

MASTER 2 ECONOMIE & INGÉNIERIE FINANCIÈRE - 272
UNIVERSITÉ PARIS DAUPHINE - PSL

Catching the curl: Wavelet thresholding improves forward curve modelling

Une application à la modélisation de la courbe forward des contrat futures
sur maïs

Auteurs :

Naïm LEHBIBEN
Cyprien TARDIVEL
Joudy BENKADDOUR
Xiyu XUE

Membres du jury :

Matthieu GARCIN
Julien ROYER
Chafic MERHY

18 mai 2024

Table des matières

1	Préliminaires	1
1.1	Présentation du sujet	1
1.2	Revue de littérature	1
1.3	Données et méthodologie	3
1.3.1	Données	3
1.3.2	Statistiques descriptives	3
1.3.3	Méthodologie	3
2	Résultats empiriques de l'article	8
2.1	Estimation du modèle espace d'état	8
2.1.1	RMSE	8
2.1.2	Erreur quadratique moyenne du modèle (RMSE)	9
2.1.3	Prévision	10
3	Nos résultats empiriques	11
3.1	Estimation du modèle espace d'état	11
3.1.1	Problèmes rencontrés	11
3.1.2	Erreur quadratique moyenne du modèle (RMSE)	11
3.1.3	Prévision	12
4	Conclusion	13
	Annexe	15
.1	Term Structure of Corn Futures Prices	15
.2	Plot of front-month corn futures daily settlement price	15
.3	Sorensen Parameters Estimate (2002)	16

Chapitre 1

Préliminaires

1.1 Présentation du sujet

Dans le cadre du cours de Gestion Quantitative du Master Economie & Ingénierie Financière de l'université Paris Dauphine - PSL nous allons répliquer l'article de recherche intitulé "*Catching the curl : Wavelet thresholding improves forward curve modelling*" de **Gabriel J. Power, James Eaves, Calum Turvey, Dmitry Vedenov** publié dans le journal *Economic Modelling* en 2017.

L'objectif de cette étude est de modéliser la structure par terme des futures sur maïs. Pour cela, les auteurs de l'article utilisent une nouvelle stratégie de seuillage par ondelette, pour débruiter les données sur les prix à terme avant l'estimation d'un modèle espace d'état à l'aide d'un filtre de Kalman afin d'améliorer l'ajustement et la prédiction du modèle.

1.2 Revue de littérature

La structure par terme des prix des contrats futures sur matières premières joue un rôle clé dans le cadre de stratégies d'investissement, illustrant les relations entre le prix au comptant et les prix à terme pour différentes dates de livraison. Au cours des dernières décennies, plusieurs modèles basés sur des variables d'état ont été développés pour décrire cette structure par terme et sa dynamique. La formulation en espace d'état établit une relation entre un vecteur de séries temporelles observables (prix futurs pour différentes échéances) et un vecteur de variables non observables. Le filtre de Kalman est alors appliqué de manière récursive pour calculer l'estimateur optimal du vecteur d'état au temps t , en utilisant les informations disponibles à ce moment.

Parmi ces modèles, celui à un facteur de BRENNAN et SCHWARTZ [1], qui traite le prix au comptant comme un mouvement brownien géométrique, est le plus connu. Le modèle à deux facteurs développé par GIBSON et SCHWARTZ [2] est un instrument fiable pour l'évaluation des contrats futures à court terme, estimant le convenience yield à l'aide d'un processus stochastique de retour à la moyenne. SCHWARTZ [3] a ensuite proposé un modèle à trois facteurs intégrant le taux d'intérêt, qui suit un processus d'Ornstein-Uhlenbeck. Compte tenu de l'incertitude de certains produits quant au prix d'équilibre vers lequel les prix tendent à revenir, SCHWARTZ et SMITH [4] ont suggéré un modèle affine à deux facteurs capturant à la fois les effets de retour à la moyenne et d'incertitude. Dans cet esprit, de nombreuses études ont appliqué la structure affine pour modéliser la structure à terme des prix des contacts futures [5] ; [6].

Comme suggéré par HASBROUCK [7], les fluctuations des prix des contrats à terme sont souvent perturbées par le bruit. Pour cette raison, il est crucial de débruiter les données avant de les utiliser dans le modèle afin d'obtenir des prédictions robustes. Bien qu'il existe différents

types de méthodes de filtrage, le seuillage par ondelettes se distingue comme une méthode particulièrement adaptée. Depuis 1980, des recherches approfondies ont été menées pour explorer l'utilisation de la transformée en ondelettes dans l'analyse des séries temporelles. Les recherches effectuées par DONOHO et JOHNSTONE [8] ont démontré l'efficacité et la précision remarquables du seuillage par ondelettes pour éliminer le bruit et restituer le signal réel des données. PERCIVAL et MONDAL [9] ont souligné que les propriétés de localisation temporelle et fréquentielle rendent cette méthode mieux appropriée pour les séries temporelles. Parmi les différentes transformées en ondelettes, l'ondelette Daubechies (Daubechies Least Asymmetrical wavelet) a été choisie en raison de ses propriétés avantageuses : nombre non nul de moments, support compact, l'orthogonalité et l'orthonormalité, et phase linéaire.

Un nombre arbitraire de moments non nuls augmente la capacité à détecter les variations de haute fréquence. Le support compact favorise une localisation précise des variations locales, améliorant ainsi l'analyse des données non périodiques ([10], [11]). L'orthogonalité et l'orthonormalité garantissent que les différentes échelles de transformation demeurent non corrélées, ce qui permet une reconstruction fidèle et précise des données [12]. Enfin, la phase linéaire contribue à une localisation temporelle exacte, indispensable pour identifier les points de changement dans les séries temporelles.

