

UBO

université
de bretagne
occidentale

Bureau d'étude :
Modélisation
stochastique
de taux de fret

Encadré par
Fabrice HAMON

Représentants université:
Franck Vermet et Rainer Buckdahn



Présenté par :

Qiu Jinhong,
Traoré Korotoumou,
Boukroute Anas,
Rajaud Léa

 VesselsValue

EURIA
Euro-Institut d'Actuariat



Remerciements.

Avant de commencer ce rapport, nous souhaitons remercier les personnes grâce auxquelles nous avons pu le réaliser.

En premier lieu, nous remercions M Hamon ainsi que VesselsValues d'avoir proposés ce sujet qui nous a permis d'explorer les mécanismes de prédictions des taux de fret et de découvrir de nouvelles connaissances telles que les séries temporelles. C'est aussi grâce à cette entreprise que nous avons pu disposer du jeu de données essentiel à nos prédictions.

Nous voudrions ensuite remercier M Vermet et M Buckdahn pour nous avoir aidés et orientés tout au long de nos recherches, mais aussi pour le temps qu'ils nous ont consacré lors de l'élaboration de ce rapport. Nous leur sommes aussi reconnaissants pour toute la documentation qu'ils ont mis à notre disposition.

Table des matières

Introduction.	3
1 Les séries temporelles et prédictions.	6
1.1 Introduction sur les séries temporelles.	6
1.1.1 Définitions et notations.	6
1.2 Modélisation d'un jeu de données via les séries temporelles.	10
1.2.1 Estimation des paramètres d'une série ARMA.	10
1.3 Modélisation du taux de fret via les séries SARIMA et prédictions.	16
1.3.1 Implémentation des prédictions :	16
1.3.2 Règle de validation du modèle.	17
1.3.3 Analyse du modèle SARIMA.	19
2 Le modèle de Vasicek.	24
2.1 Description du modèle.	24
2.2 Vasicek/ARIMA.	24
2.3 Implémentation sous R.	26
2.3.1 Régression linéaire	26
2.3.2 Prédiction	27
2.3.3 Récupération des résidus et test de blancheur.	28
2.3.4 Déroulement de l'algorithme sur 100 bateaux.	29
3 Le modèle Cox-Ingersoll-Ross et prédictions du taux de fret.	34
3.1 Introduction au modèle Cox-Ingersoll-Ross et discrétisation de l'intégrale stochastique.	34
3.2 Détermination des paramètres $\theta = (\alpha, \mu, \sigma)$.	36
3.2.1 Détermination des paramètres α et μ via une régression linéaire.	36
3.2.2 Détermination de σ .	39
3.2.3 Logarithme du maximum de vraisemblance.	39
3.3 Simulation de trajectoires du CIR.	44
3.3.1 Protocole.	44
3.3.2 Simulation de la trajectoire du CIR.	44
3.3.3 Calcul du logarithme du maximum de vraisemblance.	45
4 Comparaison des prédictions des différents modèles.	49
4.1 Comparaison entre les différents modèles.	49

4.1.1	Le modèle SARIMA.	49
4.1.2	Le modèle Vasicek.	49
4.1.3	Le modèle Cox-Ingersoll-Ross.	49
4.2	Comparaison des prédictions.	50
4.2.1	Confrontation des résultats entre les modèles.	50
4.2.2	Confrontation des résultats des modèles.	54
	Conclusions.	59
	Bibliographie.	62
	Annexes.	62

Introduction

Les armateurs de bateaux, avec leur expérience, étaient capables d'estimer le prix d'un bateau. Mais les mesures de déduction faites par les armateurs se sont avérées insuffisantes suite à la crise de 2008. En effet, cette crise financière a entraîné une chute en bourse du cours des armements, une réduction des facilités de crédits pour la construction navale et un ralentissement considérable dans les trafics de produits manufacturés. Les taux de fret ont ainsi atteint leurs niveaux les plus bas depuis 15 ans en Europe et aux États Unis. D'où l'intérêt de la mise en place d'un outil permettant à la fois de modéliser, de prédire les taux de fret et d'être capable de traduire toutes ces années de savoir et d'expériences qu'ont les armateurs de bateaux.

Lors de nos recherches, nous nous sommes rendus compte que les taux de fret dépendent de plusieurs facteurs tels que : le type, l'âge et la taille du bateau, le calendrier de la construction navale, la macroéconomie, etc. Étant donné la complexité de ces variables, nous nous sommes focalisés sur l'historique des séries temporelles. Cela nous amène à nous poser les questions suivantes : Quels sont les facteurs, outils pertinents impactant les taux de fret ? Est-il possible de modéliser les taux de fret via les outils mathématiques ? Quels sont les modèles les plus adaptés quant à l'estimation et à la prédiction des taux de fret ? Quels tests, outils mathématiques et informatiques devront nous utiliser afin d'implémenter et de valider ces modèles ?

Pour répondre à ces questions, nous nous sommes intéressés à différents modèles mathématiques tels que :

Le modèle d'équilibre partiel de **Vasicek** qui a été l'un des sujets en finance avec un impact important dans la modélisation stochastique des taux depuis 1977. L'un de ses principaux inconvénients est sa tolérance vis-à-vis des taux négatifs.

Dans un second temps nous nous sommes intéressés au modèle d'équilibre général de **Cox, Ingersoll et Ross** qui a notamment remédié au problème des taux négatifs. Cependant, il ne satisfait pas les spécialistes du fait de la difficulté d'interprétation financière des paramètres.

Quant à la dernière approche, nous avons remarqué que les données sur les taux de fret étaient recueillies trimestriellement. Ainsi, l'une de nos principales motivations a été d'étudier la tendance et la saisonnalité de nos données, afin de pouvoir estimer et prédire les valeurs futures. Intuitivement, nous nous sommes donc orientés vers les séries temporelles chronologiques **ARIMA** et **SARIMA** pouvant expliquer ces phénomènes.

