

UNIVERSITÉ LAVAL

ACT-7006 : SUJETS SPÉCIAUX I

Rapport

Par :
François PELLETIER

Remis à
M. Andrew Luong

2 novembre 2013

Table des matières

1	Introduction	3
2	Courbes paramétriques	4
2.1	Survol de la méthode et utilisation	4
2.2	Courbe de Nelson-Siegel	4
2.3	Analyse empirique	5
2.4	Conclusion	9
3	Analyse de composantes principales	10
3.1	Survol de la méthode et utilisation	10
3.2	Analyse empirique	11
3.2.1	Approche avec covariances	12
3.2.2	Approche avec corrélation	13
3.2.3	Comparaison des deux approches	14
3.2.4	Volatilité expliquée par composante (score)	14
3.3	Conclusion	15
4	Méthode des moments	16
4.1	Description de la méthode	16
4.2	Modèles de taux d'intérêt court-terme	16
4.3	Discrétisation	18
4.4	Conditions de moments utilisés	18
4.5	Matrice de pondération de Newey-West	19
4.6	Jacobien des moments	21
4.7	Statistique de Student (t)	22
4.8	Données utilisées	23
4.9	Applications	24
4.10	Conclusion	24
5	Méthode du maximum de vraisemblance	25
5.1	Description de la méthode	25
5.2	Application au modèle de Vasicek	25
5.3	Application au modèle CIR	26
5.4	Application au modèle CIR avec approximation normale	27
5.5	Analyse empirique	27
5.6	Conclusion	27
6	Bibliographie	28
7	Annexes	29
7.1	Fichier source R pour courbes de Nelson Siegel	29
7.2	Fichier source R pour PCA	30
7.3	Fichiers source R pour MMG	33
7.3.1	Fonctions	33
7.3.2	Fichier d'exécution	39

7.4 Fichiers source R pour EMV	41
--	----

1 Introduction

Ce travail porte sur différents types d'analyses empiriques pouvant être faites sur des séries de taux d'intérêt. Il vise surtout un survol de certaines des méthodes décrites dans [1] et [2]. Pour le chapitre 15, je traiterai des courbes de Nelson-Siegel pour modéliser les structures à terme. Pour le chapitre 16, je m'en tiendrai aux rudiments de l'analyse par composantes principales. Pour le chapitre 17, j'aborderai plus en détail les méthodes des moments généralisées et je terminerai par des estimations par maximum de vraisemblance.

2 Courbes paramétriques

2.1 Survol de la méthode et utilisation

La méthode de modélisation de la structure à terme basée sur des courbes paramétriques vise à fournir une approximation des taux d'intérêt à partir de points connus et observables de cette courbe, que l'on peut retrouver notamment via des services d'informations financières tels Bloomberg et CRSP. Ces courbes ne font aucunement allusion à une distribution statistique particulière et fonctionnent avec des données en coupe transversale. Pour chaque temps t , on peut calculer l'équation de la courbe, et ce indépendamment des informations passées, qui n'interviennent pas dans ces modèles. Ces modèles n'ont aucune valeur prédictive et servent strictement à faire de l'interpolation afin de peaufiner l'ajustement de modèles de taux d'intérêt tels que les modèles de la famille HJM et les modèles basés sur le modèle de Vasicek, ainsi que tout autre modèle basé sur un taux initial connu et fixé. On retrouve dans cette famille les modèles basés sur des splines cubiques et splines de lissage, ainsi que la famille de modèles basés sur les travaux de Nelson et Siegel.

Une propriété commune à ces modèles est qu'ils sont constitués d'une fonction de base $\phi_k(\tau)$ et de paramètres λ_k , composant les différents termes d'une somme, comprenant K termes, plus ou moins selon l'ajustement désiré. La forme générale est la suivante :

$$\delta(\tau) = \sum_{k=1}^K \lambda_k \phi_k(\tau)$$

L'estimation de splines est basée sur la correspondance des premières dérivées et donne des résultats précis dont on peut borner exactement l'erreur maximale, car ce sont de simples fonctions polynomiales. Cependant, ces courbes ont un comportement qui ne répond pas aux exigences d'une courbe de structure à terme car elles ne sont pas robustes, en plus de comporter de nombreux paramètres (4 par intervalle).

Quant à elles, les courbes de la famille de Nelson-Siegel ont l'avantage d'être parcimonieuses, le modèle de base comportant quatre paramètres. Elles sont construites à partir de monômes, de polynômes multipliés par une forme exponentielle. Elles sont considérées adéquates dans la plupart des situations où la courbe n'est pas complexe. Comme ces courbes ne passent pas par tous les points connus, il faut par contre les estimer par une méthode du type moindres carrés.

Dans la section suivante, on ne traitera que de la courbe de Nelson-Siegel dans sa forme originale. Par contre, il en existe plusieurs généralisations dont les courbes de Svensson, Wiseman et Bjork and Christensen.

2.2 Courbe de Nelson-Siegel

La courbe de Nelson-Siegel prend la forme suivante, pour les taux à terme :

$$f_0(\tau, t) = \beta_{0t} + (\beta_{1t} + \beta_{2t}\tau)e^{-\beta_{3t}\tau}$$

Et de même pour les taux instantanés

$$r(\tau, t) = \beta_{0t} + (\beta_{1t} + \beta_{2t}\tau)e^{-\beta_{3t}\tau}$$

Elle peut à la fois être utilisée pour modéliser la structure à terme des taux instantanés (spot rate) ou encore celle des taux à terme (forward rate). Si on utilise la forme précédente pour modéliser le taux à terme, on retrouve alors la forme suivante pour les taux instantanés, qui représentent une moyenne de taux à terme.

$$\begin{aligned} r(\tau, t) &= \frac{1}{\tau} \int_0^\tau f_0(s) ds \\ &= \beta_0 + (\beta_{1t} + \frac{\beta_{2t}}{\beta_{3t}}) \frac{1 - e^{-\beta_{3t}\tau}}{\beta_{3t}\tau} - \frac{\beta_{2t}}{\beta_{3t}} e^{-\beta_{3t}\tau} \end{aligned}$$

On remarque une certaine ressemblance entre la forme de cette équation et la structure donnée par le modèle de Vasicek, mais en pratique elle se limite à la présence de la forme exponentielle, et formellement les deux modèles ont peu de caractéristiques en commun.

2.3 Analyse empirique

On utilise pour cette étude empirique les taux composite des bons du trésor américain entre le 3 janvier 1990 et le 15 février 2012, pris à un intervalle de 30 jours ouvrables, c'est-à-dire 192 observations. Les taux considérés ici sont pour les obligations de 90 jours, 2 ans, 3 ans, 5 ans et 10 ans. Seules ces séries ont été utilisées car elles sont complètes pour toutes les journées d'ouverture des marchés.

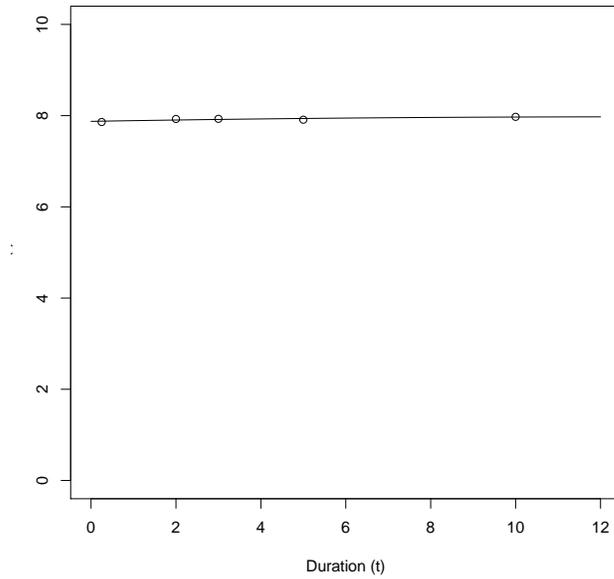
L'estimation a été effectuée à l'aide de la méthode des moindres carrés ordinaires, c'est-à-dire sans utiliser de pondération par la matrice de variance-covariance.

Si on note $O_\tau^{(t)}$ le taux instantané observé pour la durée τ au temps t , on obtient l'équation d'estimation suivante au temps t :

$$\sum_{\tau} (O_\tau^{(t)} - r(\tau, t))^2$$

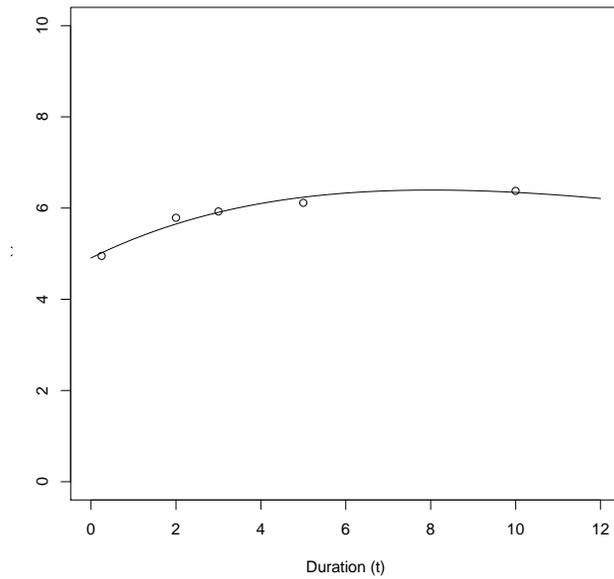
On doit minimiser la valeur de cette expression pour obtenir les meilleurs estimateurs de $\beta = [\beta_0, \beta_1, \beta_2, \beta_3]$. On effectue l'optimisation de la fonction d'estimation pour chaque temps t pour lequel on désire estimer la courbe. On peut ensuite utiliser les paramètres estimés pour tracer un graphique d'une courbe en particulier et effectuer une comparaison visuelle de la performance du modèle.

