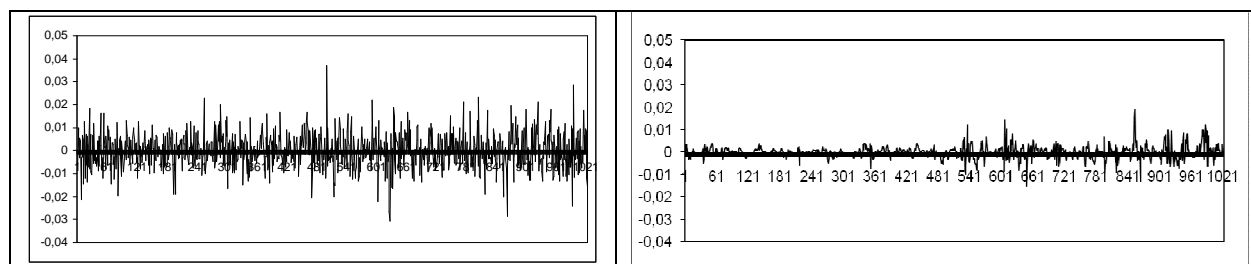
**Tableau 8***Corrélation avant et après débruitage entre quelques titres*

	Avant débruitage		Après débruitage	
	Axa	BNP	Axa	BNP
Société Générale	0.7298	0.8062	0.6456	0.7735
STM	0.5657	0.5079	0.4223	0.2906

L'ACP sur des données initiales (bruitées), a donné 5 facteurs expliquant 50.66% de l'inertie totale. Le nombre de facteurs de l'ACP sur des données débruitées est égal à 9 expliquant 56.94% de l'inertie totale.

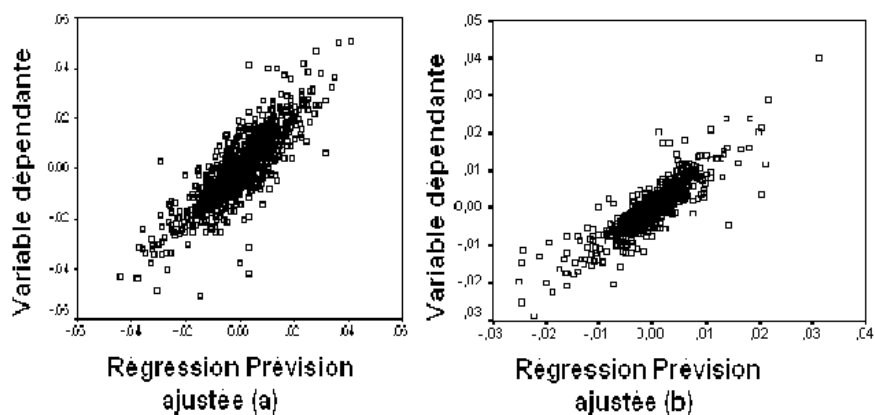
L'augmentation du nombre des facteurs après débruitage, est prévisible du fait que les coefficients de corrélation entre les variables ont baissé.

Nous observons une nette diminution des résidus de la régression après débruitage (voir figure n°4).



**Figure 4**  
*Résidu avant (droite) et après débruitage (gauche) du titre BNP PARISBAS.*

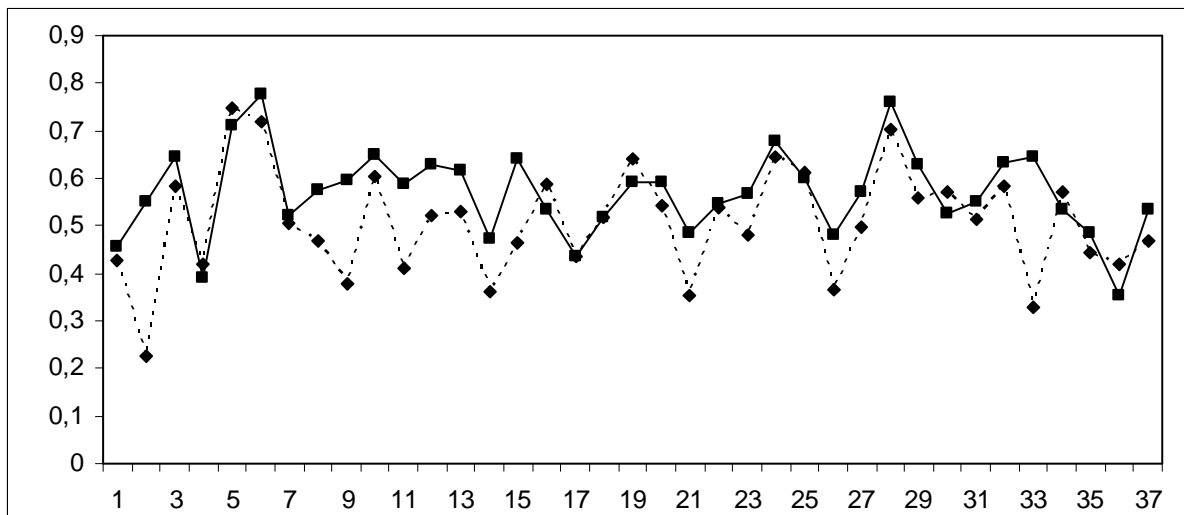
L'amélioration des résultats est aussi visible sur la valeur réelle du rendement en fonction du rendement estimé. Nous remarquons que le nuage des points est plus dispersé après débruitage (voir figure N° 5).