Dans notre étude, nous présenterons dans un premier temps les séries SARIMA. A cette étape, nous définirons les notions importantes qui seront utilisées afin d'implémenter ce modèle sur les taux de fret. Nous verrons par la suite que ce modèle permet d'avoir les meilleures prédictions pour les bateaux GAZ_LPG et GAZ_LNG. Ensuite, nous montrerons via des démonstrations mathématiques que le modèle dynamique partiel à un facteur Vasicek, est un cas particulier du modèle ARIMA. En utilisant cette équivalence, nous pourrions adapter ce modèle à l'ensemble des jeux de données. Après avoir effectué des

tests informatiques, tels que ceux de blancheur, et du Portemanteau de Box-Pierce nous pourrions modéliser 50% du jeu de données par le modèle de Vasicek. Puis, nous nous intéresserons à l'étude du modèle CIR, ce qui revient à estimer les paramètres inconnus du modèle. Pour cela, nous nous aiderons des instruments mathématiques et informatiques que nous avons à notre disposition afin de procéder à son implémentation. Pour finir, nous comparerons les résultats fournis par l'implémentation des trois méthodes afin :

- De calculer les taux d'erreurs sur les prédictions effectuées pour les trois modèles.
- De vérifier la cohérence de ces prédictions.
- De dégager une liste exhaustive des bateaux et des modèles associés.
- D'en déduire si possible le modèle le plus adapté aux taux de fret.

Le jeu de données qui nous a été donné par VesselsValue est composé du descriptif de 100 bateaux, à savoir : les taux de fret, le coût des opérations journalières, le prix de constructions de nouveaux bateaux, le type de bateaux, l'âge et la taille du bateau. Ces données nous ont été fournies du premier trimestre 1980 jusqu'au quatrième trimestre 2018, cela est normal car nous avons à notre disposition des prédictions faites par une entreprise à laquelle VesselsValue a fait appel.

1 Les séries temporelles et prédictions.

L'utilisation de séries temporelles s'est imposée à nous car elles permettent en économétrie de détecter et d'analyser les périodes de crise et de croissance, mais aussi de prédire l'évolution future de la série à partir des données passées.

Dans cette partie, nous allons traiter les étapes suivantes :

Définition des notions fondamentales sur les séries temporelles

Modélisation d'un jeu de données via ces séries.

Modélisation de la totalité du jeu de données par les séries SARIMA et réalisation des prédictions.

Mise en place d'une règle de validation du modèle via trois erreurs : deux mesures basées sur les taux d'erreurs des prédictions, une sur le maximum des erreurs et une sur les tendances de nos prédictions.

1.1 Introduction sur les séries temporelles.

L'analyse de type temporelle consiste à étudier les corrélations de la série (X_t) . Ces méthodes sont généralement paramétriques de type moyenne-mobiles (moving average MA) ou auto régressives (AR) - voir une combinaison des deux (ARMA).

1.1.1 Définitions et notations.

Dans cette partie, nous allons définir les principales notions utilisées dans notre étude.

Une **série temporelle** est une suite d'observations x_1, x_2, \dots, x_n indexée par le temps. L'indice de temps peut être en secondes, heures, années..., n est la longueur de la série. Ces observations sont la réalisation d'un processus $(X_t)_{t \in \mathbb{Z}}$. Une série temporelle peut-être constituée de plusieurs éléments :

- Une tendance représentant l'évolution à long terme de la série.
- Une saisonnalité, c'est-à-dire une évolution se répétant régulièrement tous les ans.
- Une composante stationnaire (ou résiduelle) : ce qui reste lorsque l'on a enlevé les autres composantes. Elle décrit l'évolution à court terme de la série.

Un **processus linéaire** est un processus stochastique $(X_t)_{t \in \mathbb{Z}} \subset L^2(\Omega, \mathcal{F}, P)$ formé d'une combinaison linéaire finie ou non de bruit blanc $(\eta_t)_{t \in \mathbb{Z}}$. $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire de moyenne μ s'il peut être écrit sous la forme :

$$X_t = \mu + \sum_{i=-\infty}^{\infty} b_i \eta_{t-i} \quad \text{avec} \quad \sum_{i=-\infty}^{\infty} b_i^2 < +\infty$$

Les ARMA sont des processus stationnaires et les ARIMA des processus non stationnaires intégrés : nous les rendons stationnaires par différentiation.

Nous considérons une suite de variable aléatoires $(X_t)_{t \in \mathbb{Z}}$.

$(X_t)_{t \in \mathbb{Z}}$ est un **processus stationnaire** au sens large si et seulement si :

- $\mathbb{E}(X_t) = \mu, \quad \forall t \in \mathbb{Z}$
- X_t est de carré intégrable pour tout $t \in \mathbb{Z} : \mathbb{E}(X_t) < \infty$
- $Cov(X_s, X_{s+t}) = Cov(X_{s-1}, X_{s-1+t}) = \dots = Cov(X_0, X_t), \forall t, s \in \mathbb{Z}$

Soit (X_t) un processus stationnaire, la **fonction d'auto-covariance** est la fonction γ définie par :

$$\forall h, t \in \mathbb{Z} \quad \gamma(h) = Cov(X_t, X_{t+h})$$

La fonction d'auto-covariance peut être normalisée, et la nouvelle fonction obtenue est la **fonction d'auto corrélation** ρ :

$$\forall h \in \mathbb{Z} \quad \rho(h) = corr(X_t, X_{t+h}) = \frac{\gamma(h)}{\gamma(0)}$$

L'**opérateur retard** B qui associe à tout processus $X = (X_t)_{t \in \mathbb{Z}}$ le processus $(Y_t)_{t \in \mathbb{Z}}$ est défini par :

$$\forall t \in \mathbb{Z}, \quad Y_t = BX_t = X_{t-1}$$

Nous appelons **bruit blanc** tout processus $\eta = (\eta_t)_{t \in \mathbb{Z}}$ tel que :

$$\mathbb{E}[\eta_t] = 0 \quad \text{et} \quad \mathbb{E}[\eta_t \eta_{t'}] = \sigma^2 \delta_{tt'}, \quad t, t' \in \mathbb{Z}$$

avec $\delta_{tt'}$ le symbole de Kronecker :

$$\delta_{tt'} = \begin{cases} 1 & \text{si } t = t' \\ 0 & \text{si } t \neq t' \end{cases}$$

$\sigma^2 > 0$ est appelée la variance du bruit blanc.