Temps 1



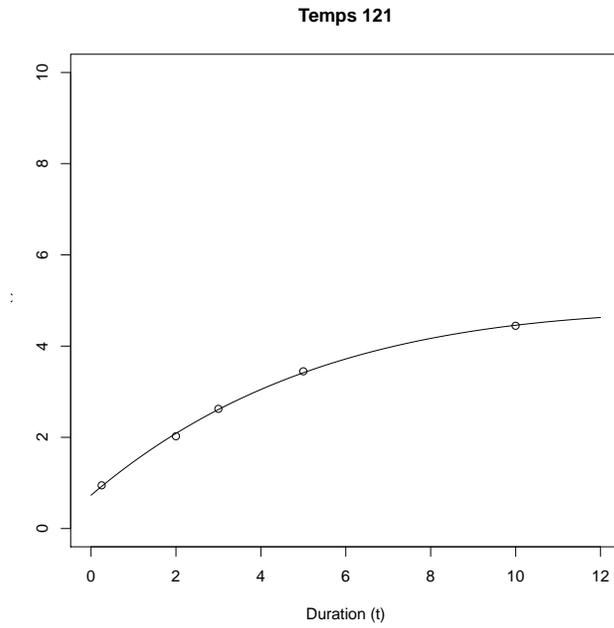
On remarque au temps 1 que la courbe est relativement plate, elle s'ajuste très bien aux données dans cette situation.

Temps 61

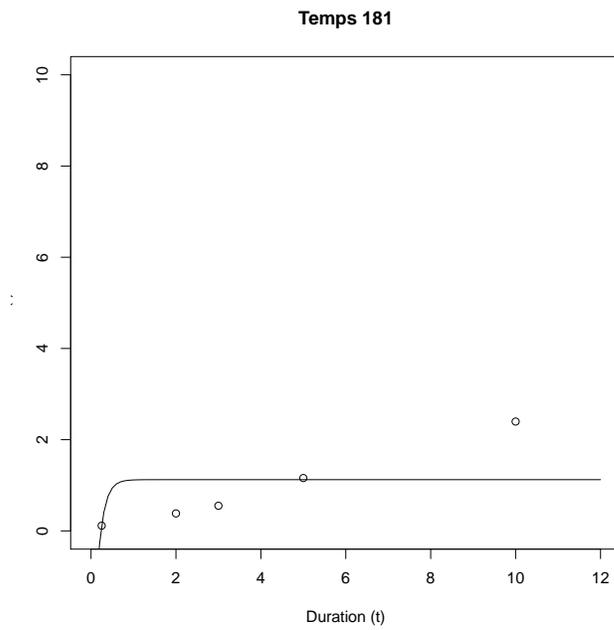


Au temps 61, la courbe est concave, et on remarque un bon ajustement par rapport aux

données observées.

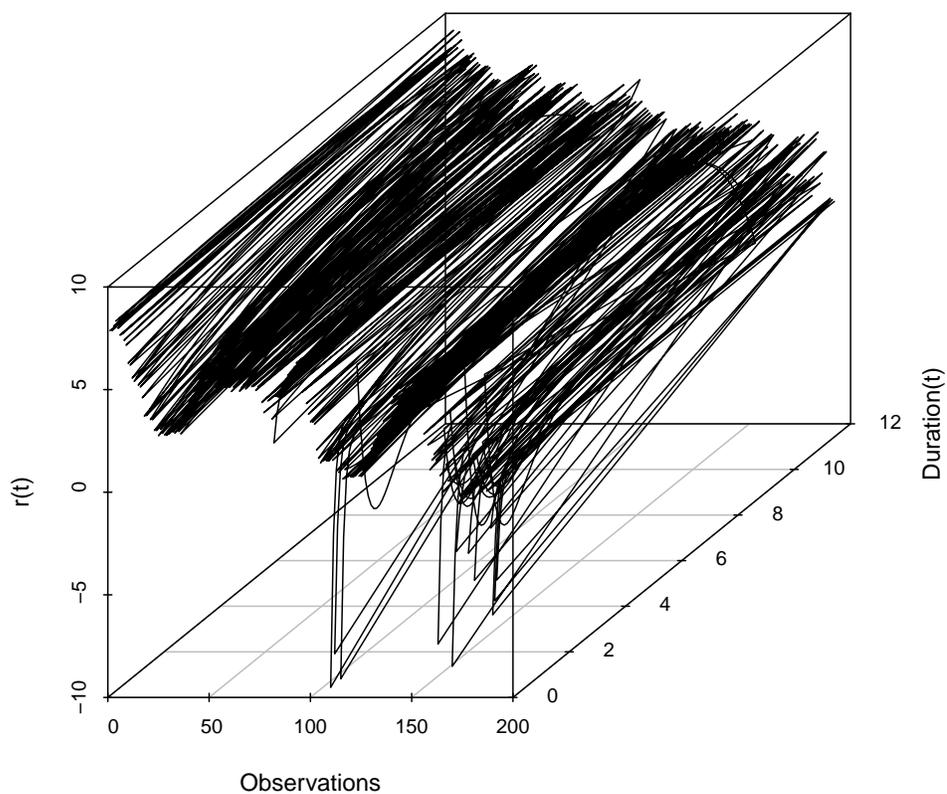


Au temps 121, la courbe est aussi concave, et on remarque encore un bon ajustement par rapport aux données observées.



Au temps 181, la courbe devrait plutôt avoir une forme convexe, on observe qu'ici, en présence de taux très faibles pour une durée de 90 jours, que l'ajustement n'est pas du tout satisfaisant. C'est d'ailleurs une des limites de la courbe de Nelson-Siegel de ne pas bien s'adapter à des structures à terme convexes ou irrégulières, ayant des points d'inflexion par exemple.

En résumé, voici les courbes pour les 192 observations, de gauche à droite



2.4 Conclusion

Les courbes paramétriques sont utiles pour faire de l'interpolation et donner une approximation de la structure à terme. Cependant, dans le cas des splines et des courbes de Nelson-Siegel, ce ne sont pas des modèles robustes et ils peuvent facilement mener à des résultats erronés si les données n'ont pas certaines caractéristiques de régularité. Cependant, les courbes de Nelson-Siegel sont faciles à estimer et selon les résultats observés, fonctionnent bien la majorité du temps.

3 Analyse de composantes principales

3.1 Survol de la méthode et utilisation

La méthode d'analyse de composantes principales est une méthode générale qui vise à identifier différents facteurs qui peuvent causer la volatilité à l'intérieur d'une série chronologique multivariée de taux d'intérêt pour des durations données. Ces composantes peuvent ensuite être représentées dans un modèle à facteurs multiples de la classe affine par exemple. Cependant, cette analyse est particulièrement complexe puisqu'elle implique de modéliser des processus stochastiques sur des espaces non euclidiens. Pour ce rapport, seule la méthode d'identification des composantes sera expliquée, et quelques conclusions pourront en être tirées.

Différenciation On considère les observations sous la forme $r_{t_i}(\tau_j)$ où t_i est le temps, de 1 à $n+1$, et j est le nombre d'observations en coupe transversale (nombre de séries de taux d'intérêt).

Comme on cherche à modéliser la volatilité, on doit alors travailler sur des données différenciées. Nous devons d'abord calculer les différences $d_{i,j}$, que nous considérons comme des observations de la variable aléatoire d_j , c'est-à-dire la variable représentant l'intensité des sauts à l'intérieur d'une série.

$$d_{i,j} = r_{t_{i+1}}(\tau_j) - r_{t_i}(\tau_j)$$

Matrice de variance-covariance On calcule ensuite la matrice de variance-covariance de ces observations Σ , qui prend la forme :

$$\Sigma = \begin{bmatrix} \text{var}(d_1) & \text{cov}(d_1, d_2) & \cdots & \text{cov}(d_1, d_k) \\ \text{cov}(d_2, d_1) & \text{var}(d_2) & \ddots & \vdots \\ \vdots & & \ddots & \vdots \\ \text{cov}(d_k, d_1) & \cdots & \cdots & \text{var}(d_k) \end{bmatrix}$$

On peut aussi calculer la matrice de corrélation correspondante et poursuivre l'analyse avec cette matrice.

Valeurs et vecteurs propres On doit maintenant trouver une matrice \mathbf{P} telle que sa transposée est également son inverse.

Cette matrice \mathbf{P} est la matrice de vecteurs propres de Σ . On a aussi un vecteur λ contenant les valeurs propres de chacun des vecteurs propres (colonnes) de \mathbf{P} .

La matrice \mathbf{P} est la matrice de composantes principales et le vecteur λ est le vecteur des variances de chacune des composantes principales, en ordre décroissant. Il est important de noter que les composantes principales sont orthogonales entre elles et donc que la covariance entre chacune d'elles est nulle.