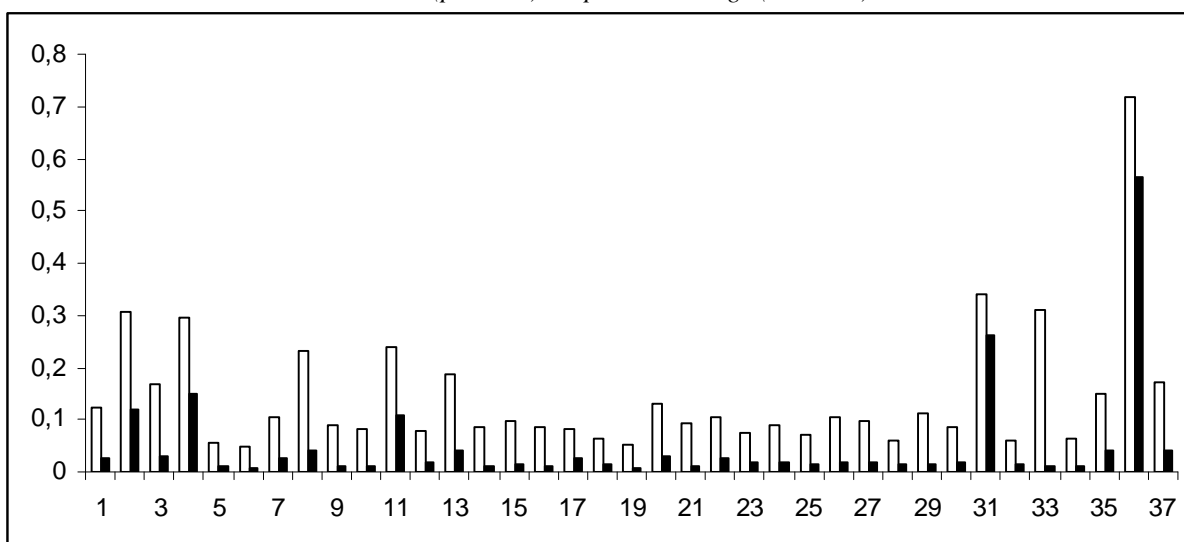
**Figure 5**  
*Ajustement de  $R_6$  par  $\hat{R}_6$  avant (a) et après débruitage (b)*

L'approche RCP-Ondelettes semble être meilleure que la RCP classique. En effet, il y a une augmentation du nombre de coefficients significatifs pour 67.56% des variables tandis que le modèle reste toujours globalement significatif.

Cette amélioration est prévisible du fait de l'amélioration de la qualité d'ajustement du modèle (voir figure n°6).



**Figure 6**  
*R<sup>2</sup> avant (pointillé) et après débruitage (continue)*



**Figure 7:**  
*SCR avant (blanc) et après débruitage (noir)*

La baisse de la somme des carrés résiduels issue de la méthode RCP-Ondelettes apparait clairement dans la figure n° 7.

## 5. Etude par simulation

D'un seul cas on ne peut tirer de généralités. Nous procédons donc par simulation. Nous opérons toujours dans le cas non classique où les variables explicatives sont des variables latentes fonctions des variables à expliquer.

Nous appliquons l'approche RCP-Ondelettes sur des réalisations d'un mouvement Brownien multidimensionnel dont les composantes sont bruitées

Nous rappelons ici qu'un mouvement brownien peut être défini de façon formelle comme un processus stochastique  $(B_t)_{(t \geq 0)}$  dont les accroissements disjoints sont indépendants et tels que la différence  $(B_{t+s} - B_t)$  suit une loi normale de moyenne nulle et de variance  $s$ .

Pour présenter ce processus, nous nous restreignons au cas  $n=2$ , la généralisation est immédiate.

Considérons le couple  $B_t = (B_t^{(1)}, B_t^{(2)})^T$  qui est un mouvement Brownien 2-dimensionnel corrélé si :

- Les incréments  $B_1(t) - B_1(s)$  et  $B_2(t) - B_2(s)$ ,  $t > s$  sont indépendants de  $B_1(y)$  et  $B_2(y)$ ,  $\forall y$  tel que  $0 \leq y \leq s$ .
- $\text{cov}(B_1(t), B_2(t)) = E(B_1(t)B_2(t)) = \rho t$  où  $-1 \leq \rho \leq 1$ . Ceci implique que :  $\text{corr}(B_1(t), B_2(t)) = \rho$ .

De plus, on a :  $\forall t \neq s$ ,  $\text{cov}(B_1(t), B_2(t)) = E(B_1(t)B_2(t)) = \rho t$  et  $\text{var}(B_1(t)) = \text{var}(B_2(t)) = t$

Dans notre cas, nous avons généré 55 réalisations d'un mouvement Brownien n-dimensionnel  $B_t = (B_t^{(1)}, B_t^{(2)}, \dots, B_t^{(n)})^T$ , avec  $n=11$  (composantes corrélées).