Nous pouvons distinguer différents types de modèles de séries temporelles tous liés les uns aux autres :

Le modèle AR Les modèles auto-régressifs forment une classe flexible de modèle pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations passées ; ils sont donc définis implicitement par la relation :

$$X_t - \phi_1 X_{t-1} + \dots - \phi_p X_{t-p} = \eta_t, \quad \forall t \in \mathbb{Z}, \quad p \geq 1$$

avec ϕ_1, \dots, ϕ_p sont des réels fixés, $(X_t)_{t \in \mathbb{Z}}$ stationnaire, et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . Un tel processus $(X_t)_{t \in \mathbb{Z}}$ est dit auto-régressif d'ordre p (AR(p)). Nous posons $\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p$, ce qui nous permet d'écrire :

$$\Phi(B)X_t = \eta_t, \quad t \in \mathbb{Z}.$$

Le modèle MA Les processus moyennes mobiles forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations d'un bruit blanc ; ils sont donc définis par la relation :

$$X_t = \eta_t - \theta_1 \eta_{t-1} + \dots - \theta_q \eta_{t-q}, \quad t \in \mathbb{Z}$$

où $\theta_1, \theta_2, \dots, \theta_q$ sont des réels fixés pour tout $p, q \geq n$ et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . Un tel processus $(X_t)_{t \in \mathbb{Z}}$ est dit MA(q). En posant : $\Theta(B) = I - \theta_1 B - \dots - \theta_q B^q$, on peut écrire :

$$X_t = \Theta(B)\eta_t$$

avec B l'opérateur retard.

Le modèle ARMA Les séries ARMA sont des combinaisons des deux premiers modèles. Elles présentent les avantages d'être plus souples à l'utilisation et de fournir en principe de bonnes approximations des séries réelles, avec moins de paramètres que les modèles purs. Un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est un processus auto-régressif et moyenne mobile d'ordre (p, q) (ARMA(p, q)), s'il vérifie l'équation suivante :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q}, \quad \forall t \in \mathbb{Z},$$

avec $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . Nous posons avec les fonctions définies précédemment :

$$\Phi(B)X_t = \Theta(B)\eta_t, \quad \forall t \in \mathbb{Z}$$

En réalité la plupart des séries économiques ne vérifient pas l'hypothèse de stationnarité. Ce problème peut être surmonté si l'on considère les différences d'ordre d , alors l'hypothèse de stationnarité devient souvent acceptable. Il est donc naturel de considérer la classe des processus dont la différence d'un certain ordre satisfait une représentation ARMA. Donc, le processus ARMA est un cas particulier de processus linéaire et joue un rôle primordial dans la modélisation concrète des processus stationnaires.

Les deux prochains modèles nous permettent de traiter des composantes non stationnaires, notamment les tendances (processus ARIMA) et, les saisonnalités (processus SARIMA).

Le modèle ARIMA On dit que $(X_t)_{t \in \mathbb{Z}}$ est un processus auto-régressif et moyenne mobile intégré ARIMA(p,d,q) s'il satisfait une équation du type :

$$\phi(B)(I - B)^d X_t = \Theta(B)\eta_t$$

où :

$$\begin{cases} \phi(B) = I + \phi_1 B + \dots + \phi_p B^p \\ \Theta(B) = I + \theta_1 B + \dots + \theta_q B^q \end{cases}$$

$\phi(B)$ et $\Theta(B)$ sont des polynômes dont les racines sont de module supérieur à 1, et où les conditions initiales : $(X_{-1}, X_{-2}, \dots, X_{-p-d}, \eta_{-1}, \eta_{-2}, \dots, \eta_{-q})$ sont non corrélées avec $\eta_0, \eta_1, \dots, \eta_t, \eta = (\eta_t)_{t \geq -q}$ étant un bruit blanc de variance σ^2 . $(I - B)^d X_t$ est un polynôme stationnaire. Et B l'opérateur retard.

Le modèle SARIMA. Lors de notre étude nous nous sommes interrogés sur la présence ou non d'une saisonnalité dans les taux de fret. Nous avons donc eu recours aux processus SARIMA.

Soient $p, q, d, s \geq 0$ et $P, Q, D \geq 0$. Un processus $(X_t)_{t \in \mathbb{Z}} \subset L^2(\Omega, \mathcal{F}, P)$ est dit auto-régressif et moyenne mobile intégré saisonnier SARIMA(s,p,d,q) si :

$$Y_t = \Delta^d \Delta_s^D X_t = (I - B)^d (I - B^s)^D X_t$$

avec $(Y_t)_{t \in \mathbb{Z}}$ un processus ARMA stationnaire de la forme :

$$\Phi(B)F(B^s)Y_t = \Theta(B)G(B^s)\eta_t$$

où Φ (resp. Θ) est le polynôme générateur d'un AR(p) (resp. d'un MA(q)) :

$$\Phi(z) = 1 - \sum_{k=1}^p \phi_k z^k \quad \text{et} \quad \Theta(z) = 1 - \sum_{k=1}^q \theta_k z^k$$

et où, pour la saisonnalité $Y_t - Y_{t-s}$, F (resp. G) est le polynôme générateur d'un AR(P) (resp. d'un MA(Q)) :

$$F(z) = 1 - \sum_{k=1}^P f_k z^k \quad \text{et} \quad G(z) = 1 - \sum_{k=1}^Q g_k z^k$$

Nous appelons $\Delta^d X_t$ la différence d'ordre d de X_t , c'est à dire $(I - B)^d X_t$.

Voici un récapitulatif des correspondances entre les différents modèles :

→ AR(p) = ARMA(p,0)

→ MA(q) = ARMA(0,q)

1.2 Modélisation d'un jeu de données via les séries temporelles.

Dans un premier temps nous avons tenté de modéliser le taux de fret pour un bateau. Ensuite, nous avons cherché à le faire pour tous les modèles de navires. Ainsi, nous avons pu comparer si cette approche était adaptée ou non. Pour pouvoir implémenter les modèles de séries temporelles, nous avons suivi un schéma similaire à celui de Box et Jenkins qui se résume par les étapes suivantes :

- Le choix de l'ordre de la série AR.
- L'estimation des paramètres.
- Le calcul des résidus.
- Mise en place de tests sur les résidus.