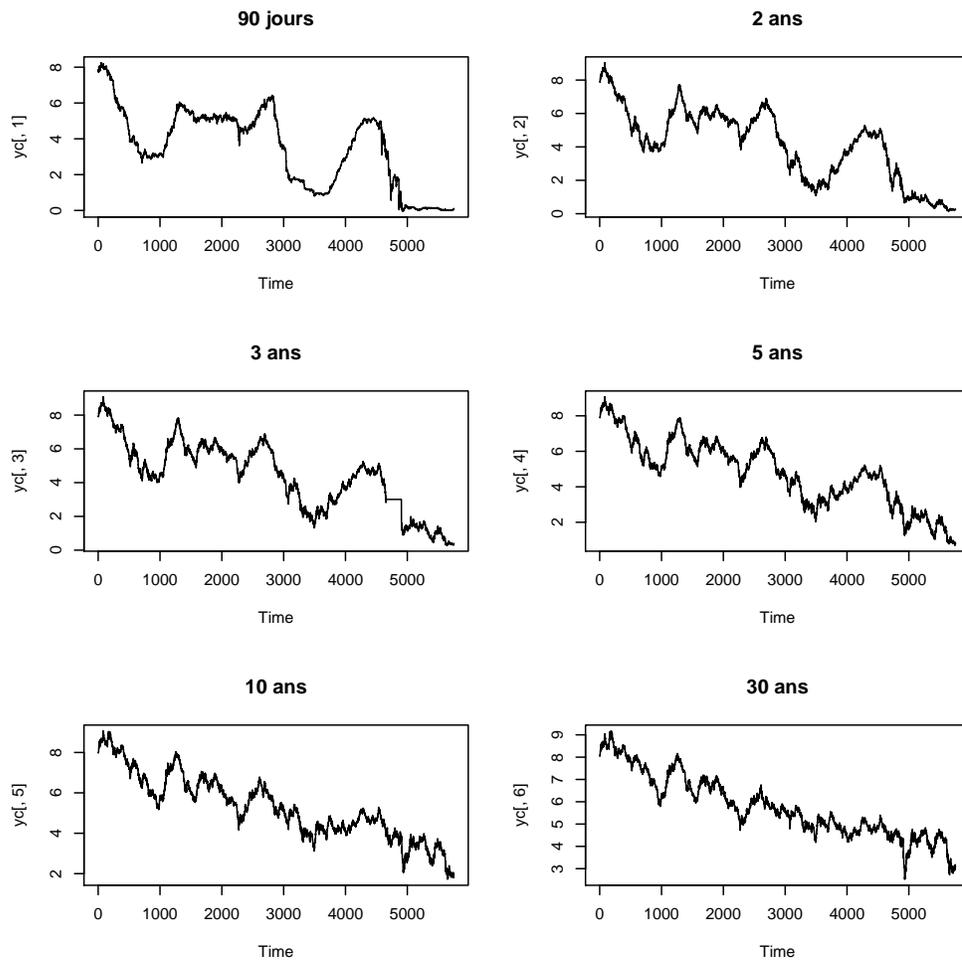
Une fois les composantes principales obtenues, il s'agit ensuite de les visualiser graphiquement pour les interpréter. La littérature affirme que les taux d'intérêts peuvent être

expliqués à l'aide des trois premières composantes principales (trois premières colonnes de la matrice \mathbf{P}).

3.2 Analyse empirique

On utilise pour cette étude empirique les taux composite des bons du trésor américain entre le 3 janvier 1990 et le 15 février 2012, pris à un intervalle quotidien. Les taux considérés ici sont pour les obligations de 90 jours, 2 ans, 3 ans, 5 ans, 10 ans et 30 ans.

Voici les séries utilisées



3.2.1 Approche avec covariances

En utilisant l'approche par covariance, nous obtenons la matrice \mathbf{P} de composantes principales suivante :

	1	2	3	4	5	6
1	-0.1576	0.9315	-0.2613	-0.1936	-0.0424	0.0027
2	-0.4388	0.1746	0.3688	0.5881	0.5189	0.1611
3	-0.4304	-0.0499	0.5864	-0.6842	0.0141	-0.0083
4	-0.4993	-0.0455	-0.0019	0.3105	-0.5847	-0.5570
5	-0.4584	-0.1913	-0.3438	-0.0074	-0.2998	0.7383
6	-0.3713	-0.2465	-0.5777	-0.2282	0.5450	-0.3443

et le vecteur de valeurs propres λ suivant :

	1	2	3	4	5	6
1	0.0170	0.0025	0.0012	0.0006	0.0002	0.0001

Nous pouvons en déduire que chacune des composantes explique cette proportion de la variance totale :

	1	2	3	4	5	6
1	0.7846	0.1167	0.0576	0.0278	0.0095	0.0037

3.2.2 Approche avec corrélation

En utilisant l'approche par corrélation, nous obtenons la matrice \mathbf{P} de composantes principales suivante :

	1	2	3	4	5	6
1	-0.1957	0.9453	-0.2133	0.1457	-0.0362	0.0017
2	-0.4367	0.1117	0.4137	-0.5828	0.5041	0.1785
3	-0.4226	-0.0719	0.5529	0.7141	0.0234	-0.0047
4	-0.4611	-0.0585	0.0472	-0.3013	-0.5541	-0.6196
5	-0.4526	-0.1715	-0.2958	-0.0381	-0.4083	0.7142
6	-0.4182	-0.2364	-0.6229	0.1921	0.5199	-0.2723

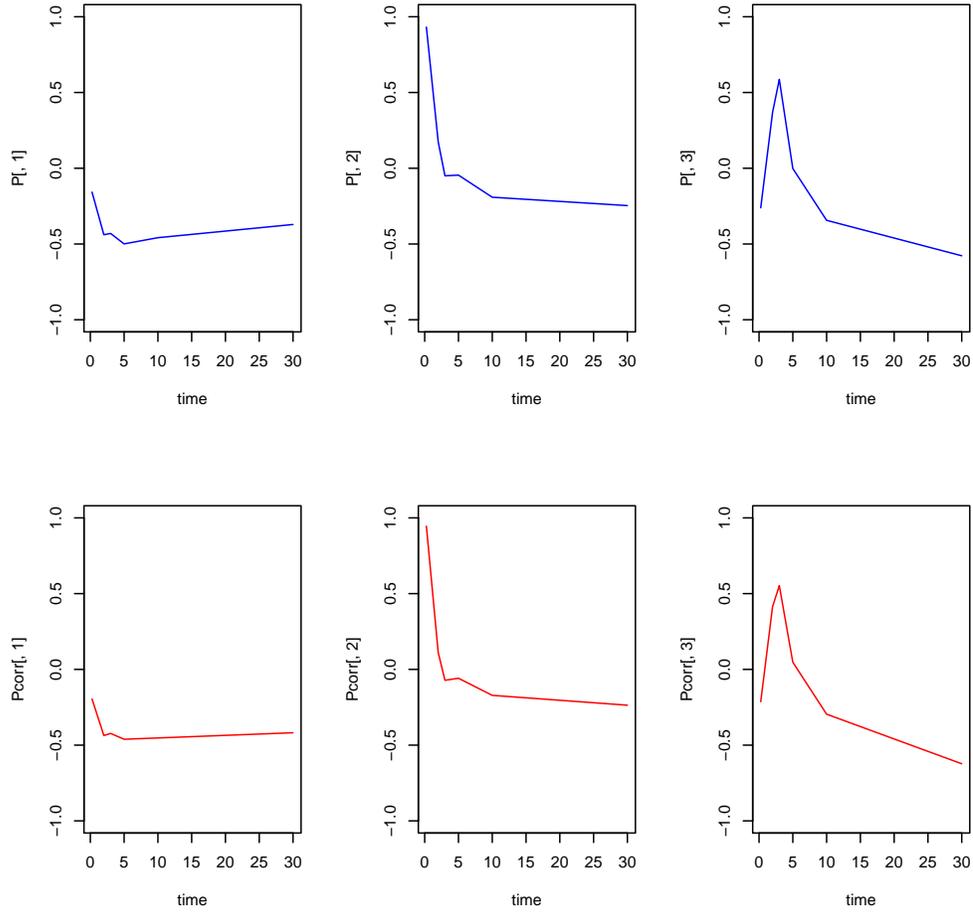
et le vecteur de valeurs propres λ suivant :

	1	2	3	4	5	6
1	4.5010	0.9038	0.3615	0.1584	0.0548	0.0205

Nous pouvons en déduire que chacune des composantes explique cette proportion de la corrélation totale :

	1	2	3	4	5	6
1	0.7502	0.1506	0.0602	0.0264	0.0091	0.0034

3.2.3 Comparaison des deux approches



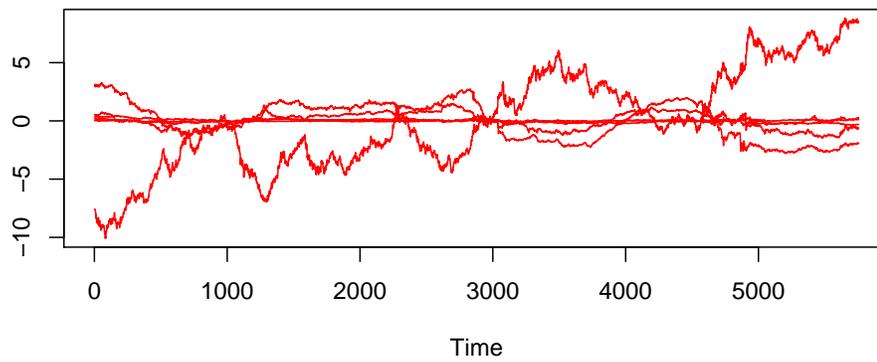
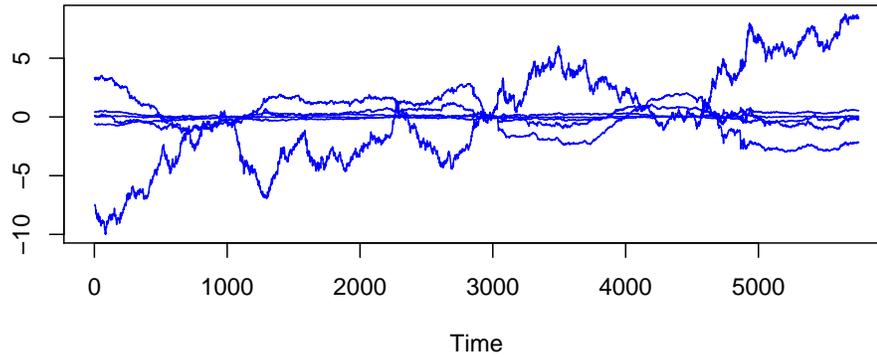
On observe sur ce graphique les trois composantes principales pour la méthode avec la covariance et la méthode avec la corrélation. On remarque que les deux méthodes donnent des résultats similaires mais non identiques.