Le problème de construction et de simulation du mouvement Brownien multidimensionnel est traité par Glasserman (2003). Benammou (2003), Haugh (2004).

L'étude par simulation est bornée de trois phases :

**Dans une première phase**, nous effectuons une RCP sur les 55 réalisations du mouvement Brownien multidimensionnel. La matrice de corrélation associée est donnée par le tableau n°9

**Tableau 9**  
*Matrice des corrélations entre vecteurs simulés*

	B1	B2	B3	B4	B5	B7	B8	B9	B10	B11
B1	1	0,68	0,84	0,63	0,65	0,88	0,71	0,70	0,72	0,91
B2	0,68	1	0,65	0,91	0,69	0,63	0,71	0,62	0,62	0,85
B3	0,84	0,65	1	0,83	0,62	0,88	0,74	0,83	0,63	0,61
B4	0,63	0,91	0,83	1	0,89	0,99	0,79	0,93	0,62	0,51
B5	0,65	0,69	0,62	0,89	1	0,74	0,77	0,82	0,76	0,64
B7	0,88	0,63	0,88	0,99	0,74	1	0,92	0,88	0,65	0,68
B8	0,71	0,71	0,74	0,79	0,77	0,92	1	0,89	0,81	0,64
B9	0,70	0,62	0,83	0,93	0,82	0,88	0,89	1	0,67	0,96
B10	0,72	0,62	0,63	0,62	0,76	0,65	0,81	0,67	1	0,64
B11	0,91	0,85	0,61	0,51	0,64	0,68	0,64	0,96	0,64	1

**La seconde phase** consiste à ajouter une composante bruitée à chacune des réalisations simulées puis effectuer une RCP sur données bruitées.

La génération du bruit se base sur la notion du ratio « signal sur bruit » qu'on désigne généralement par SNR « signal to noise ratio ». C'est le rapport entre la grandeur d'un signal et celle d'un bruit.

Ce rapport est calculé par la formule (17).

$$SNR = 10 \log_{10} \left( \frac{P_S}{P_N} \right) \quad (17)$$

$P_S$  et  $P_N$  sont les amplitudes respectives du signal et du bruit.

Les valeurs du SNR utilisées (voir tableau 10) varient d'un mouvement Brownien unidimensionnel à un autre au sein du processus multidimensionnel simulé et sont donc au nombre de 11,

Le vecteur SNR que nous avons utilisé est donné par le tableau 10

**Tableau 10**  
*Valeurs du ratio "Signal to noise"*

SNR	10	0.1	12	5	3	0.12	1.2	3.12	0.55	5	1.9
-----	----	-----	----	---	---	------	-----	------	------	---	-----

L'utilisation de valeurs différentes du SNR permettra de vérifier si la méthode de débruitage utilisée est robuste à une variation de l'intensité du bruit contaminant le signal.

La troisième phase consiste à appliquer la méthode RCP-Ondelettes aux 55 matrices bruitées et ceci afin de tester la validité de l'approche proposée dans un environnement bruitée.

Les résultats de la régression sont donnés dans le tableau n°11.

**Tableau 11**  
*Résultats de la RCP sur données simulées*

<b>R<sup>2</sup></b>			<b>SCR</b>		
initial	bruité	débruité	initial	bruité	débruité
0,9519	0,51	0,95	1.21394	25.87464	1.12291
0,7868	0,4639	0,7245	1.59524	21.97909	2.49955
0,8407	0,4761	0,797	3.4309	27.86286	4.6634
0,8597	0,6162	0,902	1.49511	19.24229	1.75334
0,8464	0,2231	0,7698	0.86283	28.59815	1.49337
0,9036	0,5994	0,893	2.63088	21.75239	2.77342
0,8248	0,4705	0,8068	4.21056	30.4824	4.75668
0,9125	0,5048	0,8943	2.53171	28.12447	2.7255
0,9119	0,5104	0,8646	1.31689	22.82679	2.13886
0,9107	0,5099	0,912	1.91476	25.81869	1.86378
0,8088	0,4139	0,8235	6.40808	40.31294	6.15128
0,8103	0,2669	0,7298	0.789	27.16662	1.58068
0,8033	0,6266	0,8176	1.71154	14.61894	1.52202
0,6935	0,2224	0,6306	2.40033	30.46717	3.22539
0,8016	0,3824	0,7784	1.7355	27.16556	2.27182
0,6835	0,2377	0,9454	1.18236	26.98912	0.24365
0,8992	0,4248	0,9026	2.08547	30.68456	2.00424
0,8138	0,6635	0,9517	3.62364	17.86735	0.93309
0,8732	0,5091	0,8625	3.59819	29.73162	4.17148
0,7592	0,5148	0,7581	1.51058	19.51831	1.95795
0,891	0,9909	0,9616	0.29191	0.29883	0.11724
0,9243	0,4436	0,876	1.78094	33.12057	3.27909