1.2.1 Estimation des paramètres d'une série ARMA.

Nous avons choisi ce schéma car c'est l'une des méthodes de traitements des séries ARMA la plus utilisée en raison de sa simplicité, de l'économie de temps qu'elle permet et de la fiabilité de ses résultats. Nous avons dû supposer que nos observations étaient celles d'une série chronologique stationnaire.

Première étape : Choix de l'ordre de la série et estimation des paramètres de la série ARMA.

L'utilisation de l'informatique permet de faciliter les différents calculs à effectuer (cf partie 1.1.1). Voici comment nous avons procédé :

- Réorganisation du jeu de données.
- Transformation des données en série temporelle.
- Calcul des coefficients d'auto-corrélation.
- Conclusions.

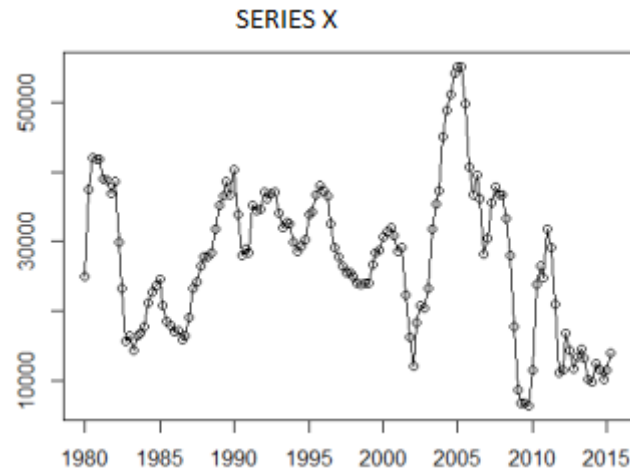
Les données nécessaires à notre étude sont regroupées dans le fichier « Time Série » : les taux de fret, le coût des opérations journalières, le prix de constructions de nouveaux bateaux, le type de bateaux, l'âge et la taille du bateau. A ce stade nous ne travaillons que sur la modélisation du taux de fret sans tenir compte des autres variables citées ci-dessus. Ensuite nous avons créé un fichier contenant uniquement les taux de fret afin de pouvoir l'importer sur R et de modéliser le jeu de données.

Pour transformer les données liées aux taux de fret en séries temporelles, nous avons utilisé la fonction "ts ()" qui sert à créer des objets de séries chronologiques.

```
x=ts(taux,start=c(1980,1),end=c(2018,4),frequency=4)
y=ts(taux,start=c(1980,1),end=c(2015,2),frequency=4)
```

La commande "start" correspond au temps de la première observation et "end" celui de la dernière. "frequency" indique le nombre d'observation par unités de temps.

Nous avons créé deux séries une première série : une regroupant les taux de fret allant de 1980 à 2020 et une seconde regroupant les taux de fret de 1980 à 2015.



Nous allons tracer les fonctions d'auto-corrélation et d'auto-corrélation partielle définies précédemment. Pour parvenir à obtenir ces fonctions voici les étapes à suivre :

En théorie : Les coefficients p et q peuvent être estimés à l'aide de la fonction d'auto-corrélation et d'auto-corrélation partielle qui nous permettent de déduire le type du processus ARMA.

Nous allons partir des équations de Yule-Walker (cf p34 du document "Renforcement statistique, Séries chronologiques") qui nous permettent d'obtenir comme modèle ARMA(p,q) d'un processus stationnaire $X = (X_t) \subset L(\Omega, \mathcal{F}, P)$:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q}, \quad \forall t \in \mathbb{Z}$$

avec $\mathbb{E}(X_t) = 0$, $\phi_0 = 0$, et $\theta_0 = 0$.

De l'espérance de $X_t X_{t-h}$, nous obtenons l'équation suivante :

$$\sum_{i=0}^p \phi_i \mathbb{E}(X_{t-i} X_{t-h}) = \sum_{j=0}^q \theta_j \mathbb{E}(\eta_{t-j} X_{t-h})$$

Or, pour $h > q + 1$, $\mathbb{E}(\eta_{t-j} X_{t-h}) = 0$. Donc, la fonction d'auto-corrélation ρ vérifie l'équation récurrente ci dessous :

$$\rho(h) + \phi_1 \rho(h-1) + \dots + \phi_p \rho(h-p) = 0, \quad \forall h \geq q+1$$

Ainsi, nous obtenons pour $q+1 \leq h \leq q+p$ un système de p équations linéaires en