3.2.4 Volatilité expliquée par composante (score)

En utilisant des séries chronologiques centrées sur leur moyenne, on peut construire une fonction appelée score qui exprime, pour chaque composante, la proportion de volatilité expliquée, sous forme de série chronologique.

$$Score(t) = r(t) \times \mathbf{P}$$

Pour les deux méthodes utilisées, on obtient les graphiques suivants :



3.3 Conclusion

Sur les graphiques précédents, on remarque bien que se détachent principalement trois composantes qui expliquent la grande majorité de la variance, est ce tout au long de la série temporelle. Dans la littérature financière, ces composantes sont souvent appelés, dans l'ordre d'importance, Parallel Shift, Tilt et Flex (Curvature). La première est relativement plate et représente les changements qui affectent l'ensemble de la courbe. La seconde explique les variations à court terme, et leur influence inverse sur les variations à long terme, et enfin, la dernière exprime la tendance qu'a la structure à terme à prendre une forme plus ou moins concave.

4 Méthode des moments

4.1 Description de la méthode

La méthode des moments généralisés est une technique d'estimation paramétrique introduite entre autres par Hansen (1982). Elle consiste à construire des fonctions d'estimation basées sur les moments empiriques qui vérifient une condition d'orthogonalité. Ainsi, le vecteur optimal de paramètres sera celui qui minimisera la valeur de la fonction, ou qui du moins l'approchera le plus de 0 dans l'éventualité de données négatives.

Formellement, on a un vecteur de k paramètres $\theta = (\theta_1, \dots, \theta_k)'$ auquel on définit $f = (f_1, \dots, f_m)$, un vecteur de $m, m \geq k$ fonctions $f_i(r_t|\theta)$ de l'échantillon r_t , qui sera pour nous une série chronologie de taux court-terme. Ces fonctions sont telles que $E[f_i(r_t|\theta)] = 0$.

On obtient le vecteur de paramètres estimés $\hat{\theta}$ en minimisant la fonction objectif $\hat{\theta} = \text{argmin}(\theta, f' \times W \times f)$ pour une matrice de pondération définie positive W . Pour la méthode des moindres carrés ordinaires, la matrice W est la matrice identité. Elle peut aussi être l'inverse de la matrice de variance-covariance empirique des moments, ou encore l'estimateur robuste de Newey and West (1987), qui tient compte de l'autocorrélation et de l'hétéroscédasticité des données.

Enfin, on pose les tests diagnostiques habituels avec le coefficient d'ajustement est les statistiques de Student pour vérifier la validité de l'estimation.

4.2 Modèles de taux d'intérêt court-terme

Les modèles suivants sont des modèles couramment utilisés en pratique pour modéliser le taux d'intérêt à court terme, souvent à des fins d'évaluations de produits financiers, notamment des obligations et des produits dérivés.

Modèle de Vasicek Le modèle de Vasicek (1977) est un des modèles les plus simples utilisés en pratique pour modéliser les taux d'intérêt à court terme. Il est basé sur un processus de Ornstein–Uhlenbeck. Une des caractéristique de ce modèle est qu'il présente une forme de retour à la moyenne.

$$dr_t = a(b - r_t) dt + \sigma dW_t \quad (1)$$

(1) présente le modèle sous la forme d'une équation différentielle stochastique, où la moyenne à long terme est b , et où a est un paramètre qui fixe la vitesse à laquelle le processus a tendance à retourner à sa moyenne. La volatilité instantanée du processus est représentée par le paramètre σ qui multiplie la variation du processus bruit blanc dW_t . La variance du processus à long terme est donnée par $\frac{\sigma^2}{2a}$. On peut retrouver ces valeurs en appliquant le lemme d'Îto, qui nous donne la moyenne et la variance du processus au temps t , et en prenant la limite lorsque $t \rightarrow \infty$.

La solution de l'équation différentielle stochastique est la variable aléatoire suivante :

$$r(t) = r(0)e^{-at} + b(1 - e^{-at}) + \sigma e^{-at} \int_0^t e^{as} dW_s.$$

La moyenne et la variance de $r(t)$ sont dérivées facilement à partir de cette équation :

$$\begin{aligned} E[r_t] &= r_0 e^{-at} + b(1 - e^{-at}) \\ \text{Var}[r_t] &= \frac{\sigma^2}{2a} (1 - e^{-2at}) \\ \lim_{t \rightarrow \infty} E[r_t] &= b \\ \lim_{t \rightarrow \infty} \text{Var}[r_t] &= \frac{\sigma^2}{2a} \end{aligned}$$

Modèle de Cox, Ingersoll et Ross Le modèle de Cox, Ingersoll et Ross (1985) (ci-après CIR) est un modèle basé sur le modèle de Vasicek et le concept de retour à la moyenne. Cependant, il n'utilise pas le même processus pour la volatilité, car le modèle de Vasicek peut produire des taux d'intérêt négatif, ce qui n'est pas possible en pratique selon les hypothèses des marchés efficients (Par contre, on a retrouvé à divers moments, des taux d'intérêt négatifs depuis la crise financière de 2008, résultant d'anomalies d'ordre calculatoires, étant donné les taux très bas de cette période.) Le modèle CIR est plutôt basé sur le processus racine carrée.

$$dr_t = a(b - r_t) dt + \sigma \sqrt{r_t} dW_t \quad (2)$$

(2) représente le modèle CIR sous sa forme d'équation différentielle stochastique. Les paramètres ont la même signification que dans le modèle de Vasicek. La moyenne et la variance de la variable aléatoire qui résout l'équation différentielle stochastique sont respectivement :

$$\begin{aligned} E[r_t|r_0] &= r_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) \\ \text{Var}[r_t|r_0] &= r_0 \frac{\sigma^2}{\theta} (e^{-\theta t} - e^{-2\theta t}) + \frac{\mu\sigma^2}{2\theta} (1 - e^{-\theta t})^2 \end{aligned}$$

Modèle de Chan, Karolyi, Longstaff et Sanders Le modèle de Chan, Karolyi, Longstaff et Sanders (1992) est une généralisation de plusieurs modèles de taux d'intérêt à court terme. Il constitue ainsi une base pour évaluer les différents modèles et les comparer entre eux. Sa particularité, par rapport au modèle CIR, est que l'exposant de r_t est aussi un paramètre du modèle, γ , alors qu'il est fixé dans le modèle de CIR ($\gamma = 0.5$). Pour le modèle de Vasicek, on a ($\gamma = 0$).

$$dr_t = a(b - r_t) dt + \sigma r_t^\gamma dW_t \quad (3)$$

4.3 Discrétisation

Afin de calibrer ces modèles, nous faisons appel à la discrétisation, étant donné que le modèle sera calibré sur des données discrètes. Une des façons les plus simples de discrétiser une équation différentielle stochastique est la méthode d'Euler. Pour le modèle CKLS, on obtient (4)

$$r_{t+1} = a + br_t + \sigma r_t^\gamma u_{t-1} \quad (4)$$

Si on fixe

$$\begin{aligned} \epsilon_{t+1} &= r_{t+1} - (a + br_t) \\ &= \sigma r_t^\gamma u_{t-1} \end{aligned}$$

On obtient alors une variable aléatoire ϵ_{t+1} qui suit une loi normale centrée de variance $\sigma^2 r_t^{2\gamma} \Delta t$. De plus, ϵ_{t+1} n'est pas corrélé avec r_t . Ces deux conditions vont permettre de construire les fonctions d'estimation pour la méthode des moments. Pour les modèles de Vasicek et CIR, il suffit de remplacer la valeur de γ par 0 ou 0.5 respectivement.

4.4 Conditions de moments utilisés

À partir des conditions précédentes, nous pouvons facilement construire au moins quatre équations basées sur les moments de ϵ_{t+1} . Comme les moments d'ordre supérieur peuvent produire un phénomène d'oscillation en présence de données extrêmes, on préférera utiliser des moments de premier et second ordre. On a donc les moments en (5) et (6).

$$E[\epsilon_{t+1}] = 0 \quad (5)$$

$$E[\epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \Delta t] = 0 \quad (6)$$

On peut aussi utiliser le fait que ϵ_{t+1} n'est pas corrélé avec r_t (7 et 8).

$$E[\epsilon_{t+1} r_t] = 0 \quad (7)$$

$$E[(\epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \Delta t) r_t] = 0 \quad (8)$$

Pour une valeur de θ donnée, on peut calculer les moments empiriques équivalents (9),(10),(11) et (12).

$$f_1 = \frac{1}{n} \sum_{i=1}^n (r_{t+1} - a - br_t) \quad (9)$$

$$f_2 = \frac{1}{n} \sum_{i=1}^n (r_{t+1} - a - br_t)^2 - \sigma^2 r_t^{2\gamma} \Delta t \quad (10)$$

$$f_3 = \frac{1}{n} \sum_{i=1}^n (r_{t+1} - a - br_t) r_t \quad (11)$$

$$f_4 = \frac{1}{n} \sum_{i=1}^n ((r_{t+1} - a - br_t) - \sigma^2 r_t^{2\gamma} \Delta t) r_t \quad (12)$$

On rassemble ces équations en une somme de moindres carrée (méthode des moindres carrés ordinaires) et on obtient l'équation d'estimation (13)

$$J(a, b, \sigma, \gamma) = f'f = \sum_{i=1}^4 f_i^2 \quad (13)$$

Cependant, cette équation fait une hypothèse forte d'absence de corrélation et d'homoscédasticité des erreurs, que nous allons éviter en utilisant les moindres carrés pondérés (14).

$$J(a, b, \sigma, \gamma) = f'Wf \quad (14)$$

Une matrice optimale pour $W = S^{-1}$ est la matrice qui minimise la covariance. Il est suggéré d'utiliser la matrice d'information de Fisher empirique (15) de l'estimateur f .

$$\hat{\Omega}_0 = \frac{1}{T} \sum_{t=1}^T \hat{f}_t^2 \quad (15)$$

Cependant, il existe un meilleur choix qui tient compte à la fois de l'autocovariance et de l'hétéroscédasticité des erreurs, l'estimateur de Newey et West (1987). C'est cet estimateur que nous utiliserons.