0,8198	0,5577	0,7755	0,93289	16,48965	1,40442
0,8794	0,4692	0,8658	1,77485	24,58808	2,0391
0,776	0,4393	0,7745	2,38874	24,32138	3,01749
0,9578	0,7352	0,9561	3,51985	29,46491	3,52042
0,8967	0,4052	0,823	1,62731	28,73987	2,65562
0,9097	0,2533	0,8427	0,55638	28,59297	1,11684
0,9062	0,328	0,8191	1,31014	32,4062	2,80458
0,9592	0,5583	0,935	1,27959	28,72608	2,17561
0,8735	0,5202	0,8679	2,36089	25,66878	2,64321
0,5323	0,1697	0,484	4,16696	34,44944	4,83568
0,7971	0,4267	0,692	2,7544	25,48723	4,25449
0,8158	0,3212	0,7442	0,91642	24,83403	1,40026
0,9221	0,5002	0,9279	1,4428	25,54091	1,37803
0,8754	0,4263	0,8597	2,45956	29,0283	2,8899
0,8671	0,4434	0,8304	2,1729	26,25899	2,49189
0,8511	0,5644	0,8454	1,10874	16,21745	1,1672
0,9324	0,551	0,8726	0,93809	19,09653	1,48465
0,3916	0,3563	0,5105	2,4329	22,64652	2,36205
0,8859	0,4217	0,8481	1,49759	26,8842	2,17857
0,9238	0,484	0,9071	1,75691	28,54517	2,19896
0,9374	0,4124	0,8966	0,68684	25,23205	1,14056
0,7569	0,4101	0,7473	3,12972	24,88083	2,4435
0,7953	0,5678	0,783	2,24944	19,48027	3,31352
0,7944	0,1541	0,6499	2,46498	30,31899	1,21304
0,8185	0,5489	0,7502	0,92961	16,08339	1,00533
0,7149	0,456	0,6782	0,57203	23,17704	3,48006
0,9535	0,6832	0,9452	3,05518	21,65876	2,07219
0,7701	0,372	0,7172	3,05518	19,69272	1,14192
0,8112	0,2041	0,7182	0,73471	31,85267	2,34898
0,6607	0,3296	0,6189	2,13978	27,50776	3,0468
0,9439	0,2757	0,8469	0,56794	32,61808	1,65242

Nous remarquons que les résultats ( $R^2$  et SCR) obtenues par de application de la méthode RCP-Ondelettes sur données bruitées sont plus proches de ceux du modèle de référence que ceux de la RCP sur les données bruitées (tableau n° 11).

La figure 8 donne les  $R^2$  associés aux cas simulés fournis par la RCP avant et après introduction du bruit.

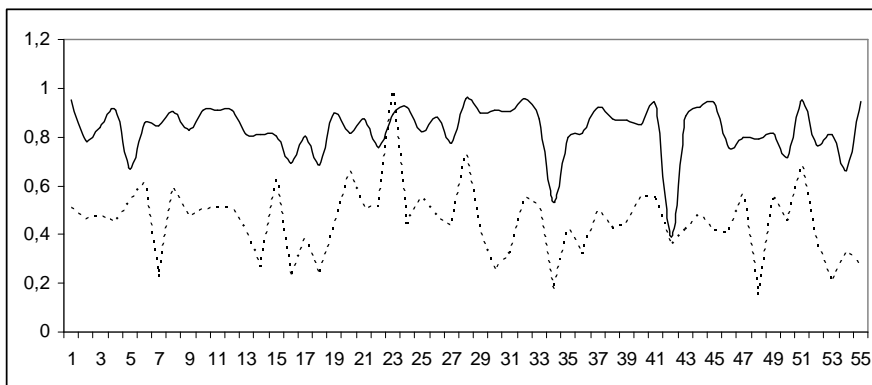
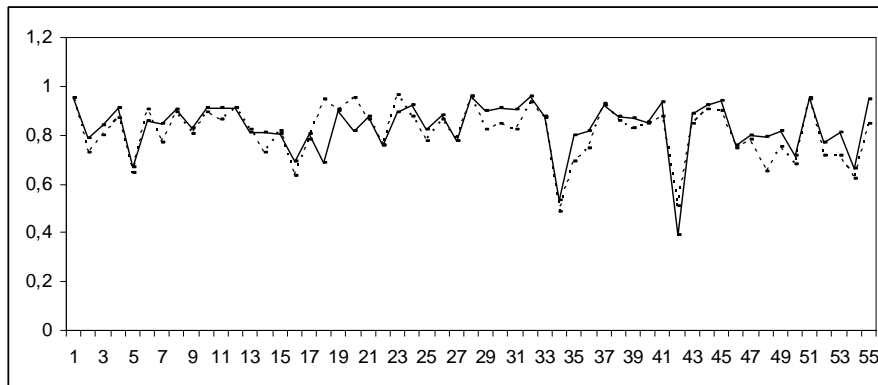


Figure 8 :  $R^2$  avant (plein) et après introduction du bruit (pointillé)



La figure 9 donne les  $R^2$  associés aux cas simulés fournis par les deux méthodes : RCP et RCP-Ondelettes.