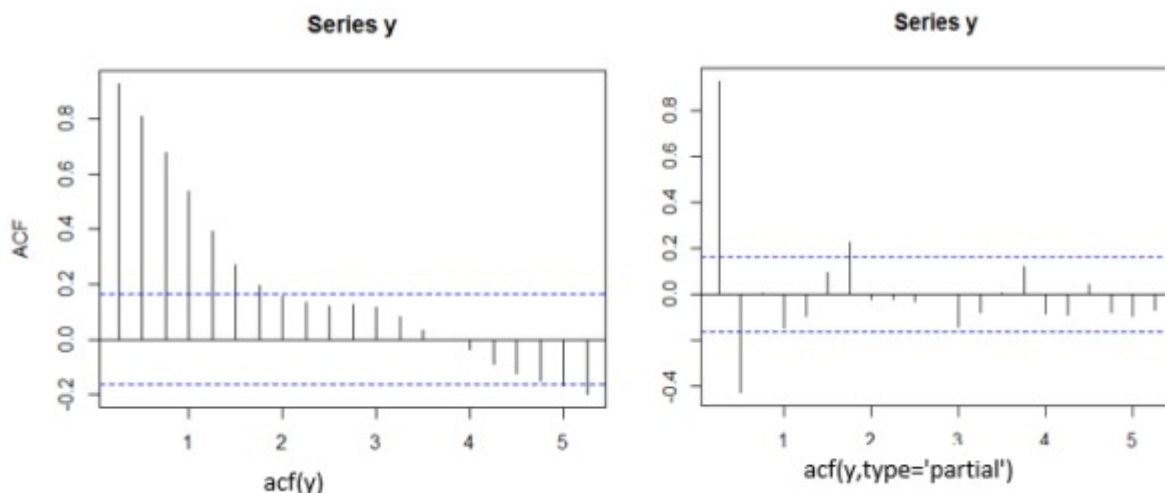
$\phi_1, \dots, \phi_p :$

$$\begin{pmatrix} \rho(q+1) \\ \cdot \\ \cdot \\ \cdot \\ \rho(q+p) \end{pmatrix} = \begin{pmatrix} \rho(q) & \rho(q-1) & \dots & \rho(q-p+1) \\ \rho(q+1) & \rho(q) & \dots & \rho(q-p+2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho(q+p-1) & \rho(q+p-2) & \dots & \rho(q) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \phi_p \end{pmatrix}$$

Ainsi, nous pouvons exprimer $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ à l'aide de la fonction d'auto-corrélation et en déduire des estimateurs de ces paramètres.

La fonction "acf" sous R permet de résoudre ce système et d'en déduire les coefficients d'auto-corrélation.

Informatiquement : Nous avons donc créé deux séries temporelles X_t et Y_t . Dans X_t , nous avons stocké les taux de fret de 1980 à 2020 et dans Y_t ceux de 1980 à 2015, ce qui revient à enlever les valeurs futures c'est à dire celles de 2016 à 2020. En utilisant, "acf" nous avons donc obtenu les graphiques liés aux fonctions d'auto-corrélation, et d'auto-corrélation partielle :



L'auto-corrélation d'une série temporelle est égale à la corrélation de celle-ci par rapport à une version décalée dans le temps d'elle-même. L'auto-corrélation partielle est la même chose sauf que l'on a enlevé l'influence des variables intermédiaires. Nous pouvons remarquer que pour la fonction d'auto-corrélation partielle le premier pic est très haut comparé au second qui est plus petit. Il est intéressant de noter que les autres coefficients sont pratiquement nuls à partir du deuxième pic.

Donc l'une de nos premières conclusions est que la série pourrait être une série AR(2). En effet, l'ACF partiel donne un résultat similaire à celui des AR(2). Afin de vérifier si cette hypothèse est valable, nous devons ensuite effectuer des tests sur les résidus.

Nous avons l'équivalence :

$$AR(2) \Leftrightarrow ARIMA(2, 0, 0)$$

Deuxième étape : Calculs et tests sur les résidus.

Nous testons cette hypothèse sur le premier jeu de données. Une approche intéressante est de prédire les valeurs futures et de comparer nos prédictions à celles fournies par le jeu de données. Cette étape est informatique. Nous avons procédé comme suit :

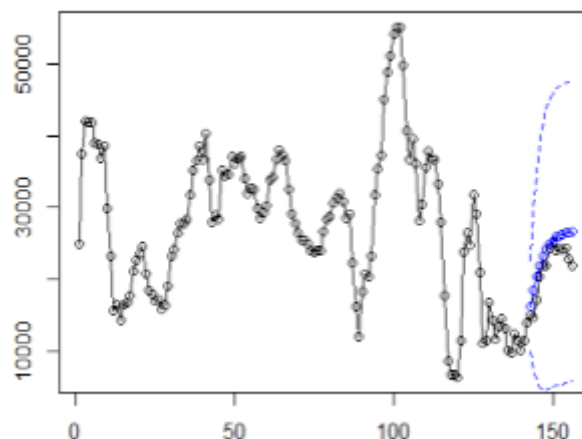
Nous avons créé une série AR(2) : $a=ARIMA(2,0,0)$ ¹. Cette fonction nous renvoie une série AR(2).

Puis, nous avons fait une prédiction de la série "a" sur 14 valeurs, c'est à dire sur les valeurs de taux de fret entre 2015-2020. Ensuite nous avons stocké les valeurs prédites dans une matrice "c" et les écarts types dans une autre matrice "s" comme suit :

```
b=predict(a,n.ahead=L)
c=as.matrix(b$pred)
s=as.matrix(b$se)
```

Nous traçons un graphique de nos prédictions ainsi qu'un intervalle de confiance, c'est à dire une marge d'erreur pour celles-ci :

```
plot(1 :lx,x)
lines(1 :lx,x)
points((l+1) :lx,c, col='blue')
lines((l+1) :lx,c, col='blue')
lines((l+1) :lx,c+1.96*s,lty="dashed", col='blue')
lines((l+1) :lx,c-1.96*s,lty="dashed", col='blue')
```

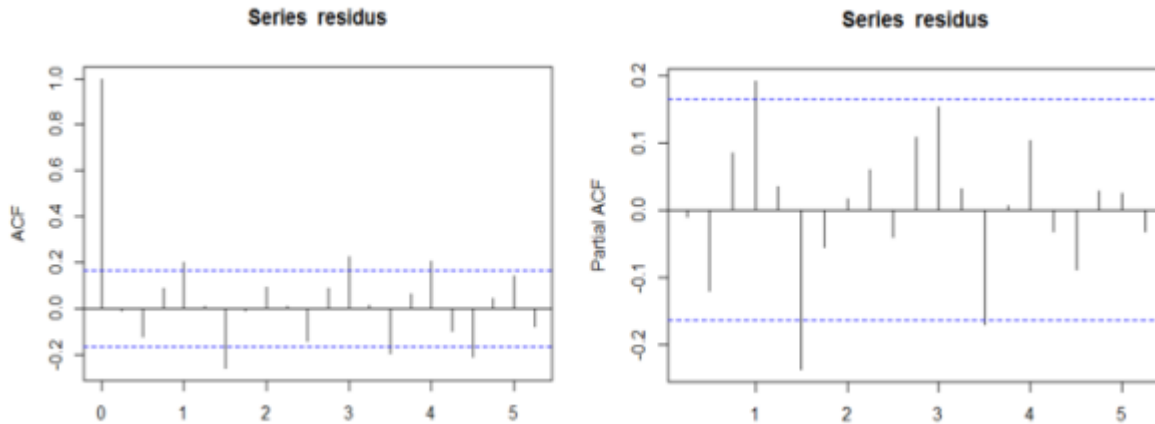


Sur ce graphique, nous pouvons observer nos prédictions représentées par les points bleus et l'intervalle de confiance en pointillés bleus. Nous pouvons notamment remarquer un retour à la moyenne de la série. Ce qui est normal. La prochaine étape consiste à effectuer