Ces littératures montrent clairement la rationalité des décisions de l'auteur, allant du débruitage des données à la construction de modèle de structure par terme. Dans notre étude, nous avons également adopté la même approche pour réaliser le débruitage par ondelettes et estimer les paramètres du modèle avec des facteurs en le représentant sous forme d'espace d'état et en utilisant le filtre de Kalman.

1.3 Données et méthodologie

1.3.1 Données

Les données sont les prix de règlement des contrats futures sur maïs échangé sur le Chicago Board of Trade. Les prix journaliers des futures sur maïs ont été récupéré à l'aide de l'API fournit par **Bloomberg** (BL API) et stockés dans un fichier en dur (.pkl) dans le repertoire data. La plage de données va de 1988 à 2024 pour les 5 échéances des futures sur maïs (Mars, Mai, Juillet, Septembre, Decembre). La série est assemblée en passant au deuxième contrat à terme le plus proche deux semaines avant l'expiration du premier contrat. Le 11 septembre 2001 est considéré comme un outliers et enlevé de la plage de données.

1.3.2 Statistiques descriptives

Voici un tableau présentant les statistiques descriptives en utilisant comme série les contracts aux maturités la plus proche (first nearby futures contract) :

TABLE 1.1 – Statistiques descriptives des prix et des rendements logarithmiques

Statistic	Prices (c/bu)	Log returns
Mean	328.0151	0.0001
Median	269.1875	0.0
Minimum	174.75	-0.272
Maximum	831.25	0.128
Std deviation	143.9133	0.0172
Skewness	1.53	-1.04
Kurtosis	1.57	19.59
Autocorrelation	0.999	0.05
ADF test p-value (10 lags)	0.16	0.0
Nb obs	7312	7311

Nous présentons des statistiques descriptives dans le tableau 1 et un tracé de séries chronologiques de le prix à terme dans la figure 2. Le tableau 1 montre que les prix à terme du maïs les prix sont non stationnaires, asymétriques à droite et platikurtique.

Les rendements logarithmiques des contrats à terme sont stationnaires et faiblement autocorrélés (premier retard statistiquement mais non économiquement significatif), asymétrique à gauche et fortement leptokurtique. La normalité des log rendement est rejetée, mais cela n'est ni un problème pour le seuillage d'ondelettes ni pour l'estimation d'un modèle espace d'états, qui utilise le quasi-MLE.

1.3.3 Méthodologie

Modèle théorique de la courbe forward

Les auteurs supposent que log du prix spot est une fonction affine de N variables d'état et d'une fonction périodique déterministe. Cette dernière capture les variations saisonnières du prix des matières premières. Ainsi, le logarithme du prix spot P_t est la somme de deux composantes, à savoir une fonction cyclique $s(t)$ et de variables d'état x_i

$$\ln(F_{t,T_j}) = s(t) + \sum_{i=1}^N x_i(t; T_j) \quad (1)$$

Les variables d'état sont très souvent inobservables, mais elles incluent dans certains modèles le prix spot et le taux d'intérêt. Cependant, cette approche n'est pas retenue par les auteurs, car des recherches antérieures ont montré que, pour les contrats à terme des produits agricoles, le risque de taux d'intérêt a peu d'impact sur le modèle.

La dynamique de chaque variable d'état x_t est décrite par une équation différentielle stochastique (EDS) résolue à l'aide d'une équation aux dérivées partielles de Feynman-Kac. Dans le cas multivarié l'équation est notée :

$$dx_t = -Kx_t dt + \Sigma dw_t \quad (2)$$

où x_t est la variable d'état, K une matrice de drift, Σ est la matrice des termes de diffusion, et w_t un mouvement brownien (Wiener process).

Pour capturer la saisonnalité des matières premières, une fonction cyclique s_t est utilisée :

$$s_t = \sum_{k=1}^K \gamma_k \cos(2\pi kt) + \bar{\gamma}_k \sin(2\pi kt) \quad (3)$$

Dans l'équation (3), chaque coefficient prend une valeur constante qui doit d'abord être estimée à partir des données avant l'estimation du modèle espace d'état (Sans précision supplémentaire). Selon SORENSEN [13], $K=2$ est suffisant pour capturer cycles saisonniers des prix des matières premières agricoles.

Dans ce modèle la première variable d'état est un mouvement brownien géométrique (Wiener process) et les $N-1$ autres variables d'état sont des processus mean-reverting (Ornstein-Uhlenbeck) :

$$dx_1(t) = \mu x_1(t) dt + \sigma_1 x_1(t) dw_1(t) \quad (4)$$

Les variables d'état x_2 à x_N sont définies comme des processus d'Ornstein-Uhlenbeck, chacun avec une vitesse différente de retour à la moyenne capturée par κ_n et la moyenne à long terme vers laquelle le processus est attiré est C :

$$dx_n(t) = -\kappa_n(x_n(t) - C) dt + \sigma_n dw_n(t) \quad (5)$$

Les mouvements browniens sont supposés être corrélés deux à deux à travers un coefficient ρ_{ij} . La structure temporelle de la volatilité des prix à terme est obtenue en utilisant les paramètres de diffusion et de corrélation estimés :

$$\sigma_F^2(T-t) = \sum_{i=1}^N \sum_{j=1}^N \sigma_i \sigma_j \rho_{ij} e^{-(\kappa_i + \kappa_j)(T-t)} \quad (6)$$

Le papier utilise l'absence d'opportunité d'arbitrage pour obtenir la relation analytique entre entre prix spot et prix future [14] :

$$\begin{aligned} \log F(x, t, T) = & s(t) + x_1(t) + \sum_{i=2}^N e^{-\kappa_i(T-t)} x_i(t) + \mu t + (\mu - \lambda_1 + \frac{1}{2}\sigma_1^2)(T-t) \\ & - \sum_{i=2}^N \frac{(1 - e^{-\kappa_i(T-t)})}{\kappa_i} \lambda_i + \frac{1}{2} \sum_{i,j \neq i} \frac{(1 - e^{-(\kappa_i + \kappa_j)(T-t)})}{\kappa_i + \kappa_j} \sigma_i \sigma_j \rho_{ij} \end{aligned} \quad (7)$$

Debruitage via shrinkage

L'objectif du seuillage par ondelettes est de déterminer une valeur optimale valeur du seuil en utilisant un critère précis, tel qu'une fonction de perte ou valeur de risque minimale (STEIN [15]). Le choix du seuil et de la règle du seuil doivent être choisis de manière parcimonieuse.