**Figure 9**  
 *$R^2$  avant introduction du bruit (plein) et après débruitage (pointillé)*

Les  $R^2$  avant le bruitage du signal sont très proches de ceux fournis par la méthode RCP-Ondelettes après introduction du bruit. Ceci confirme les résultats obtenus sur les données réelles (voir §4).

## 6. Conclusion

Nous avons proposé ici une approche dite « RCP-Ondelettes » combinant la technique de débruitage des données par ondelettes et la Régression sur Composantes Principales. De point de vue qualité de l'ajustement du modèle, cette approche s'avère meilleure que la RCP classique. Nous avons testé cette méthode sur des données boursières françaises. Nos résultats sont confirmés par simulation.

## Bibliographie

- AMINGHAFARI, M. CHEZE, N. & POGGI, J.M. (2006). Multivariate denoising using wavelets and principal component analysis. *Computational Statistics & Data Analysis*, vol. 50, issue 9, pp. 2381-2398.
- BENAMMOU, S. (2003). Simulation & Monte Carlo Methods. *Master Course Notes, ISG Sousse*.
- DAUBECHIES, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math* 41, pp. 909-996.
- DAUBECHIES, I. (1992). Ten lectures on wavelets. *SIAM. Philadelphia*.
- DONOHO, D. & JOHNSTONE, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 81, pp. 425-455
- DONOHO, D. & JOHNSTONE, I. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* 26 (3), pp. 879-921.
- DONOHO, D. (1999). Wedgelets: nearly minimax estimation of edges. *Ann. Statist.* 27 (3). pp. 859-897.
- GABOR, D. (1964). Theory of communication. *J. Inst. Elect. Eng.* 93, pp.429-457.
- GENCAI, R. SELUK, F. & WHITCHER, B. (2002). An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. *Academic Press*.
- GLASSERMAN, P. (2003). Monte Carlo methods in financial engineering. *Springer-Verlag*.
- HAUGH, M. (2004). The Monte Carlo Framework, Examples from Finance and Generating Correlated Random Variables. *Course Notes*.
- MALLAT, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 11, pp. 674-693.
- TRUCHETET, F. (1998). Ondelettes pour le signal numérique. *Editions Hermes*.
- VIGNERON, V.; PARASCHIV-IONESCU, A.; AZANCOT, A.; JUTTEN, C. & SIBONY, O. (2003). Fetal electrocardiogram extraction based on non-stationary ICA and wavelet denoising, *Seventh IEEE International Symposium on Signal Processing and its Applications*, vol.2, pp. 69-72.