1. cette fonction crée une série ARIMA avec en paramètres $p=2$, $d=0$ et $q=0$.

des tests sur les résidus. Il s'agit de récupérer les résidus de la série a (ARIMA) que nous avons créé précédemment et de calculer les auto corrélations des résidus.

```
residus=a$res
acf(residus)
acf(residus,type='partial')
```



Si les résidus $\eta_t, t \in \mathbb{Z}$ obéissent à un bruit blanc, il ne doit pas exister d'auto-corrélation dans la série. Nous pouvons alors utiliser les différents tests suivants : celui du portemanteau proposé par Box et Pierce, et celui de Ljung-Box.

Théoriquement Le premier cherche à tester l'indépendance des résidus. Nous notons $\hat{\rho}_{\hat{\eta}}$ la fonction d'auto-corrélation empirique des résidus estimés $\hat{\eta}$. Nous considérons la statistique de Box-Pierce Q_{BP} défini par la somme des H premières auto-corrélations empiriques au carré :

$$Q_{BP} = \sqrt{T} \sum_h^H \hat{\rho}_{\hat{\eta}}(h)^2$$

Les auto-corrélations $\hat{\rho}_{\hat{\eta}}$ d'un bruit blanc sont approximativement indépendantes et de loi $\mathcal{N}(0, \frac{1}{T})$, lorsque T est assez grand. Les tests concernant le bruit blanc, ont pour but de vérifier si l'estimation des résidus est bien cohérente avec les hypothèses relatives au bruit blanc. Nous pouvons donc montrer sous l'hypothèse $H_0 : \rho_{\hat{\eta}}(1) = \dots = \rho_{\hat{\eta}}(H) = 0$, si $T > H > 2$, alors :

$$Q_{BP} \xrightarrow{T \rightarrow +\infty} \mathcal{X}^2(H - 2)$$

L'indice H est généralement choisi entre 15 et 20 (en pratique, on ne considère pas de processus ARMA d'ordre très élevé).

Cette statistique de test peut être améliorée (dans le sens où le test est plus puissant) en tenant compte du fait que la variance de $\hat{\rho}_{\hat{\eta}}$ vaut $(T - h)/(T(T + 2))$. On obtient alors la statistique de Ljung-Box :

$$Q_{LB} = T(T + 2) \sum_{h=1}^H (T - h)^{-1} \hat{\rho}_{\hat{\eta}}(h)^2$$

qui sous l'hypothèse H_0 , converge également vers une loi du χ^2 à H-2 degrés de liberté, quand T tend vers l'infini. Si on rejette l'hypothèse nulle, alors il faut essayer de trouver un autre modèle qui pourrait mieux convenir.

Informatiquement Sous R, les calculs liés à la méthode de Box et Jenkins se font de manière automatique à l'aide de fonction que nous allons présenter.

Nous effectuons un test de **Box-Pierce** qui évalue la corrélation existant entre les résidus. Nous voulons tester l'hypothèse H_0 : "les résidus sont de bruit blanc fort".

```
Box.test(residus, lag = 12, type = "Box-Pierce")
```

```
Box.test(residus, lag = 12, type = "Ljung-Box")
```

Nous obtenons les résultats suivants :

Test du portemanteau	X-squared	df	p-value
Box-Pierce	30.9472	12	0.002007
Ljung-Box	33.1023	12	0.0009331

Ici, la pvalue est inférieure à 5% donc nous rejetons l'hypothèse H_0 . Les résidus ne sont pas des bruits blancs donc le bateau CO_MED_00 ne peut pas être modélisé par un AR(2).

1.3 Modélisation du taux de fret via les séries SARIMA et prédictions.

Les modèles SARIMA (Seasonal Integrated AutoRegressive Moving-Average) sont utilisés pour modéliser les données saisonnières (mensuelles, trimestrielles) ou cycliques. **Nous pouvons souligner que le modèle SARIMA est un modèle ARIMA avec une composante saisonnière.** Dans notre étude nous nous sommes finalement focalisés sur ce modèle car c'est lui qui s'adapte le plus à la modélisation des taux de fret.

1.3.1 Implémentation des prédictions :

Afin de procéder à l'implémentation de ce modèle sous R , nous avons choisi parmi le jeu de données les deux premiers trimestres de 2015 car nous possédons les valeurs réelles que nous avons reçu de la part de VesselsValue et ainsi nous avons pu comparer nos résultats avec les leurs.

Les séries SARIMA sont de la forme SARIMA(s,p,d,q), il nous faut donc dans un premier temps estimer ces paramètre afin de faire nos prédictions. Pour cela nous utilisons la fonction "auto.arima" qui retourne le meilleur modèle ARIMA selon la valeur de l'AIC et du BIC , la fonction effectue une recherche sur les modèles possibles en fonction des contraintes d'ordre imposées. Ces deux valeurs se calculent ainsi :