Pour déterminer ce seuil il faut utiliser une transformation en ondelettes discrète (Discrete Wavelet Transform) aux données pour produire une matrice de coefficient en ondelette puis appliquer une transformation en ondelettes discrète inverse aux coefficients de la matrice de coefficient en ondelette donnant une version filtrée de l'original sans perte d'informations.

Supposons une variance des innovations (erreurs) σ_ϵ^2 et un nombre d'observations T . Le seuil universel est donné par :

$$\delta = \sqrt{2\sigma_\epsilon^2 \ln(T)} \quad (8)$$

La règle de seuillage doux appliquée aux coefficients d'ondelette w est donnée par :

$$w^{\text{soft}} = \text{sgn}(w) \left(\frac{1}{2} (|w| - \delta + (|w| - \delta)) \right) \quad (9)$$

Le rapport de la médiane des coefficients d'ondelettes à l'échelle de temps la plus fine sur un facteur de normalisation s'est avéré optimal :

$$\hat{\sigma}_{\text{MAD}} = \frac{\text{median}(w^{j=1})}{0.6745} \quad (10)$$

Estimation du modèle

Le filtre de Kalman est utilisé pour estimer la vraisemblance des paramètres du modèle espace-état des prix à terme. Le modèle espace-état est basé sur une équation de transition (état) et une équation de mesure définie ci-dessous :

$$\begin{aligned} X_{t+1} &= a + AX_t + \eta_t \\ Y_t &= c_t + C_t X_t + \varepsilon_t \end{aligned} \quad (11)$$

Les composantes du vecteur de transition a et des matrices de transition A pour des modèles à 1, 2, 3, et 4 facteurs sont définies comme suit, présentées ensemble pour chaque configuration de facteur :

$$\begin{aligned} \text{1 facteur : } a &= \left(\mu - \frac{1}{2}\sigma_1^2 \right), & A &= 1 \\ \text{2 facteurs : } a &= \begin{pmatrix} \mu - \frac{1}{2}\sigma_1^2 \\ 0 \end{pmatrix}, & A &= \begin{pmatrix} 1 & 0 \\ 0 & e^{-k_2\Delta} \end{pmatrix} \\ \text{3 facteurs : } a &= \begin{pmatrix} \mu - \frac{1}{2}\sigma_1^2 \\ 0 \\ 0 \end{pmatrix}, & A &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-k_2\Delta} & 0 \\ 0 & 0 & e^{-k_3\Delta} \end{pmatrix} \\ \text{4 facteurs : } a &= \begin{pmatrix} \mu - \frac{1}{2}\sigma_1^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, & A &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-k_2\Delta} & 0 & 0 \\ 0 & 0 & e^{-k_3\Delta} & 0 \\ 0 & 0 & 0 & e^{-k_4\Delta} \end{pmatrix} \end{aligned}$$

Le vecteur c_t est défini comme un vecteur colonne où chaque composante est donnée par :

$$c_t = \begin{pmatrix} s(t) + (\mu + \lambda_1 - \frac{1}{2}\sigma_1^2)(T_1 - t) \\ s(t) + (\mu + \lambda_2 - \frac{1}{2}\sigma_1^2)(T_2 - t) \\ s(t) + (\mu + \lambda_3 - \frac{1}{2}\sigma_1^2)(T_3 - t) \\ s(t) + (\mu + \lambda_4 - \frac{1}{2}\sigma_1^2)(T_4 - t) \\ s(t) + (\mu + \lambda_1 - \frac{1}{2}\sigma_1^2)(T_5 - t) \end{pmatrix}$$

La matrice de mesure C_t est défini de la manière suivante :

Pour un modèle à plusieurs facteurs, où k_i est la constante de décroissance pour le second facteur, et $T(i)$ représente la maturité correspondant à chaque terme :

$$C_t = \begin{pmatrix} 1 & e^{-k_2(T_1-t)} & e^{-k_3(T_1-t)} & e^{-k_4(T_1-t)} \\ 1 & e^{-k_2(T_2-t)} & e^{-k_3(T_2-t)} & e^{-k_4(T_2-t)} \\ 1 & e^{-k_2(T_3-t)} & e^{-k_3(T_3-t)} & e^{-k_4(T_3-t)} \\ 1 & e^{-k_2(T_4-t)} & e^{-k_3(T_4-t)} & e^{-k_4(T_4-t)} \\ 1 & e^{-k_2(T_5-t)} & e^{-k_3(T_5-t)} & e^{-k_4(T_5-t)} \end{pmatrix}$$

Notez que T_i représente le temps jusqu'à la maturité dans le modèle, et $s(t)$ représente une fonction saisonnière ou une autre forme de modification basée sur le temps t .