4.5 Matrice de pondération de Newey-West

L'estimateur de Newey-West [3] prend la forme (16)

$$\hat{S}_T = \sum_{j=1}^m \left(1 - \frac{j}{m-1}\right) [\hat{\Omega}_j + \hat{\Omega}_j'] \quad (16)$$

$$\hat{\Omega}_j = \frac{1}{T} \sum_{t=j+1}^T \hat{f}_t \hat{f}_{t-j} \quad (17)$$

Où m est le nombre de pas de temps de décalage utilisés, et $\hat{\Omega}_j$ est une matrice d'autocovariance entre la série de données et la même série mais décalée de j pas de temps (17). Le rôle du coefficient $(1 - \frac{j}{m-1})$ est de s'assurer que la matrice est semi-définie positive. L'estimateur, sans ce facteur de pondération, donne parfois des résultats erronés. C'était l'estimateur proposé dans Hansen (1982) lors de la proposition du modèle GMM.

4.6 Jacobien des moments

Nous aurons besoin, pour former les statistiques de diagnostic, du gradient des moments empiriques. Pour le modèle CKLS :

$$\frac{df}{d\theta} = \begin{bmatrix} \frac{\partial f_1}{\partial a} & \frac{\partial f_1}{\partial b} & \frac{\partial f_1}{\partial \sigma} & \frac{\partial f_1}{\partial \gamma} \\ \frac{\partial f_2}{\partial a} & \frac{\partial f_2}{\partial b} & \frac{\partial f_2}{\partial \sigma} & \frac{\partial f_2}{\partial \gamma} \\ \frac{\partial f_3}{\partial a} & \frac{\partial f_3}{\partial b} & \frac{\partial f_3}{\partial \sigma} & \frac{\partial f_3}{\partial \gamma} \\ \frac{\partial f_4}{\partial a} & \frac{\partial f_4}{\partial b} & \frac{\partial f_4}{\partial \sigma} & \frac{\partial f_4}{\partial \gamma} \end{bmatrix}$$

$$\frac{df}{d\theta} = \begin{bmatrix} -2 \sum_{t=1}^{T-1} (r_{t+1} - a - br_t) & -\sum_{t=1}^{T-1} (r_t) & 0 & 0 \\ -\sum_{t=1}^{T-1} (r_t) & -2 \sum_{t=1}^{T-1} ((r_{t+1} - a - br_t) r_t) & -\Delta t \sum_{t=1}^{T-1} (r_t^{2\gamma}) & -2\sigma^2 \Delta t \sum_{t=1}^{T-1} (\log(r_t) r_t^{2\gamma}) \\ -2 \sum_{t=1}^{T-1} ((r_{t+1} - a - br_t) r_t) & -\sum_{t=1}^{T-1} (r_t^2) & 0 & 0 \\ -2 \sum_{t=1}^{T-1} ((r_{t+1} - a - br_t) r_t) & -2 \sum_{t=1}^{T-1} ((r_{t+1} - a - br_t) r_t^2) & -\Delta t \sum_{t=1}^{T-1} (r_t^{2\gamma+1}) & -2\sigma^2 \Delta t \sum_{t=1}^{T-1} (\log(r_t) r_t^{2\gamma+1}) \end{bmatrix}$$

4.7 Statistique de Student (t)

On utilise ici la méthode delta pour construire la matrice de variance-covariance des paramètres à partir de celle des équations (estimateur de Newey-West) (18)

$$V = \frac{df}{d\hat{\theta}} W \frac{df'}{d\hat{\theta}} \quad (18)$$

On prend la diagonale de cette matrice, qui contient les variances des estimateurs, pour calculer les statistiques de Student

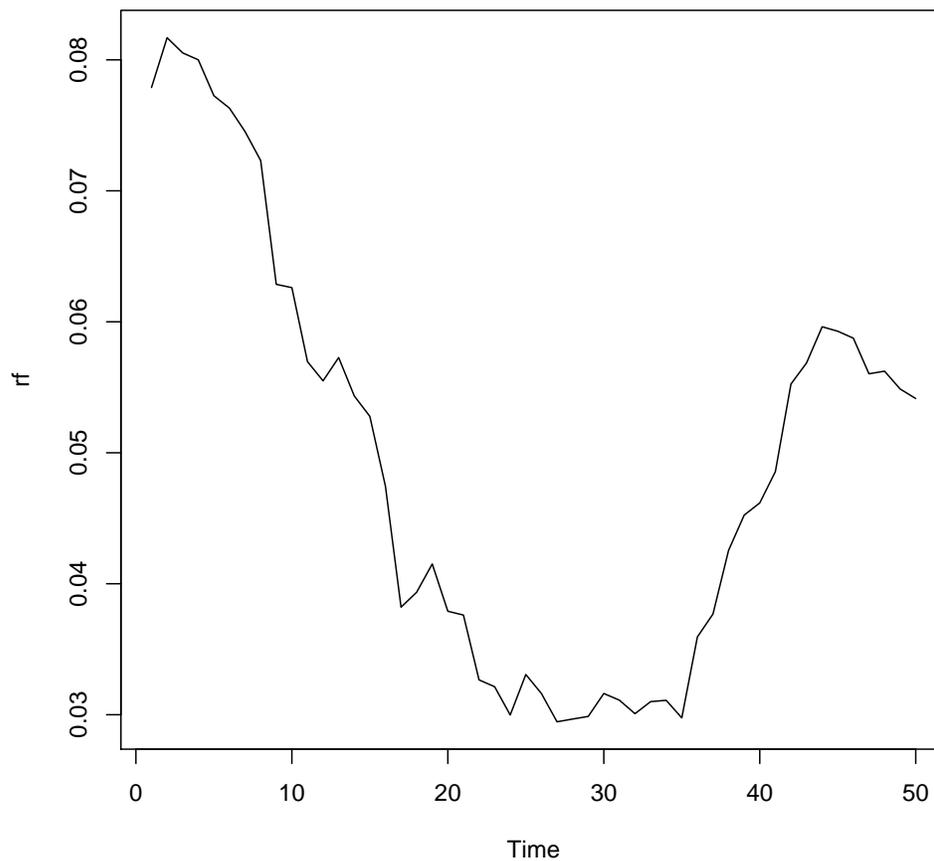
$$t = \frac{\theta_i}{\sqrt{V_{ii}}}$$

4.8 Données utilisées

Les données utilisées sont des données mensuelles entre les dates suivantes

[1] "1990-01-03" "1995-10-18"

La série se décrit visuellement comme suit :



Les statistiques descriptives de base de la série sont :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02946	0.03276	0.04800	0.04930	0.05838	0.08169

Ces statistiques peuvent servir de point de comparaison pour évaluer la pertinence des résultats d'un modèle étant donné que l'on travaille avec des méthodes numériques.

4.9 Applications

Les résultats des estimations, les statistiques de test ainsi que les matrices de variance-covariance des paramètres, pour chacun des modèles de taux d'intérêt.

	Est. param.	T-Stat	p-value
a	0.02320	22.30001	0.00000
b	-0.58973	14.27554	0.00000
sigma	0.03416	72953440.38807	0.00000
gamma	0.96593	372870417.22069	0.00000

TABLE 1 – Modèle CKLS estimé avec GMM

	Est. param.	T-Stat	p-value
a	0.02320	21.94401	0.00000
b	-0.58973	1.59672	0.05859
sigma	0.00011	10.98051	0.00000

TABLE 2 – Modèle Vasicek estimé avec GMM

	Est. param.	T-Stat	p-value
a	0.02320	22.28222	0.00000
b	-0.58973	1.62643	0.05535
sigma	0.00227	2.39356	0.01041

TABLE 3 – Modèle CIR estimé avec GMM

On remarque que les modèles semblent cohérents les uns envers les autres, notamment par rapport à la moyenne à long terme ainsi qu'à la vitesse de retour à l'équilibre.

4.10 Conclusion

L'application de la méthode est facile à implémenter mais il semble avoir des difficultés au niveau du calcul de la matrice de variance-covariance qui explose pour le modèle CKLS. Cette méthode donne tout du moins des estimateurs semblables, sauf en ce qui concerne la volatilité. On peut donc dire que cette analyse est cohérente avec l'intuition que l'on peut avoir des modèles de taux d'intérêt à court terme.

	a	b	sigma	gamma
a	0.00010	-0.00138	0.00007	-0.00001
b	-0.00138	0.02636	-0.00267	0.00052
sigma	0.00007	-0.00267	0.00402	-0.00079
gamma	-0.00001	0.00052	-0.00079	0.00015

TABLE 4 – Matrice de Var-Cov des par. pour modèle CKLS avec GMM

	a	b	sigma
a	0.00010	-0.00138	0.00000
b	-0.00138	0.02636	-0.00001
sigma	0.00000	-0.00001	0.00000

TABLE 5 – Matrice de Var-Cov des par. pour modèle Vasicek avec GMM

5 Méthode du maximum de vraisemblance

5.1 Description de la méthode

La méthode du maximum de vraisemblance cherche les paramètres qui maximisent la probabilité que l'échantillon obtenu provienne de la distribution en question. La fonction de vraisemblance $L(\theta)$ est la densité conjointe de tous les éléments de l'échantillon. On utilise souvent le logarithme de la fonction de vraisemblance $\ln L(\theta)$ comme fonction à maximiser car elle est plus simple à évaluer et à différencier. Pour une description de la méthode en profondeur et les propriétés statistiques des estimateurs on peut consulter toute monographie de statistique mathématique.