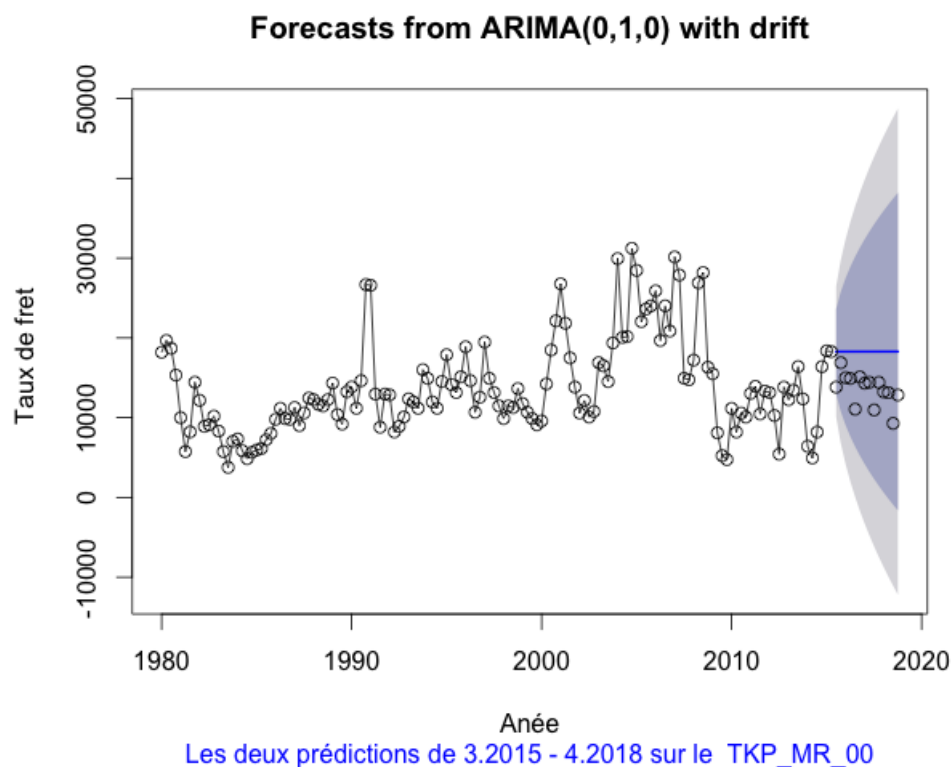
- $AIC = -2\ln(L(\hat{\theta})) + 2d$
- $BIC = -2\ln(L(\hat{\theta})) + \log(n)$

avec $\hat{\theta}$ l'estimateur du maximum de vraisemblance, L la fonction de vraisemblance (nous détaillons leur calcul dans la partie sur le modèle CIR) , d le nombre de paramètres inconnus et n le nombre d'observations. Plus l'AIC et le BIC sont petits, meilleur est le modèle.

Puis nous utilisons "forecast" qui donne les prédictions liées à ce modèle. Celle-ci utilise des méthodes particulières qui dépendent de la classe du premier argument . Nous avons créé un fichier taux2 qui contient toutes les séries liées à chaque type de bateaux.

```
n=dim(taux2)[2]    #nombre de types de bateaux
for (j in 1:n)
{y=ts(taux2[,j],start=c(1980,1),end=c(2015,2),frequency=4)
#Valeurs connues
y1=ts(taux2[,j],start=c(1980,1),end=c(2014,4),frequency=4)
#Valeurs connues sans les deux premiers trimestres 2015
a1=auto.arima(y1) #paramètres modèle arima
z=forecast(a1,h=2) #fonction de prédiction
pred=z$mean      #résultat des prédictions ponctuelles
}
```

Voici le type de graphique que nous obtenons, par exemple pour le bateau TKP_MR_00 (série 76) :



Les boules noires et la courbe noire sont les données et les prédictions données par VesselsValue, la courbe bleue est notre prédiction, et le halot gris est l'intervalle de confiance à 95 % pour nos prévisions. Nous observons sur ce graphique un drift, nous détaillons plus tard dans ce rapport ce point.

1.3.2 Règle de validation du modèle.

Nous avons effectué des tests sur l'ensemble des séries afin de déterminer quels types de bateaux pourraient être modélisés par le modèle SARIMA et les calculs des **taux d'erreur** pour chaque navire. Voici les différents tests que nous avons faits :

- **Type d'erreur n°1 (Erreur1)** : cette erreur représente le taux de pourcentage des erreurs de prédictions pour le premier trimestre de 2015. Nous calculons les taux d'erreurs pour le premier trimestre de 2015 en utilisant la formule suivante :

$$erreur1 = \frac{y - \hat{y}}{y} \quad (1)$$

avec y les prédictions données par VesselsValue et \hat{y} nos prédictions obtenues avec la fonction "forecast" sous R. Ci dessous le code de ce test.

```
l=length(y1)
```

```

err=matrix(0,nrow=n,ncol=4)
modele=NULL

for (j in 1:n)
{y=ts(taux2[,j],start=c(1980,1),end=c(2015,2),frequency=4)
y1=ts(taux2[,j],start=c(1980,1),end=c(2014,4),frequency=4)

a1=auto.arima(y1) #paramètre modèle arima
z=forecast(a1,h=2) #fonction de prédiction
pred=z$mean      #predictions ponctuelles
err[j,1]=(y[1+1]-pred[1])/y[1+1] #calcul de l'erreur 1
err[j,2]=(y[1+2]-pred[2])/y[1+2] #calcul de l'erreur 2
modele=c(modele,z$method)
}

```

L'erreur 2 relève du même calcul mais pour le deuxième trimestre 2015.

- **Type d'erreur n°3 (Erreur sur le maximum)** : Nous choisissons la valeur absolue du maximum entre les erreurs 1 et 2 que nous stockons dans un vecteur :

$$erreur3 = |max(erreur1, erreur2)|$$

Si le max est inférieur à 15% c'est que notre prédiction est bonne, le code nous retourne "1" dans ce cas, sinon "0". Ci dessous le code traduisant ce test :

```

err_max=cbind(err[j,1],err[j,2])
#nous créons une matrice avec les erreurs du 1er et du 2ème trimestre.
A=apply(abs(err_max),1,max)
#nous calculons les valeurs absolues des erreurs précédentes
err_max2=cbind(err_max,A<0.15)
#nous créons la matrice avec les valeurs absolues inférieures à 15%

```

- **Type d'erreur n°4 (Erreur de tendance)** : Ce test consiste à comparer les tendances que nous avons prédites et celles données par VesselsValue. Nous regardons si lorsque nous prédisons une hausse ou une baisse, l'entreprise prévoit la même chose pour le premier et le deuxième trimestres 2015.