La matrice de variance covariance est défini de la manière suivante :

Pour un modèle à 1 facteur, la matrice Q est une matrice 1×1 :

$$Q = \sigma_1^2 \cdot \Delta$$

Pour un modèle à 2 facteurs, la matrice Q est une matrice 2×2 :

$$Q = \begin{pmatrix} \sigma_1^2 \cdot \Delta & \frac{\rho_{12} \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - e^{-(\kappa_1 + \kappa_2) \cdot \Delta})}{\kappa_1 + \kappa_2} \\ \frac{\rho_{12} \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - e^{-(\kappa_1 + \kappa_2) \cdot \Delta})}{\kappa_1 + \kappa_2} & \frac{\sigma_2^2 \cdot (1 - e^{-2\kappa_2 \cdot \Delta})}{2\kappa_2} \end{pmatrix}$$

Pour un modèle à 3 facteurs, la matrice Q est une matrice 3×3 :

$$Q = \begin{pmatrix} \sigma_1^2 \cdot \Delta & \frac{\rho_{12} \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - e^{-(\kappa_1 + \kappa_2) \cdot \Delta})}{\kappa_1 + \kappa_2} & \frac{\rho_{13} \cdot \sigma_1 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_1 + \kappa_3) \cdot \Delta})}{\kappa_1 + \kappa_3} \\ \frac{\rho_{12} \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - e^{-(\kappa_1 + \kappa_2) \cdot \Delta})}{\kappa_1 + \kappa_2} & \frac{\sigma_2^2 \cdot (1 - e^{-2\kappa_2 \cdot \Delta})}{2\kappa_2} & \frac{\rho_{23} \cdot \sigma_2 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_2 + \kappa_3) \cdot \Delta})}{\kappa_2 + \kappa_3} \\ \frac{\rho_{13} \cdot \sigma_1 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_1 + \kappa_3) \cdot \Delta})}{\kappa_1 + \kappa_3} & \frac{\rho_{23} \cdot \sigma_2 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_2 + \kappa_3) \cdot \Delta})}{\kappa_2 + \kappa_3} & \frac{\sigma_3^2 \cdot (1 - e^{-2\kappa_3 \cdot \Delta})}{2\kappa_3} \end{pmatrix}$$

Pour un modèle à 4 facteurs, la matrice Q est une matrice 4×4 :

$$Q = \begin{pmatrix} \sigma_1^2 \cdot \Delta & \frac{\rho_{12} \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - e^{-(\kappa_1 + \kappa_2) \cdot \Delta})}{\kappa_1 + \kappa_2} & \frac{\rho_{13} \cdot \sigma_1 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_1 + \kappa_3) \cdot \Delta})}{\kappa_1 + \kappa_3} & \frac{\rho_{14} \cdot \sigma_1 \cdot \sigma_4 \cdot (1 - e^{-(\kappa_1 + \kappa_4) \cdot \Delta})}{\kappa_1 + \kappa_4} \\ \frac{\rho_{12} \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - e^{-(\kappa_1 + \kappa_2) \cdot \Delta})}{\kappa_1 + \kappa_2} & \frac{\sigma_2^2 \cdot (1 - e^{-2\kappa_2 \cdot \Delta})}{2\kappa_2} & \frac{\rho_{23} \cdot \sigma_2 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_2 + \kappa_3) \cdot \Delta})}{\kappa_2 + \kappa_3} & \frac{\rho_{24} \cdot \sigma_2 \cdot \sigma_4 \cdot (1 - e^{-(\kappa_2 + \kappa_4) \cdot \Delta})}{\kappa_2 + \kappa_4} \\ \frac{\rho_{13} \cdot \sigma_1 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_1 + \kappa_3) \cdot \Delta})}{\kappa_1 + \kappa_3} & \frac{\rho_{23} \cdot \sigma_2 \cdot \sigma_3 \cdot (1 - e^{-(\kappa_2 + \kappa_3) \cdot \Delta})}{\kappa_2 + \kappa_3} & \frac{\sigma_3^2 \cdot (1 - e^{-2\kappa_3 \cdot \Delta})}{2\kappa_3} & \frac{\rho_{34} \cdot \sigma_3 \cdot \sigma_4 \cdot (1 - e^{-(\kappa_3 + \kappa_4) \cdot \Delta})}{\kappa_3 + \kappa_4} \\ \frac{\rho_{14} \cdot \sigma_1 \cdot \sigma_4 \cdot (1 - e^{-(\kappa_1 + \kappa_4) \cdot \Delta})}{\kappa_1 + \kappa_4} & \frac{\rho_{24} \cdot \sigma_2 \cdot \sigma_4 \cdot (1 - e^{-(\kappa_2 + \kappa_4) \cdot \Delta})}{\kappa_2 + \kappa_4} & \frac{\rho_{34} \cdot \sigma_3 \cdot \sigma_4 \cdot (1 - e^{-(\kappa_3 + \kappa_4) \cdot \Delta})}{\kappa_3 + \kappa_4} & \frac{\sigma_4^2 \cdot (1 - e^{-2\kappa_4 \cdot \Delta})}{2\kappa_4} \end{pmatrix}$$

Ces matrices utilisent les paramètres de volatilité σ_i , les taux de réversion κ_i , les corrélations ρ_{ij} et l'intervalle de temps Δ qui est égal à 1 jour ouvré :

$$\Delta = 1/250 = 0.004$$

Procédure d'estimation et d'optimisation

Concernant la procédure d'estimation, le filtre de Kalman est initialisé avec les valeurs initiales, puis calcule des prévisions à la période suivante et les corrige à partir des observations réelles. Pour les valeurs de départ les auteurs utilisent les valeurs de l'estimation de [13] disponible en annexe.

Pour l'optimisation, les auteurs utilisent l'algorithme de Méthode de Nelder-Mead (5000 iterations) pour obtenir une première approximation des valeurs des paramètres, puis la méthode BFGS (200 iterations) pour estimer la véritable valeur des paramètres ainsi que leurs standard errors.

Calcul de l'erreur quadratique moyenne (RMSE)

Dans notre modèle le log des prix futures sont une fonction linéaire de variable d'état non stationnaire, de variable d'état stationnaire et de paramètres. A partir de la relation (7) les auteurs calculent le **RMSE** comme l'écart entre la prédiction du modèle et les prix de marché. La formule pour le Root Mean Squared Error **RMSE** est donnée par :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (12)$$

où :

- $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ sont les valeurs prédites par le modèle
- y_1, y_2, \dots, y_n sont les valeurs observées,
- n est le nombre d'observations.