## Annexe

### Matrices des corrélations avant et après débruitage :

#### Matrice des corrélations avant débruitage :

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17
V1	1.0000	0.2568	0.3844	0.3381	0.5294	0.4824	0.4127	0.4009	0.3726	0.3914	0.2186	0.3896	0.2838	0.2996	0.3384	0.4642	0.3568
V2	0.2568	1.0000	0.2529	0.1919	0.3634	0.3353	0.2713	0.2430	0.2207	0.2857	0.0395	0.2414	0.1995	0.2188	0.2148	0.2980	0.2474
V3	0.3844	0.2529	1.0000	0.3116	0.5625	0.4885	0.3466	0.4836	0.3353	0.4137	0.1755	0.3776	0.3614	0.2571	0.4382	0.4306	0.3619
V4	0.3381	0.1919	0.3116	1.0000	0.3890	0.3998	0.3479	0.2934	0.2625	0.3517	0.1114	0.3071	0.3211	0.2607	0.2905	0.3822	0.3100
V5	0.5294	0.3634	0.5625	0.3890	1.0000	0.7453	0.5273	0.5282	0.4769	0.6453	0.2793	0.5621	0.4341	0.3628	0.5151	0.6013	0.4704
V6	0.4824	0.3353	0.4885	0.3998	0.7453	1.0000	0.4926	0.4588	0.4594	0.6887	0.2540	0.6059	0.3694	0.3446	0.4562	0.5884	0.4340
V7	0.4127	0.2713	0.3466	0.3479	0.5273	0.4926	1.0000	0.3894	0.3125	0.4460	0.2131	0.4532	0.3556	0.3094	0.3897	0.5192	0.3693
V8	0.4009	0.2430	0.4836	0.2934	0.5282	0.4588	0.3894	1.0000	0.3675	0.3964	0.2104	0.3946	0.3522	0.2655	0.3541	0.4215	0.3324
V9	0.3726	0.2207	0.3333	0.2625	0.4769	0.4584	0.3125	0.3675	1.0000	0.4030	0.2187	0.3082	0.2731	0.2457	0.3294	0.3828	0.2870
V10	0.3914	0.2857	0.4137	0.3517	0.6453	0.6887	0.4460	0.3964	0.4030	1.0000	0.2526	0.5103	0.3259	0.3150	0.3991	0.5214	0.3940
V11	0.2186	0.0395	0.1755	0.1114	0.2793	0.2540	0.2131	0.2104	0.2187	0.2526	1.0000	0.1675	0.1681	0.1159	0.1730	0.2078	0.1942
V12	0.3896	0.2414	0.3776	0.3071	0.5621	0.6059	0.4532	0.3946	0.3082	0.5103	0.1675	1.0000	0.3121	0.2661	0.3281	0.4875	0.3158
V13	0.2838	0.1995	0.3614	0.3211	0.4341	0.3694	0.3556	0.3522	0.2731	0.3259	0.1681	0.3121	1.0000	0.2491	0.2990	0.4085	0.4229
V14	0.2996	0.2188	0.2571	0.2607	0.3628	0.3446	0.3094	0.2655	0.2457	0.3150	0.1159	0.2661	0.2491	1.0000	0.2043	0.3362	0.2377
V15	0.3384	0.2148	0.4382	0.2905	0.5151	0.4562	0.3897	0.3541	0.3294	0.3991	0.1730	0.3281	0.2990	0.2043	1.0000	0.3700	0.3842
V16	0.4642	0.2980	0.4306	0.3822	0.6013	0.5884	0.5192	0.4215	0.3828	0.5214	0.2078	0.4875	0.4085	0.3362	0.3700	1.0000	0.4060
V17	0.3568	0.2474	0.3619	0.3100	0.4704	0.4340	0.3693	0.3324	0.2870	0.3940	0.1942	0.3158	0.4229	0.2377	0.3842	0.4060	1.0000
V18	0.3586	0.3001	0.3790	0.2773	0.5372	0.4994	0.3655	0.3669	0.4737	0.4295	0.2556	0.3373	0.2748	0.3317	0.3599	0.4291	0.3100
V19	0.4845	0.3464	0.5133	0.3956	0.6712	0.6187	0.4935	0.5307	0.4482	0.5466	0.3091	0.4773	0.3860	0.3631	0.4410	0.5717	0.4582
V20	0.4497	0.2787	0.3684	0.3822	0.5264	0.4510	0.4630	0.3770	0.3510	0.4046	0.1674	0.3971	0.3409	0.2717	0.3158	0.5058	0.3522
V21	0.2071	0.1628	0.2129	0.1842	0.3027	0.2853	0.2305	0.1919	0.2145	0.2829	0.1597	0.2135	0.2157	0.2240	0.1852	0.2997	0.2159
V22	0.3874	0.2482	0.4047	0.3495	0.5387	0.4955	0.4441	0.3938	0.3525	0.4253	0.1663	0.3978	0.3306	0.3023	0.3619	0.4951	0.3666
V23	0.4187	0.2511	0.3934	0.3409	0.5174	0.4971	0.4164	0.4095	0.3878	0.4605	0.1570	0.4254	0.2922	0.2920	0.3146	0.4563	0.3797
V24	0.4457	0.3028	0.4275	0.3923	0.5902	0.5656	0.4780	0.4080	0.3899	0.4978	0.1103	0.4585	0.3574	0.3118	0.3736	0.5386	0.3754
V25	0.5151	0.3097	0.4715	0.3966	0.6298	0.6212	0.5374	0.4418	0.4312	0.5399	0.2535	0.4833	0.3904	0.3376	0.4374	0.6515	0.4323
V26	0.2726	0.1941	0.2775	0.1918	0.4165	0.3864	0.2705	0.2409	0.2926	0.3693	0.1867	0.3039	0.1713	0.2136	0.3103	0.3174	0.2588
V27	0.4627	0.3055	0.4108	0.3448	0.5464	0.5305	0.4471	0.4243	0.3662	0.4844	0.1900	0.4296	0.3409	0.3321	0.3324	0.5487	0.3458
V28	0.4717	0.3101	0.4814	0.3679	0.7298	0.8062	0.5009	0.4640	0.4337	0.6468	0.2539	0.6156	0.3792	0.3409	0.4243	0.5978	0.4423
V29	0.3828	0.2611	0.5783	0.3139	0.5657	0.5079	0.3755	0.4883	0.3521	0.4234	0.2063	0.3311	0.3413	0.2521	0.4241	0.4275	0.3519
V30	0.3695	0.2936	0.3776	0.2575	0.5491	0.4974	0.3986	0.3356	0.3301	0.4392	0.2570	0.3415	0.2920	0.2784	0.3831	0.4568	0.3355
V31	0.1860	0.1437	0.1379	0.2036	0.2284	0.2263	0.1870	0.1654	0.1470	0.2642	0.1118	0.1704	0.1986	0.0956	0.1331	0.2070	0.1441
V32	0.3010	0.2387	0.2638	0.2866	0.4082	0.3792	0.3751	0.2659	0.2679	0.3327	0.1685	0.2757	0.3547	0.2661	0.3076	0.3502	0.3204
V33	0.2324	0.1233	0.1723	0.1857	0.2774	0.2730	0.2770	0.1991	0.1978	0.2427	0.1144	0.2117	0.1322	0.1968	0.1919	0.2730	0.1664
V34	0.4198	0.3132	0.4843	0.3166	0.5949	0.5383	0.4748	0.4321	0.4248	0.4538	0.2335	0.4317	0.3879	0.3114	0.5415	0.4913	0.4670
V35	0.3437	0.1788	0.2300	0.2781	0.3656	0.3657	0.4696	0.2870	0.2517	0.3766	0.1505	0.3519	0.2264	0.2360	0.2320	0.3852	0.2387
V36	0.2053	0.1312	0.1864	0.2200	0.2051	0.2240	0.2202	0.2512	0.1667	0.2110	0.1226	0.1711	0.2146	0.1704	0.1588	0.2709	0.1449
V37	0.4591	0.2290	0.4382	0.2920	0.4896	0.4783	0.3895	0.3846	0.3818	0.3993	0.1664	0.4212	0.3450	0.2425	0.3558	0.4722	0.3804