5.2 Application au modèle de Vasicek

La résolution de l'équation différentielle stochastique qui caractérise le modèle de Vasicek permet d'identifier la distribution (19) du taux à un temps $t_2, t_2 > t_1$ sachant la valeur du taux au temps t_1 .

$$r_{t_2}|r_{t_1} \sim N\left(\mu + (r_{t_1} - \mu)e^{-\alpha\Delta t}, \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\Delta t})\right) \quad (19)$$

$N(\mu, \sigma^2)$ est distribution normale de moyenne μ et de variance σ^2 . On voit donc minimiser (20) en θ , où $f(\cdot)$ est la densité de la loi normale de r_{t_2} .

$$-\sum_{i=1}^{n-1} \ln f(\cdot|r_{t_1}, \theta) \quad (20)$$

La solution est obtenue directement en utilisant un algorithme d'optimisation numérique.

	a	b	sigma
a	0.00010	-0.00138	0.00000
b	-0.00138	0.02636	-0.00018
sigma	0.00000	-0.00018	0.00001

TABLE 6 – Matrice de Var-Cov des par. pour modèle CIR avec GMM

5.3 Application au modèle CIR

La résolution de l'équation différentielle stochastique qui caractérise le modèle CIR permet d'identifier la distribution (21) du taux à un temps $t_2, t_2 > t_1$ sachant la valeur du taux au temps t_1 .

$$p(t_2, r_{t_2}; t_1, r_{t_1} | \theta) = ce^{-u-\nu} \left(\frac{\nu}{u}\right)^{\frac{q}{2}} I_q(2\sqrt{u\nu}) \quad (21)$$

$I_q()$ est la fonction de Bessel modifiée de type 1

$$I_q(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-i(q\tau - x \sin \tau)} d\tau.$$

Et on retrouve les constantes suivantes :

$$\begin{aligned} c &= \frac{2\alpha}{\sigma^2(1 - e^{-\alpha\Delta t})} \\ u &= cr_{t_1}e^{-\alpha\Delta t} \\ \nu &= cr_{t_2} \\ q &= \frac{2\alpha\mu}{\sigma} - 1 \end{aligned}$$

Cette fonction de distribution peut aussi être vue comme une χ^2 non centrée (22). Cette approche est plus facile à programmer étant donné qu'elle ne requiert pas d'approximation de la fonction de Bessel modifiée de type 1.

$$r_{t_2} | r_{t_1} \sim \chi^2(2cr_{t_2}; 2q + 2, 2u) \quad (22)$$

On doit donc minimiser (23) en θ , où $f()$ est la densité de la loi χ^2 non centrée de r_{t_2} .

$$-\sum_{i=1}^{n-1} \ln f(\cdot | r_{t_1}, \theta) \quad (23)$$

La solution peut être obtenue directement en utilisant un algorithme d'optimisation numérique. Cet algorithme n'a pas été implémenté car il est numériquement très instable et cet estimation se fait habituellement avec d'autres méthodes dont le filtre de Kalman.

5.4 Application au modèle CIR avec approximation normale

L'approche par approximation normale pose comme hypothèse que la distribution de $r_{t_2}|r_{t_1}$ est normale, suivant la moyenne et la variance de la distribution de χ^2 non centrée.

$$\begin{aligned} E[r_{t_2}|r_{t_1}] &= r_{t_1}e^{-\alpha\Delta t} + \mu(1 - e^{-\alpha\Delta t}) \\ V[r_{t_2}|r_{t_1}] &= r_{t_1} \frac{\sigma^2}{\alpha} (e^{-\alpha\Delta t} - e^{-2\alpha\Delta t}) + \mu \frac{\sigma^2}{2\alpha} (1 - e^{-\alpha\Delta t})^2 \end{aligned}$$

5.5 Analyse empirique

Les données considérées sont les données de série chronologique mensuelles du taux composite des bons du trésor américains d'échéance 3 mois, entre les dates suivantes : [1] "1990-02-13" "1995-10-18"

Les paramètres estimés avec les deux méthodes sont resumés dans la table 7 :

	Param	Vasicek	CIR.QL
1	alpha	0.604539835935924	0.642262538740856
2	mu	0.0393379847421409	0.0398958894018834
3	sigma	0.000117483898235261	0.00250688987001702

TABLE 7 – Paramètres estimés par maximum de vraisemblance pour 2 modèles

5.6 Conclusion

En effectuant un test de ratio de vraisemblance, pour H_0 : le modèle X est meilleur que le modèle Y, on constate que le modèle CIR ajusté avec l'approximation normale est meilleur que le modèle de Vasicek, avec une p-value de 6.694339e-06. On remarquera par contre dans la littérature que le modèle CIR, lorsque l'on surmonte les problèmes numériques, est meilleur que ces deux modèles.

6 Bibliographie

Références

- [1] J. James and N. Webber. *Interest rate modelling*. Wiley series in financial engineering. John Wiley & Sons, 2000.
- [2] T.L. Lai and H. Xing. *Statistical models and methods for financial markets*. Springer texts in statistics. Springer, 2008.
- [3] W K Newey and K D West. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 1987.

7 Annexes

7.1 Fichier source R pour courbes de Nelson Siegel

```
library("scatterplot3d")

##
## implémentation de la courbe de Nelson-Siegel
##

## Lecture des taux spot
taux <- t(read.table("usgg.csv",header=T, sep=";",na.strings = "#NA")
          [seq(from=1,to=5758,by=30),2:6])
length(taux)
temps <- c(.25,2,3,5,10)

## fonction Nelson-Siegel originale
f0t <- function(PAR,t)
{
  PAR[1] + (PAR[2]+PAR[3]*t)*exp(-PAR[4]*t)
}

## fonction objectif
f0obj <- function(PAR,temps,taux)
{
  sum((taux-f0t(PAR,temps))^2)
}

## fonction qui retourne les paramètres
NelsonSiegel <- function(taux,temps)
{
  m1 <- optim(c(1,1,1,1),f0obj,gr=NULL,temps,taux)
  m1$par
}

## dimensions pour le graphique
duration <- seq(from=0,to=12,by=0.1)
timeline <- 1:dim(taux)[2]

## calcul des paramètres pour chaque courbe
parametres <- apply(taux,2,NelsonSiegel,temps)

## calcul des points du graphique
pointsgraph <- cbind(rep(timeline,each=length(duration)),duration,
                    as.vector(apply(parametres,2,f0t,duration)))
```

```

## tracer le graphique
pdf("nelsonsiegel-plots.pdf")
scatterplot3d(pointsgraph,type="l", pch=20,
              xlab="Observations",
              ylab="Duration(t)", zlab="r(t)")
plot(pointsgraph[pointsgraph[,1]==1,],[-1],type="l",main="Temps 1",
      xlab="Duration (t)",
      ylab="r(t)",ylim=c(0,10))
points(temps,taux[,1])
plot(pointsgraph[pointsgraph[,1]==61,],[-1],type="l",main="Temps 61",
      xlab="Duration (t)",
      ylab="r(t)",ylim=c(0,10))
points(temps,taux[,61])
plot(pointsgraph[pointsgraph[,1]==121,],[-1],type="l",main="Temps 121",
      xlab="Duration (t)",
      ylab="r(t)",ylim=c(0,10))
points(temps,taux[,121])
plot(pointsgraph[pointsgraph[,1]==181,],[-1],type="l",main="Temps 181",
      xlab="Duration (t)",
      ylab="r(t)",ylim=c(0,10))
points(temps,taux[,181])
dev.off()

```

7.2 Fichier source R pour PCA

```

## principal component analysis
## données .25,2,3,5,10,30 ans USGG pris sur Bloomberg

## librairies
library("xtable")

yc <- data.matrix(read.table("usgg.csv",header=T, sep=";",na.strings = "#NA"))[,,-1]
dt <- yc[,7]
yc <- yc[,,-7]
time <- c(.25,2,3,5,10,30)
yc.center <- t(t(yc)-apply(yc,2,mean))

## plot des séries
pdf("PCA-tseries.pdf")
par(mfrow=c(3,2))
ts.plot(yc[,1], main="90 jours")
ts.plot(yc[,2], main="2 ans")
ts.plot(yc[,3], main="3 ans")
ts.plot(yc[,4], main="5 ans")
ts.plot(yc[,5], main="10 ans")
ts.plot(yc[,6], main="30 ans")

```

```

dev.off()

## plot de l'observation 1
pdf("PCA-observation1.pdf")
plot(time,yc[1,],type="l")

## plot de l'observation 1 avec spline
lines(spline(time,yc[1,]),type="l", col="red")
dev.off()
## différenciation
yc.diff <- diff(yc.center)

##
## Approche #1 (cov)
##

## matrice de variance covariance des différences
yc.diff.vcov <- cov(yc.diff)

## matrice P (vecteurs propres)
P <- eigen(yc.diff.vcov)$vectors
sink("PCA-Pcov.tex",append=FALSE,split=FALSE)
xtable(P,digits=4)
sink()
## valeurs propres (variance par composant)
lambda <- eigen(yc.diff.vcov)$values
sink("PCA-lambdacov.tex",append=FALSE,split=FALSE)
xtable(t(as.matrix(lambda)),digits=4)
sink()
## écart-type par composant
sq.lambda <- sqrt(lambda)

## variance totale (trace de la matrice de valeurs propres)
totvar <- sum(lambda)

## fraction expliquée par composante
lambda/totvar
sink("PCA-prcov.tex",append=FALSE,split=FALSE)
xtable(t(as.matrix(lambda/totvar)),digits=4)
sink()

##
## Approche #2 (corr)
##

## matrice de corrélations des différences