Prévision

Après avoir mis de côté 250 observations (l'ensemble des observations de l'année 2016), les auteurs utilisent le modèle espace d'état pour réaliser des prévisions sur le prix des contrats futures à 5 jours.

Ils réalisent donc 50 prévisions et calculent encore une fois le RMSE entre la prévision des prix des futures et ceux sur l'année 2016. Cette méthode leur permet d'évaluer l'efficacité du pouvoir prédictif du modèle qui est à la base de la modélisation espace d'état et du filtre de Kalman.

Chapitre 2

Résultats empiriques de l'article

2.1 Estimation du modèle espace d'état

2.1.1 RMSE

TABLE 2.1 – Ce tableau présente les résultats d'estimation de modèles de un à quatre facteurs de la structure par terme des prix à terme, en utilisant les 5 premières échéances des contrats à terme sur le maïs du Chicago Board of Trade (CME), de 1988 à 2015.

	One-factor model		Two-factor model		Three-factor model		Four-factor model	
	original	de-noised	original	de-noised	original	de-noised	original	de-noised
μ	.0053	.0048	.0047	.0041	.0091	.0039	.0048	.0013
k_2			.144	.096	1.06	.074	1.27	2.19
k_3					1.18	2.65	.634	2.24
k_4							.49	2.21
σ_1	.13	.11	.23	.11	.17	.15	.12	.18
σ_2			.045	.042	.038	.041	.14	.23
σ_3					.087	.036	.092	.19
σ_4							.031	.18
λ_1	.014	−.0085	−.102	−.131	−.207	−.196	−.068	.026
λ_2			.064	.163	.129	.168	.071	.017
λ_3					.136	.145	.048	.029
λ_4							.092	−.056
p_{12}			−.27	.21	−.0208	−.0124	.72	.93
p_{13}					−.763	−.586	.38	.91
p_{14}							.26	−.90
p_{23}							−.92	.93
p_{24}							−.71	.89
p_{34}							.42	.94

Les estimations des paramètres suggèrent que le trend et la prime des risques sont très petit, ce qui était attendu en théorie. Le terme de diffusion σ_1 est cohérent avec les estimations antérieures trouvées dans la littérature économique.

En regardant le modèle avec filtrage par ondelette, la principale différence est que la prime de

risque est maintenant proche de zéro. Cela peut être interprété comme une preuve que la prime de risque est captée par une variation à très court terme. Comme prévu, La convergence de l'estimation s'améliore, car la variance des données filtrées par ondelette est inférieure à celle des données originelles.

Le modèle à trois facteurs fournit un ajustement supérieur aux données. Les résultats suggérant que la variable non stationnaire a un drift faible mais non nulle car ainsi qu'une diffusion importante, tandis que la vitesse de retour à la moyenne pour les deux autres variables d'état est rapide et cohérent avec les résultats précédents [13]. Les trois paramètres de prime de risque confirment les résultats de la littérature que l'impact global des primes de risque sur les contrats à terme sur matières premières agricoles est petit.

La variable d'état non stationnaire est faiblement corrélée à la première variable d'état de retour à la moyenne, mais fortement corrélée négativement avec la deuxième. Les deux variables d'état stationnaire ne sont que faiblement corrélées.

Enfin, nous considérons les résultats de l'estimation d'un modèle à quatre facteurs, Les estimations filtrées par ondelettes sont globalement meilleures. En particulier, les paramètres de retour à la moyenne de diffusions et prix du risque sur le marché prennent des valeurs plus raisonnables. Cependant, les coefficients de corrélation sont plus élevés en valeur absolue.

2.1.2 Erreur quadratique moyenne du modèle (RMSE)

TABLE 2.2 – Ce tableau présente les calculs des erreurs quadratiques moyennes de suivi en échantillon (RMSE, en pourcentage) pour les modèles à un à quatre facteurs appliqués aux prix des contrats à terme sur le maïs, en utilisant des données quotidiennes pour les cinq premières échéances.

Maturity	Original data				De-noised data (wavelet)			
	1-factor	2-factor	3-factor	4-factor	1-factor	2-factor	3-factor	4-factor
First	1.25	1.12	0.931	0.824	1.07	0.817	0.732	0.711
Second	1.36	1.44	1.04	0.943	1.13	0.921	0.776	0.728
Third	1.47	1.52	1.26	1.08	1.43	1.13	0.964	0.943
Fourth	1.81	1.76	1.32	1.14	1.48	1.38	1.26	1.22
Fifth	1.90	1.81	1.38	1.18	1.86	1.81	1.30	1.27

L'utilisation d'un modèle à un facteur conduit invariablement à de mauvaises performances. L'estimation des modèles en utilisant des données préalablement débruitées à l'aide du seuillage par ondelettes conduit à de meilleurs résultats, comme le montre le tableau 2.2 L'amélioration des performances est plus visible pour les échéances les plus proches, et moins impressionnant pour les échéances plus lointaines.

2.1.3 Pr vision

TABLE 2.3 – Ce tableau pr sente les calculs des erreurs quadratiques moyennes de suivi en  chantillon (RMSE, en pourcentage) pour les mod les   un   quatre facteurs appliqu s aux prix des contrats   terme sur le ma s, en utilisant des donn es quotidiennes pour les cinq premi res  ch ances.

Maturity	Original data				De-noised data (wavelet)			
	1-factor	2-factor	3-factor	4-factor	1-factor	2-factor	3-factor	4-factor
First	2.79	1.62	1.38	1.34	2.05	1.31	1.10	1.06
Second	3.02	1.64	1.35	1.32	2.34	1.36	1.08	1.05
Third	3.40	1.82	1.39	1.35	3.06	1.65	1.15	1.12
Fourth	3.64	1.96	1.51	1.47	3.32	1.83	1.23	1.20
Fifth	3.98	2.00	1.83	1.76	3.78	1.89	1.47	1.41

Pour compl ter les preuves de l' chantillon, les auteurs calculent  galement dans le tableau 2.3 quelques r sultats de pr visions hors  chantillon.