	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36	V37
V1	0.2614	0.4368	0.4573	0.1267	0.3188	0.4247	0.3597	0.4446	0.1270	0.4566	0.3944	0.3066	0.2542	0.0718	0.2391	0.1618	0.2662	0.3813	0.1248	0.3888
V2	0.0868	0.1353	0.0842	0.0574	0.0444	0.0717	0.0948	0.0797	0.0704	0.0882	0.0808	0.0866	0.0718	0.0311	0.1343	0.0053	0.1124	0.0433	0.0431	0.0443
V3	0.2023	0.3479	0.2432	0.1375	0.2744	0.2557	0.2638	0.3298	0.0741	0.2306	0.2730	0.4923	0.1806	0.0466	0.1444	0.0421	0.3522	0.1082	0.0961	0.3132
V4	0.1818	0.2184	0.3040	0.0746	0.2257	0.2552	0.3307	0.2953	0.0809	0.2833	0.2497	0.1523	0.1359	0.0990	0.1608	0.0538	0.1693	0.2292	0.1098	0.2023
V5	0.4368	0.5685	0.4591	0.2246	0.4086	0.4658	0.5114	0.5649	0.1794	0.5439	0.6456	0.4223	0.4155	0.0794	0.3659	0.1726	0.4149	0.3714	0.1103	0.4147
V6	0.4047	0.3950	0.3891	0.1460	0.3588	0.4479	0.4778	0.5378	0.2109	0.5297	0.7735	0.2906	0.3204	0.0281	0.2908	0.1109	0.3387	0.3939	0.1273	0.4195
V7	0.2561	0.3565	0.4092	0.1701	0.3762	0.4164	0.4385	0.4876	0.1803	0.4449	0.4191	0.3164	0.2557	0.0982	0.3199	0.1377	0.3692	0.4861	0.1530	0.3192
V8	0.3021	0.4507	0.4340	0.1701	0.2627	0.4000	0.3364	0.3725	0.1106	0.4136	0.3190	0.4269	0.2197	0.0692	0.2145	0.0809	0.2842	0.2653	0.2077	0.3824
V9	0.3485	0.2900	0.3241	0.0757	0.2580	0.3900	0.3055	0.3282	0.1313	0.3211	0.2987	0.1669	0.1927	0.0237	0.1829	0.1157	0.2503	0.2631	0.0794	0.3975
V10	0.2522	0.4051	0.3297	0.1733	0.2963	0.3918	0.4182	0.4475	0.1831	0.4777	0.5532	0.2638	0.2565	0.1861	0.2810	0.1392	0.3153	0.3507	0.1272	0.3435
V11	0.0995	0.1326	0.0593	0.0778	0.0314	0.0382	-0.8001	0.0641	0.0522	0.0723	0.0868	0.0821	0.1123	0.0309	0.0737	0.0195	0.0780	0.0663	0.0493	0.0924
V12	0.2091	0.3504	0.3859	0.1370	0.3172	0.4357	0.3742	0.3818	0.1877	0.3609	0.6206	0.2337	0.2083	0.0648	0.2279	0.1084	0.2933	0.3534	0.0982	0.4237
V13	0.1582	0.2972	0.2606	0.1897	0.2167	0.1739	0.2688	0.2742	0.0122	0.2138	0.2297	0.2231	0.1732	0.0697	0.3704	0.0126	0.2423	0.1001	0.0879	0.1926
V14	0.2803	0.3849	0.2692	0.2801	0.2380	0.2444	0.3459	0.3303	0.1029	0.3117	0.3037	0.1903	0.2367	0.0665	0.2752	0.1113	0.2892	0.2626	0.1435	0.1971
V15	0.1784	0.2153	0.1629	0.0100	0.2135	0.1028	0.1681	0.2102	0.1557	0.0632	0.1274	0.3160	0.2024	0.0645	0.2326	0.0223	0.4100	0.0952	0.0462	0.1705
V16	0.3295	0.4189	0.4450	0.1812	0.3869	0.4004	0.4679	0.6057	0.1594	0.5018	0.5050	0.2498	0.3011	0.0664	0.2881	0.1284	0.3433	0.3420	0.1955	0.3989
V17	0.2429	0.3401	0.3112	0.1268	0.2420	0.2568	0.2666	0.3612	0.1476	0.2894	0.3383	0.2850	0.2515	0.0637	0.2914	0.0450	0.3360	0.1875	0.0908	0.2569
V18	1.0000	0.4523	0.2556	0.2292	0.2103	0.3080	0.3125	0.3905	0.1433	0.3168	0.3620	0.2962	0.3425	0.0527	0.2952	0.1014	0.2224	0.2295	0.0946	0.2372