```

```

yc.diff.corr <- cor(yc.diff)
## matrice P (vecteurs propres)
Pcorr <- eigen(yc.diff.corr)$vectors
sink("PCA-Pcorr.tex",append=FALSE,split=FALSE)
xtable(Pcorr,digits=4)
sink()
## valeurs propres (variance par composant)
lambda.corr <- eigen(yc.diff.corr)$values
sink("PCA-lambdacorr.tex",append=FALSE,split=FALSE)
xtable(t(as.matrix(lambda.corr)),digits=4)
sink()
## écart-type par composant
sq.lambda.corr <- sqrt(lambda.corr)
## variance totale (trace de la matrice de valeurs propres)
totvar.corr <- sum(lambda.corr)
## fraction expliquée par composante
lambda.corr/totvar.corr
sink("PCA-prcorr.tex",append=FALSE,split=FALSE)
xtable(t(as.matrix(lambda.corr/totvar.corr)),digits=4)
sink()

## graphiques (3 premières composantes: parallel shift, tilt, curvature)
pdf("PCA-composantes1-2-3.pdf")
par(mfrow=c(2,3))
plot(time,P[,1],col="blue", ylim=c(-1,1), type="l")
plot(time,P[,2],col="blue", ylim=c(-1,1), type="l")
plot(time,P[,3],col="blue", ylim=c(-1,1), type="l")
plot(time,Pcorr[,1],col="red", ylim=c(-1,1), type="l")
plot(time,Pcorr[,2],col="red", ylim=c(-1,1), type="l")
plot(time,Pcorr[,3],col="red", ylim=c(-1,1), type="l")
dev.off()

## produits Score (volatilité expliquée par composante)
Score <- yc.center %*% P
Scorecorr <- yc.center %*% Pcorr
pdf("PCA-score.pdf")
par(mfrow=c(2,1))
ts.plot(Score,col="blue")
ts.plot(Scorecorr,col="red")
dev.off()

## utilisation de la fonction princomp pour fins de comparaison
yc.princomp <- prcomp(yc.diff)
pdf("PCA-verif-princomp.pdf")
plot(yc.princomp, type="l")
dev.off()

```

7.3 Fichiers source R pour MMG

7.3.1 Fonctions

```
## general methods of moments (GMM)
library(MASS)

MMGpoidsNW <- function(param,modele)
{
  donnees <- modele$donnees
  q <- modele$q
  donneesF <- donnees[-1]
  donneesL <- donnees[-length(donnees)]
  deltaTemps <- modele$deltaTemps
  a <- param[1]
  b <- param[2]
  Gamma <- array(0,c(4,4,q+1))
  if(modele$nomModele=="CKLS")
  {
    sigma <- param[3]
    gamma <- param[4]
    g1t <- donneesF - a - b * donneesL
    g2t <- (donneesF - a - b * donneesL) ^ 2 - sigma^2 * donneesL ^ (2*gamma) *
      deltaTemps
    g3t <- (donneesF - a - b * donneesL) * donneesL
    g4t <- ((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * donneesL ^ (2*gamma) *
      deltaTemps) * donneesL
  }

  if(modele$nomModele=="CIR")
  {
    sigma <- param[3]
    g1t <- donneesF - a - b * donneesL
    g2t <- (donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * donneesL * deltaTemps
    g3t <- (donneesF - a - b * donneesL) * donneesL
    g4t <- ((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * donneesL * deltaTemps) *
      donneesL
  }

  if(modele$nomModele=="Vasicek")
  {
    sigma <- param[3]
    g1t <- donneesF - a - b * donneesL
    g2t <- (donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * deltaTemps
    g3t <- (donneesF - a - b * donneesL) * donneesL
    g4t <- ((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * deltaTemps) * donneesL
  }
}
```

```

    }
gt <- cbind(g1t,g2t,g3t,g4t)
n <- length(g1t)
## en attendant
#W <- solve(cov(gt))
## Newey-West
gtc <- apply(gt,2,function(x) x-mean(x))
for(v in 0:q)
  {
    gtF <- gtc[(1+v):n,]
    gtL <- gtc[1:(n-v),]
    Gamma[,,(v+1)] <- t(gtF) %*% gtL / n
  }
S <- Gamma[, ,1]
for(v in 1:q)
  {
    Snext <- (1-v/(q+1)) * (Gamma[, ,v+1] + t(Gamma[, ,v+1]))
    S <- S+Snext
  }
W <- ginv(S)
}

MomentsJacobien <- function(param, modele)
{
  donnees <- modele$donnees
  donneesF <- donnees[-1]
  donneesL <- donnees[-length(donnees)]
  n <- length(donneesL)
  deltaTemps <- modele$deltaTemps
  if(modele$nomModele=="CKLS")
  {
    a <- param[1]
    b <- param[2]
    sigma <- param[3]
    gamma <- param[4]

    g1a <- -n
    g2a <- -2*sum(donneesF - a - b*donneesL)
    g3a <- -sum(donneesL)
    g4a <- -2*sum((donneesF - a - b*donneesL)*donneesL)

    g1b <- -sum(donneesL)
    g2b <- -2*sum((donneesF - a - b*donneesL)*donneesL)
    g3b <- -sum(donneesL^2)
    g4b <- -2*sum((donneesF - a - b*donneesL)*donneesL^2)
  }
}

```

```

g1s <- 0
g2s <- -deltaTemps*sum(donneesL^(2*gamma))
g3s <- 0
g4s <- -deltaTemps*sum(donneesL^(2*gamma+1))

g1g <- 0
g2g <- -2*sigma^2*deltaTemps*sum(log(donneesL)*donneesL^(2*gamma))
g3g <- 0
g4g <- -2*sigma^2*deltaTemps*sum(log(donneesL)*donneesL^(2*gamma+1))

  d <- cbind(c(g1a, g1b, g1s, g1g),
             c(g2a, g2b, g2s, g2g),
             c(g3a, g3b, g3s, g3g),
             c(g4a, g4b, g4s, g4g))/n
}
if(modele$nomModele=="CIR")
{
  a <- param[1]
  b <- param[2]

  g1a <- -n
  g2a <- -2*sum(donneesF - a - b*donneesL)
  g3a <- -sum(donneesL)
  g4a <- -2*sum((donneesF - a - b*donneesL)*donneesL)

  g1b <- -sum(donneesL)
  g2b <- -2*sum((donneesF - a - b*donneesL)*donneesL)
  g3b <- -sum(donneesL^2)
  g4b <- -2*sum((donneesF - a - b*donneesL)*donneesL^2)

  g1s <- 0
  g2s <- -sum(deltaTemps*donneesL)
  g3s <- 0
  g4s <- -sum(deltaTemps*donneesL*donneesL)

  d <- cbind(c(g1a, g1b, g1s),
             c(g2a, g2b, g2s),
             c(g3a, g3b, g3s),
             c(g4a, g4b, g4s))/n
}
if(modele$nomModele=="Vasicek")
{
  a <- param[1]
  b <- param[2]

  g1a <- -n

```

```

g2a <- -2*sum(donneesF - a - b*donneesL)
g3a <- -sum(donneesL)
g4a <- -2*sum((donneesF - a - b*donneesL)*donneesL)

g1b <- -sum(donneesL)
g2b <- -2*sum((donneesF - a - b*donneesL)*donneesL)
g3b <- -sum(donneesL^2)
g4b <- -2*sum((donneesF - a - b*donneesL)*donneesL^2)

g1s <- 0
g2s <- -deltaTemps*n
g3s <- 0
g4s <- -sum(deltaTemps*donneesL)

d <- cbind(c(g1a, g1b, g1s),
           c(g2a, g2b, g2s),
           c(g3a, g3b, g3s),
           c(g4a, g4b, g4s))/n
}
d
}

MMGobjectif<- function(param, modele, W)
{
  donnees <- modele$donnees
  donneesF <- donnees[-1]
  donneesL <- donnees[-length(donnees)]
  n <- length(donnees)-2
  deltaTemps <- modele$deltaTemps
  a <- param[1]
  b <- param[2]

  if(modele$nomModele=="CKLS")
  {
    sigma <- param[3]
    gamma <- param[4]
    g1 <- sum(donneesF - a - b * donneesL)
    g2 <- sum((donneesF - a - b * donneesL) ^ 2 - sigma^2 * donneesL ^ (2*gamma) *
              deltaTemps)
    g3 <- sum((donneesF - a - b * donneesL) * donneesL)
    g4 <- sum(((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * donneesL ^ (2*gamma) *
              deltaTemps) * donneesL )
  }

  if(modele$nomModele=="CIR")
  {

```

```

    sigma <- param[3]
    g1 <- sum(donneesF - a - b * donneesL)
    g2 <- sum((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * donneesL * deltaTemps)
    g3 <- sum((donneesF - a - b * donneesL) * donneesL)
    g4 <- sum(((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * donneesL * deltaTemps) *
              donneesL)
  }

if(modele$nomModele=="Vasicek")
{
  sigma <- param[3]
  g1 <- sum(donneesF - a - b * donneesL)
  g2 <- sum((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * deltaTemps)
  g3 <- sum((donneesF - a - b * donneesL) * donneesL)
  g4 <- sum(((donneesF - a - b * donneesL) ^ 2 - sigma ^ 2 * deltaTemps) *
            donneesL)
}
g <- c(g1,g2,g3,g4)/n
t(g) %*% W %*% g
}

MMGestimation <- function(modele)
{
  deltaTemps <- modele$deltaTemps
  ## paramètres initiaux
  if(modele$nomModele=="CKLS")
  {
    alpha <- 0.5
    beta <- -0.5
    sigma <- 0.5
    gamma <- 0.5
    a <- alpha * deltaTemps
    b <- beta * deltaTemps + 1
    Initialparam <- c(a,b,sigma,gamma)
  }
  if(modele$nomModele=="CIR" || modele$nomModele=="Vasicek")
  {
    alpha <- 0.5
    beta <- -0.5
    sigma <- 0.5
    a <- alpha * deltaTemps
    b <- beta * deltaTemps + 1
    Initialparam <- c(a,b,sigma)
  }
  ## Première étape avec matrice identité
  W <- diag(4)