Pour cela, ils utilisent toutes les donn es disponibles   la date t afin de pr voir la date $t+k$, o  $k=5$ jours. Apr s avoir mis de c t  les 250 derni res observations, les auteurs calculent   l'aide du mod le espace d' tat des pr visions   5 jours et calculent les RMSE   l'aide de ces pr visions. Nous constatons que les RMSE sont plus  lev s pour les r sultats hors  chantillon que pour les r sultats dans l' chantillon.

De plus, les performances sont am lior es   mesure que nous passons de mod les monofactoriels   des mod les multifactoriels. Ils montrent  galement que les r sultats sont sensiblement am lior s lorsque nos donn es sont pr alablement d bruit    l'aide du seuillage par ondelettes.

En r sum , les auteurs montrent que le d bruitage des donn es brutes conduit au suivi et des r sultats de pr vision qui sont au moins aussi bons que les r sultats utilisant les donn es bruit es, au moins pour les cinq premi res  ch ances.

Chapitre 3

Nos résultats empiriques

3.1 Estimation du modèle espace d'état

3.1.1 Problèmes rencontrés

Le problème majeur auxquels nous avons fait face est la précision de l'estimation. En effet, à partir du modèle à deux facteurs la matrice de variance covariance devient particulièrement instable ce qui compromet la significativité des coefficients estimés. Pire encore, cette instabilité rend l'estimation impossible à mesure que le nombre d'itération augmente.

Une version précédente de l'article publié par Gabriel J. Power en 2008 indiquait : "The present paper reports only the preliminary results of this research, which need to be further refined and revised, because at this stage they are numerically unstable and lead to possibly unreliable standard errors (and thus, hypothesis tests), even if one computes the robust QML standard errors".

Après plusieurs tentatives (initialisation de la variance avec une valeur importante, exclusion des premières observations, discussion avec l'auteur de l'article), nous n'avons à ce jour pas trouver de moyen de stabiliser l'estimation. Les résultats qui suivent sont donc à remettre en perspective car sont issue d'estimation avec peu d'itérations.

3.1.2 Erreur quadratique moyenne du modèle (RMSE)

TABLE 3.1 – Ce tableau présente les calculs des erreurs quadratiques moyennes de suivi en échantillon (RMSE, en pourcentage) pour les modèles à un à quatre facteurs appliqués aux prix des contrats à terme sur le maïs, en utilisant des données quotidiennes pour les cinq premières échéances.

Maturity	Original data				De-noised data (wavelet)			
	1-factor	2-factor	3-factor	4-factor	1-factor	2-factor	3-factor	4-factor
First	1.76	0.748	0.862	1.85	1.71	0.721	0.837	1.78
Second	1.68	0.920	1.19	3.26	1.62	0.893	1.18	3.17
Third	1.74	1.12	1.53	4.55	1.68	1.08	1.51	4.42
Fourth	1.93	1.28	1.85	5.78	1.86	1.23	1.82	5.63
Fifth	2.24	1.51	2.21	7.02	2.15	1.45	2.18	6.83

En l'absence de convergence on observe tout de même plus ou moins les résultats des auteurs à savoir que les ondelettes permettent de réduire la RMSE du modèle et que l'augmentation de la maturité augmente la RMSE. En revanche, en l'absence de convergence du modèle nous ne retrouvons pas l'amélioration de la RMSE par l'augmentation du nombre de facteur.

3.1.3 Prédiction

TABLE 3.2 – Ce tableau présente les calculs des erreurs quadratiques moyennes de suivi en échantillon (RMSE, en pourcentage) pour les modèles à un à quatre facteurs appliqués aux prix des contrats à terme sur le maïs, en utilisant des données quotidiennes pour les cinq premières échéances.

Maturity	Original data				De-noised data (wavelet)			
	1-factor	2-factor	3-factor	4-factor	1-factor	2-factor	3-factor	4-factor
First	4.58	5.43	5.39	4.46	4.59	5.46	5.42	4.47
Second	4.47	5.37	5.19	3.09	4.49	5.40	5.23	3.14
Third	4.37	5.29	5.00	1.81	4.39	5.33	5.06	1.88
Fourth	4.28	5.24	4.85	0.61	4.31	5.30	4.93	0.71
Fifth	4.22	5.25	4.76	0.56	4.27	5.32	4.86	0.46

Chapitre 4

Conclusion

Dans cette étude, nous avons exploré l'impact du seuillage par ondelettes sur la modélisation de la courbe des contrats futures sur le maïs. En répliquant l'article "*Catching the Curl : Wavelet Thresholding Improves Forward Curve Modelling*", nous avons démontré que cette technique de débruitage améliore significativement l'ajustement et la prédiction des modèles espace d'état appliqués aux prix des contrats futures.

Notre analyse empirique a révélé plusieurs points clés :

- **Débruitage des données** : L'application du seuillage par ondelettes a permis de réduire la variance des données, entraînant des estimations plus stables et fiables des paramètres du modèle. Ce débruitage est particulièrement crucial pour les données de prix futures, souvent perturbées par des fluctuations à court terme. Les changements de prix contiennent à la fois de l'information et du bruit, et ce bruit réduit l'efficacité des estimations et la précision des prévisions de la structure par terme.
- **Estimation du modèle** : Les résultats montrent que les modèles avec trois et quatre facteurs fournissent un meilleur ajustement aux données que les modèles à un ou deux facteurs, surtout lorsque les données sont débruitées. Les paramètres de diffusion et les coefficients de réversion à la moyenne sont cohérents avec la littérature existante, confirmant la pertinence du modèle choisi. L'innovation méthodologique réside dans le débruitage au niveau des coefficients d'ondelettes plutôt que des données brutes, ce qui évite le sur-lissage ou le sous-lissage.
- **Erreur quadratique moyenne (RMSE)** : La performance du modèle, mesurée par la RMSE, s'est améliorée de manière significative après le débruitage des données. Les gains de performance sont particulièrement notables pour les échéances les plus proches, où la précision des prédictions est essentielle pour les stratégies de couverture et de trading. Nous avons réussi à montrer que l'augmentation du nombre de facteurs dans le modèle réduit la RMSE, ce qui indique une meilleure précision, bien que cela augmente également les coûts de calcul.
- **Prévisions hors échantillon** : Les prévisions à cinq jours réalisées avec des données débruitées ont également montré une réduction substantielle des erreurs par rapport aux prévisions faites avec des données brutes. Cela souligne l'importance du débruitage pour des prévisions robustes et fiables dans le contexte des marchés de futures.