V19	0.4523	1.0000	0.3993	0.2301	0.2848	0.4348	0.4264	0.4391	0.1575	0.4727	0.3930	0.4218	0.2833	0.0973	0.3005	0.1222	0.3814	0.3044	0.2115	0.3799
V20	0.2556	0.3993	1.0000	0.1911	0.4635	0.4744	0.5742	0.4972	0.1126	0.4313	0.3747	0.2933	0.2229	0.0764	0.2428	0.1381	0.2978	0.4157	0.1264	0.4069
V21	0.2292	0.2301	0.1911	1.0000	0.0938	0.1380	0.1882	0.2037	0.0349	0.1641	0.1656	0.1604	0.1997	0.0424	0.2228	0.0753	0.1512	0.1245	0.1007	0.0659
V22	0.2103	0.2848	0.4635	0.0938	1.0000	0.3626	0.6009	0.4762	0.1674	0.3377	0.3647	0.3555	0.2124	0.0378	0.1687	0.1297	0.2714	0.2458	0.1040	0.3558
V23	0.3080	0.4348	0.4744	0.1380	0.3626	1.0000	0.4772	0.4689	0.1611	0.5358	0.4366	0.3456	0.1862	0.0721	0.1768	0.1508	0.2532	0.3392	0.1182	0.4706
V24	0.3125	0.4264	0.5742	0.1882	0.6009	0.4772	1.0000	0.5301	0.1418	0.5090	0.4679	0.2677	0.2027	0.0579	0.2095	0.1356	0.3475	0.3135	0.1162	0.3857
V25	0.3905	0.4391	0.4972	0.2037	0.4762	0.4689	0.5301	1.0000	0.1328	0.5829	0.5287	0.3309	0.3289	0.0975	0.2740	0.1929	0.3786	0.3648	0.1764	0.4478
V26	0.1433	0.1575	0.1126	0.0349	0.1674	0.1611	0.1418	0.1328	1.0000	0.1844	0.1999	0.1703	0.1256	0.0293	0.1424	0.0118	0.2104	0.1587	0.0491	0.1384
V27	0.3168	0.4727	0.4313	0.1641	0.3377	0.5358	0.5090	0.5829	0.1844	1.0000	0.4995	0.3122	0.2431	0.0839	0.2501	0.1576	0.2819	0.3383	0.1853	0.4045
V28	0.3620	0.3930	0.3747	0.1656	0.3647	0.4366	0.4679	0.5287	0.1999	0.4995	1.0000	0.2766	0.2793	0.0134	0.2770	0.1558	0.3131	0.3861	0.1283	0.4261
V29	0.2962	0.4218	0.2933	0.1604	0.3555	0.3456	0.2677	0.3309	0.1703	0.3122	0.2766	1.0000	0.1881	0.0897	0.1863	0.0201	0.3796	0.1893	0.0337	0.2808
V30	0.3425	0.2833	0.2229	0.1997	0.2124	0.1862	0.2027	0.3289	0.1256	0.2431	0.2793	0.1881	1.0000	0.0489	0.4321	0.0947	0.3150	0.2213	0.0838	0.2597
V31	0.0527	0.0973	0.0764	0.0424	0.0378	0.0721	0.0579	0.0975	0.0293	0.0839	0.0134	0.0897	0.0489	1.0000	0.0956	0.0210	0.0506	0.0780	0.0913	0.0350
V32	0.2952	0.3005	0.2428	0.2228	0.1687	0.1768	0.2095	0.2740	0.1424	0.2501	0.2770	0.1863	0.4321	0.0956	1.0000	0.0719	0.2960	0.2367	0.0881	0.1725
V33	0.1014	0.1222	0.1381	0.0753	0.1297	0.1508	0.1356	0.1929	0.0118	0.1576	0.1558	0.0201	0.0947	0.0210	0.0719	1.0000	0.0822	0.1412	0.0530	0.0984
V34	0.2224	0.3814	0.2978	0.1512	0.2714	0.2532	0.3475	0.3786	0.2104	0.2819	0.3131	0.3796	0.3150	0.0506	0.2960	0.0822	1.0000	0.2101	0.0700	0.3365
V35	0.2295	0.3044	0.4157	0.1245	0.2458	0.3392	0.3135	0.3648	0.1587	0.3383	0.3861	0.1893	0.2213	0.0780	0.2367	0.1412	0.2101	1.0000	0.1105	0.3619
V36	0.0946	0.2115	0.1264	0.1007	0.1040	0.1182	0.1162	0.1764	0.0491	0.1853	0.1283	0.0337	0.0838	0.0913	0.0881	0.0530	0.0700	0.1105	1.0000	0.0615
V37	0.2372	0.3799	0.4069	0.0659	0.3558	0.4706	0.3857	0.4478	0.1384	0.4045	0.4261	0.2808	0.2597	0.0350	0.1725	0.0984	0.3365	0.3619	0.0615	1.0000