```

```

estim <- nlminb(Initialparam,MMGobjectif,gr=NULL,hessian=NULL,modele,W)
param <- estim$par
Fval <- estim$objective
Exitflag <- estim$convergence
if(modele$nomModele=="CKLS")
{
  Ralpha <- estim$par[1] / deltaTemps
  Rbeta <- (estim$par[2] - 1) / deltaTemps
  Rsigma2 <- estim$par[3] ^ 2
  Rgamma <- estim$par[4]
}
if(modele$nomModele=="CIR")
{
  Ralpha <- estim$par[1] / deltaTemps
  Rbeta <- (estim$par[2] - 1) / deltaTemps
  Rsigma2 <- estim$par[3] ^ 2
  Rgamma <- 0.5
}

if(modele$nomModele=="Vasicek")
{
  Ralpha <- estim$par[1] / deltaTemps
  Rbeta <- (estim$par[2] - 1) / deltaTemps
  Rsigma2 <- estim$par[3] ^ 2
  Rgamma <- 0
}
## Seconde étape avec matrice W
if(modele$Iterations>0)
{
  for (i in 1:modele$Iterations)
  {
    Initialparam <- param
    W <- MMGpoidsNW(param, modele)
    estim <- nlminb(Initialparam,MMGobjectif,gr=NULL,hessian=NULL,modele,W)
    param <- estim$par
    Fval <- estim$objective
    Exitflag <- estim$convergence
    if(modele$nomModele=="CKLS")
    {
      Ralpha <- estim$par[1] / deltaTemps
      Rbeta <- (estim$par[2] - 1) / deltaTemps
      Rsigma2 <- estim$par[3] ^ 2
      Rgamma <- estim$par[4]
    }
    if(modele$nomModele=="CIR")
    {

```

```

        Ralpha <- estim$par[1] / deltaTemps
        Rbeta  <- (estim$par[2] - 1) / deltaTemps
        Rsigma2 <- estim$par[3] ^ 2
        Rgamma <- 0.5
    }

    if(modele$nomModele=="Vasicek")
    {
        Ralpha <- estim$par[1] / deltaTemps
        Rbeta  <- (estim$par[2] - 1) / deltaTemps
        Rsigma2 <- estim$par[3] ^ 2
        Rgamma <- 0
    }
}

## Statistique T
n <- length(modele$donnees)-1
d <- MomentsJacobien(param,modele)
Varparam <- ginv(d %*% W %*% t(d))
Tstat <- solve(chol(Varparam),param)/sqrt(n)
Tpvalue <- 1-pt(Tstat,n-length(param))
list(Tstat=Tstat,Tpvalue=Tpvalue,
     Varparam=Varparam,param = c(Ralpha,Rbeta,Rsigma2,Rgamma), Fval=Fval,
     Exitflag=Exitflag)
}

```

7.3.2 Fichier d'exécution

```

source("gmm.r")
library(r2lh)
library(xtable)
## Execution

## mes données
rf <- data.matrix(read.table("usgg.csv",
                           header=T, sep=";",na.strings = "#NA"))[(1:50)*30,2]/100
sink("GMM-dates.txt")
as.vector(read.table("usgg.csv",
                   header=T, sep=";",na.strings = "#NA")[c(1,1500),1])
sink()

pdf("serieGMM.pdf")
ts.plot(rf)
dev.off()

```

```

sink("summaryDonneesGMM.txt",append=FALSE,split=FALSE)
summary(rf)
sink()

## donnees ckls
rf2 <- scan("ckls.csv")/100

ModeleCKLS <- list(donnees=rf,nomModele="CKLS",
                  deltaTemps=1/12, Iterations=2, q=12)
ModeleVASICEK <- list(donnees=rf,nomModele="Vasicek",
                     deltaTemps=1/12, Iterations=2, q=12)
ModeleCIR <- list(donnees=rf,nomModele="CIR",
                  deltaTemps=1/12, Iterations=2, q=12)
nomsParam <- c("a","b","sigma","gamma")
nomsParam2 <- c("a","b","sigma")
m1 <- MMGestimation(ModeleCKLS)
m2 <- MMGestimation(ModeleVASICEK)
m3 <- MMGestimation(ModeleCIR)

par1 <- cbind(m1$param,m1$Tstat,m1$Tpvalue)
par2 <- cbind(m2$param,m2$Tstat,m2$Tpvalue)[-4,]
par3 <- cbind(m3$param,m3$Tstat,m3$Tpvalue)[-4,]

cov1 <- m1$Varparam
cov2 <- m2$Varparam
cov3 <- m3$Varparam

sink("MMGestimation.tex")

xtable(matrix(par1,nrow=4,ncol=3,
              dimnames=list(nomsParam,c("Est. param.,"T-Stat","p-value"))),
        caption="Modèle CKLS estimé avec GMM",digits=5)

xtable(matrix(par2,nrow=3,ncol=3,
              dimnames=list(nomsParam2,c("Est. param.,"T-Stat","p-value"))),
        caption="Modèle Vasicek estimé avec GMM",digits=5)

xtable(matrix(par3,nrow=3,ncol=3,
              dimnames=list(nomsParam2,c("Est. param.,"T-Stat","p-value"))),
        caption="Modèle CIR estimé avec GMM",digits=5)

xtable(matrix(cov1,nrow=4,ncol=4,
              dimnames=list(nomsParam,nomsParam)),
        caption="Matrice de Var-Cov des par. pour modèle CKLS avec GMM",digits=5)

```

```

xtable(matrix(cov2,nrow=3,ncol=3,
              dimnames=list(nomsParam2,nomsParam2)),
        caption="Matrice de Var-Cov des par. pour modèle Vasicek avec GMM",digits=5)
xtable(matrix(cov3,nrow=3,ncol=3,
              dimnames=list(nomsParam2,nomsParam2)),
        caption="Matrice de Var-Cov des par. pour modèle CIR avec GMM",digits=5)

sink()

```

7.4 Fichiers source R pour EMV

```

##
## maximum de vraisemblance avec Vasicek
##
library("xtable")

## importation des données
data <- read.csv("usgg.csv",header=TRUE, sep=";")[(1:50)*30,]
delta <- 1/12
rf <- data$usgg3m/100
rf[which(rf<=0)] <- 0.001

##
## graphique quantile quantile
##
pdf("MLE-qqplot-norm.pdf")
qqnorm(diff(rf))
qqline
dev.off()

## fonction objectif de log-vraisemblance négatif

VASICEK.FUNOBJ <- fonction(PAR,RF,DELTA)
{
  mu <- PAR[2]
  alpha <- PAR[1]
  sigma <- PAR[3]
  n <- length(RF)
  moyenne <- RF[-n] * exp(-alpha*delta)+mu*(1-exp(-alpha*delta))
  variance <- sigma^2 * (1-exp(-2*alpha*delta)) / (2*alpha)
  -sum(dnorm(RF[-1],moyenne,sqrt(variance),log=TRUE))
}

```

```

ES1 <- optim(c(.5,-.5,.5),fn=VASICEK.FUNOBJ,RF=rf,DELTA=delta)

##
## maximum de vraisemblance avec CIR (quasi-vraisemblance normale)
##

CIR.QV.FUNOBJ <- fonction(PAR,RF,DELTA)
{
  mu <- PAR[2]
  alpha <- PAR[1]
  sigma <- PAR[3]
  n <- length(RF)
  moyenne <- RF[-n] * exp(-alpha * DELTA) +
    mu * (1-exp(-alpha * DELTA))
  variance <- RF[-n] * sigma^2/alpha * (exp(-alpha * DELTA) - exp(-2*alpha * DELTA)) +
    mu * sigma^2/2/alpha*(1-exp(-alpha * DELTA))^2
  -sum(dnorm(RF[-1],moyenne,sqrt(variance),log=TRUE))
}

ES3 <- optim(c(alpha.st,mu.st,sigma.st),fn=CIR.QV.FUNOBJ,RF=rf,DELTA=delta)

##
## sorties
##
ES1$par[3] <- ES1$par[3]^2
ES3$par[3] <- ES3$par[3]^2

sink("MLE-dates.tex",append=FALSE,split=FALSE)
as.vector(data$date[c(1,length(data$date))])
sink()

sink("MLE-param.tex",append=FALSE,split=FALSE)
xtable(cbind(Param=c("alpha","mu","sigma"),
               Vasicek=ES1$par,CIR.QL=ES3$par),
       caption="Paramètres estimés par maximum de vraisemblance pour 2 modèles",
       label="tab:estimParam")
sink()

ES1$par
ES3$par

##
## test de ratio de vraisemblance
##

loglik <- c(ES1$value,ES3$value)

```

```
n <- length(rf)-1
1-pchisq(-2*n*(loglik[2]-loglik[1]),4)
```



Cette création est mise à disposition selon le contrat Paternité-Partage à l'identique 2.5 Canada de Creative Commons disponible à l'adresse
<http://creativecommons.org/licenses/by-sa/2.5/ca/deed.fr>

En vertu de ce contrat, vous êtes libre de :

- **partager** — reproduire, distribuer et communiquer l'œuvre ;
- **remixer** — adapter l'œuvre ;
- utiliser cette œuvre à des fins commerciales.

Selon les conditions suivantes :



Attribution — Vous devez attribuer l'œuvre de la manière indiquée par l'auteur de l'œuvre ou le titulaire des droits (mais pas d'une manière qui suggérerait qu'ils vous soutiennent ou approuvent votre utilisation de l'œuvre).



Partage à l'identique — Si vous modifiez, transformez ou adaptez cette œuvre, vous n'avez le droit de distribuer votre création que sous une licence identique ou similaire à celle-ci.