En revanche, nous n'avons pas réussi à démontrer de manière claire que la RMSE diminue en fonction de l'échéance du contrat à terme. L'estimation devient plus complexe à mesure que la maturité s'éloigne, ce qui souligne les défis inhérents à la modélisation de la structure à terme pour des échéances plus lointaines.

Tant bien que mal, nous avons réussi à répliquer certaines conclusions principales du papier original. Bien que tout n'ait pas été parfait et que nous n'ayons pas collé exactement aux résultats du papier, nous avons pu en capturer les idées principales. Globalement, nous avons confirmé que le modèle avec plus de facteurs offre une meilleure précision, bien qu'à un coût computationnel plus élevé. Ces résultats montrent la pertinence de l'approche et suggèrent des pistes pour des améliorations futures.

Pour conclure, les traders et les gestionnaires de portefeuille ont un grand intérêt à mieux comprendre l'ensemble de la structure à terme des futures. Améliorer la modélisation des courbes à terme des matières premières est essentiel pour les décisions de budgétisation du capital et de gestion des risques des entreprises, en particulier celles impliquées dans le pétrole et le gaz ou d'autres matières premières (Geman, 2009).

Ainsi, cette étude confirme que le seuillage par ondelettes peut être une technique très utile pour aider à améliorer la modélisation de la courbe forward des contrats futures. L'intégration de cette méthode dans les modèles espace d'état permet non seulement de mieux capturer la dynamique des prix en fournissant des prévisions plus précises et fiables, mais aussi d'améliorer la gestion des risques, ce qui permet aux acteurs du marché de prendre des décisions plus éclairées et optimales. Ces résultats ouvrent la voie à des applications plus larges de la technique de seuillage, particulièrement pour les marchés caractérisés par une forte volatilité et des données bruitées.

Annexe

.1 Term Structure of Corn Futures Prices

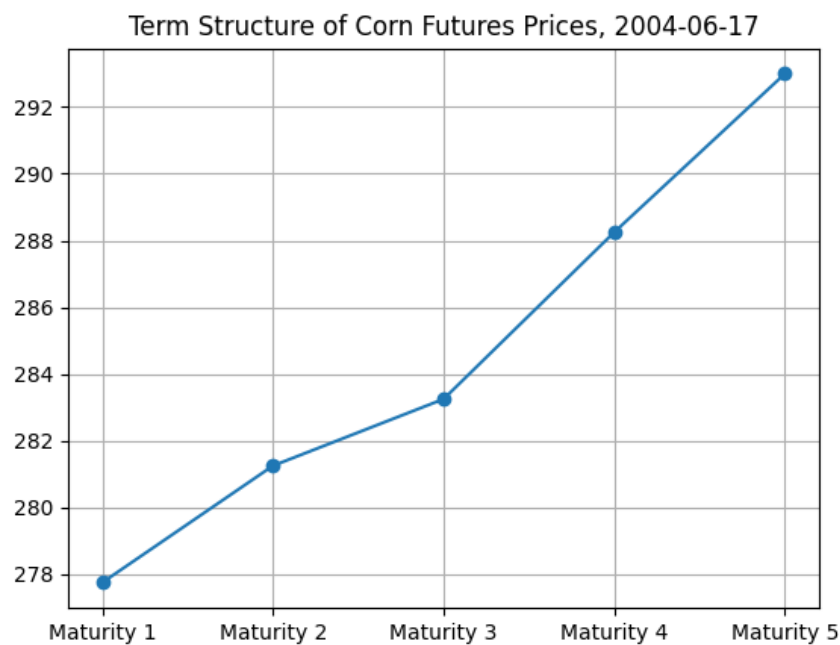


FIGURE 1 – Term Structure of Corn Futures Prices - 17/06/2004. Source : Bloomberg.

.2 Plot of front-month corn futures daily settlement price

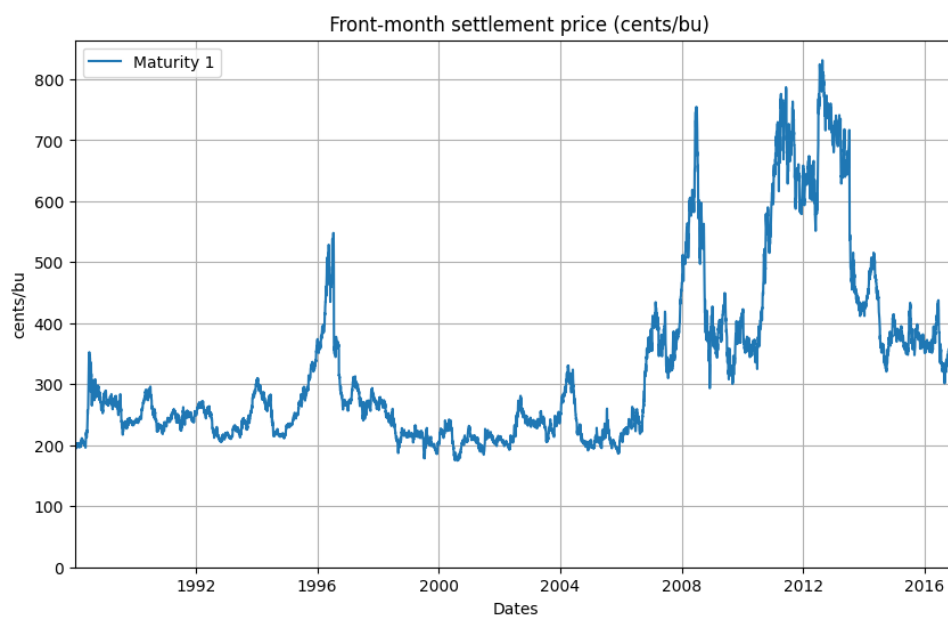


FIGURE 2 – Time series plot of front-month corn futures price 1988-2017. Source : Bloomberg.

.3 Sorensen Parameters Estimate (2002)

	Corn futures	Soybean futures	Wheat futures
μ	0.0416 (0.0206)	0.0433 (0.0267)	0.0531 (0.0336)
κ	0.7744 (0.0239)	1.0366 (0.0299)	0.7640 (0.0333)
σ	0.1585 (0.0052)	0.1785 (0.0050)	0.1794 (0.0063)
v	0.2201 (0.0077)	0.2363 (0.0072)	0.2253 (0.0099)
ρ	-0.3116 (0.0427)	-0.1344 (0.0379)	-0.2590 (0.0523)
α	-0.0386 (0.0025)	-0.0204 (0.0027)	-0.0028 (0.0036)
λ_2	-0.1011 (0.0376)	-0.0292 (0.1472)	0.0002 (0.0131)
x_1	4.8738 (0.0558)	5.7225 (0.1403)	4.9257 (0.0399)
σ_ϵ	0.0171 (0.0001)	0.0187 (0.0001)	0.0178 (0.0002)
γ_1	-0.0228 (0.0003)	-0.0182 (0.0003)	0.0162 (0.0003)
γ_1^*	0.0081 (0.0003)	0.0085 (0.0003)	0.0202 (0.0003)
γ_2	0.0029 (0.0002)	0.0031 (0.0003)	-0.0057 (0.0003)
γ_2^*	0.0054 (0.0004)	0.0058 (0.0002)	-0.0081 (0.0004)
log-likelihood function	21856	27591	18654

FIGURE 3 – Sorensen Parameters Estimate (2002) - Seasonality in Agricultural Commodity Futures

Bibliographie

- [1] BRENNAN et SCHWARTZ. « Evaluating natural resource investments ». In : *J. Bus.* 58.2 (1985), p. 135-157 (page 1).
- [2] GIBSON et SCHWARTZ. « Stochastic convenience yield and the pricing of oil contingent claims ». In : *J. Financ.* 45 (1990), p. 959-976 (page 1).
- [3] SCHWARTZ. « The stochastic behavior of commodity prices: implications for valuation and hedging ». In : *J. Financ.* 52 (1997), p. 923-973 (page 1).
- [4] SCHWARTZ et SMITH. « Short-term variations and long-term dynamics in commodity prices ». In : *Manag. Sci.* 46 (2000), p. 893-911 (page 1).
- [5] Julio J. LUCIA et Eduardo S. SCHWARTZ. « Electricity Prices and Power Derivatives: Evidence from the Nordic Power Exchange ». In : *Review of Derivatives Research* 5 (jan. 2002), p. 5-50 (page 1).
- [6] Samuel HIKSPOORS et Sebastian JAIMUNGAL. « Energy Spot Price Models and Spread Options Pricing ». In : *International Journal of Theoretical and Applied Finance* 10 (07 2007), p. 1111-1135 (page 1).
- [7] J. HASBROUCK. « High frequency quoting: short-term volatility in bids and offers ». In : (2013) (page 1).
- [8] D. L. DONOHO et I. M. JOHNSTONE. « Ideal de-noising in an orthonormal basis chosen from a library of bases ». In : *CRAS Paris, Ser I* 319 (1994), p. 1317-1322 (page 2).
- [9] D. B. PERCIVAL et D. MONDAL. « A Wavelet Variance Primer ». In : 30 (2012). Sous la dir. de T. SUBBA RAO, S. SUBBA RAO et C. R. RAO, p. 623-657 (page 2).
- [10] I. DAUBECHIES. *Ten Lectures on Wavelets*. Philadelphia : Society for Industrial et Applied Mathematics, 1992 (page 2).
- [11] I. DAUBECHIES. « Orthonormal bases of compactly supported wavelets II: Variations on a theme ». In : *SIAM J. Math. Anal.* 24.2 (1993), p. 499-519 (page 2).
- [12] Brani VIDAKOVIC. *Statistical Modeling by Wavelets*. New York : John Wiley & Sons, 1999 (page 2).
- [13] Carsten SORENSEN. « Seasonality in Agricultural Commodity Futures ». In : *J. Futur.Mark* 22 (2002), p. 223-232 (pages 4, 7, 9).
- [14] G. CORTAZAR et L. NARANJO. « An N-factor Gaussian model of oil futures prices. » In : *J. Futur.Mark* 26 (2006), p. 243-268 (page 4).
- [15] STEIN. « Speculative price: economic welfare and the idiot of chance ». In : *Rev. Econ.Stat.* (1981), p. 223-232 (page 5).
- [16] DONOHO et JOHNSTONE. « Ideal de-noising in an orthonormal basis chosen from a library of bases ». In : *CRAS Paris, Ser I* 319 (1994), p. 1317-1322.
- [17] DONOHO et JOHNSTONE. « Ideal spatial adaptation by wavelet shrinkage ». In : *Biometrika* 81 (1994), p. 425-455.

- [18] DONOHO et JOHNSTONE. « Wavelet shrinkage: asymptopia ». In : *J. R. Stat. Soc. Ser. B* 57.2 (1995), p. 301-369.