Si les prédictions sont concordantes, notre code nous renvoie "1" et si les résultats sont contraires "0" comme nous pouvons le voir dans ce code :

```
n=100
```

```

l=length(y1)
err=matrix(0,nrow=n,ncol=4)
modele=NULL
err_tend=NULL
id=rep(1,100)

for (j in 1:n)
{y=ts(taux2[,j],start=c(1980,1),end=c(2015,2),frequency=4)
y1=ts(taux2[,j],start=c(1980,1),end=c(2014,4),frequency=4)

a1=auto.arima(y1)
z=forecast(a1,h=2)
pred=z$mean          # predictions ponctuelles
err[j,1]=(y[1+1]-pred[1])/y[1+1]
err[j,2]=(y[1+2]-pred[2])/y[1+2]
err[j,3]=sign(y[1+1]-y[1])*sign(pred[1]-y[1])
#(S'il y a prediction d' une hausse (resp. baisse)
#et s'il y avraiment une hausse(resp. baisse) =1, sinon =-1)
err[j,4]=sign(y[1+2]-y[1+1])*sign(pred[2]-y[1+1])
modele=c(modele,z$method)
err_tend[(err[j,3]==id) & (err[j,4]==id)]=1
}

```

Nous avons décidé d'accorder plus d'importance à l'erreur sur le pourcentage que celle sur la tendance, car c'est la première qui a le rôle plus important. Après cela, nous pouvons conclure si le modèle ARIMA est adapté ou non, selon le type de bateau.

1.3.3 Analyse du modèle SARIMA.

Nous avons fait les calculs de ces tests pour tous les bateaux ; nos résultats sont fournis en annexe.

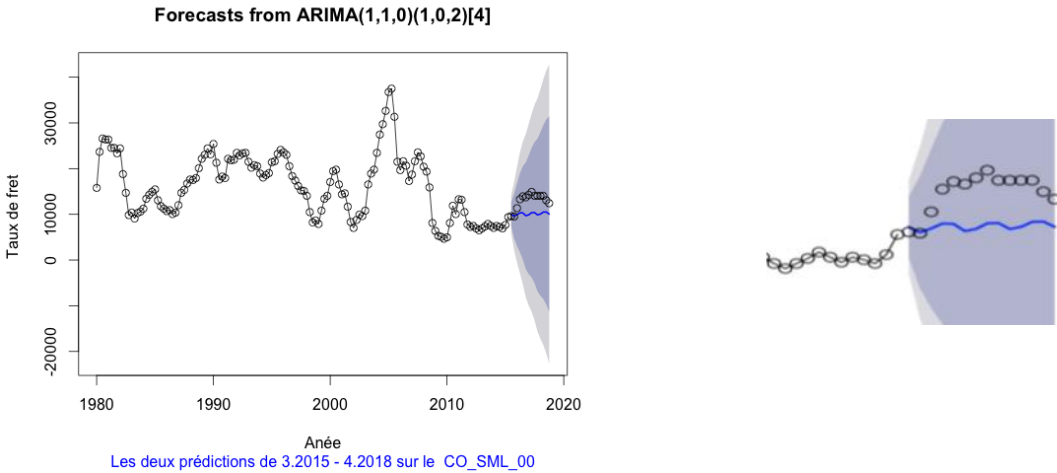
Pour illustrer cette partie, nous avons regroupé les résultats donnés par nos codes et tracé les représentations associées pour les bateaux qui ont les meilleures modélisations : les bateaux CO_SMI_00 (série 1), TKP_MR_00 (série 76), GAS_LNG_00 (série 91), et GAS_LPG_00 (série 96) :

Erreur 1: représente le taux de pourcentage des erreurs de prédiction pour le premier trimestre 2015.
 Erreur 2: représente le taux de pourcentage des erreurs de prédiction pour le deuxième trimestre 2015.

Erreur max: est la valeur absolue du maximum entre les erreurs 1 et 2.
 Erreur de tendance: variable qui vaut 1 si le modèle prédit une hausse lorsqu'il y a une hausse, ou une baisse lorsqu'il y en a une, et 0 sinon.

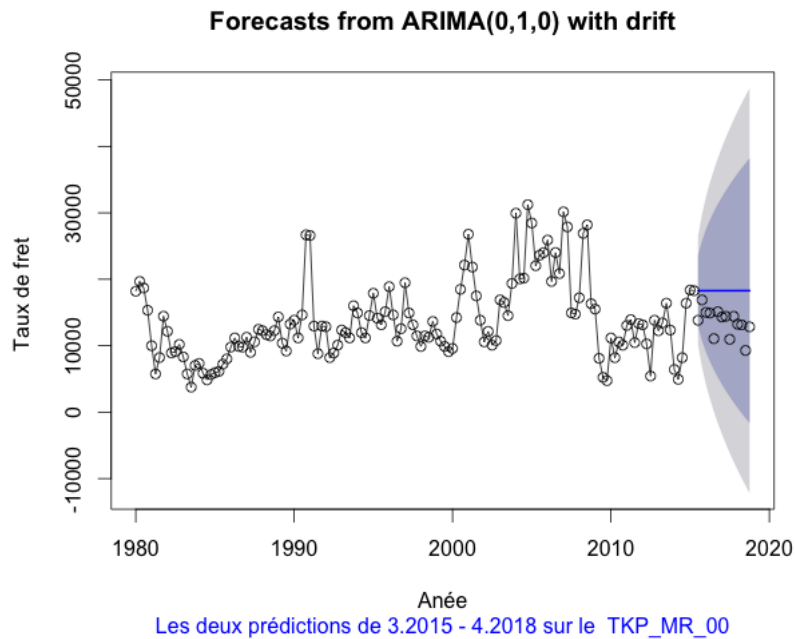
Nom	Erreur1	Erreur2	Err_Max	Err_Tend	Modèle
CO_SML_00	0,013	0,135	1	1	ARIMA(1,1,0)(1,0,2)[4]
TKP_MR_00	0,108	0,104	1	0	ARIMA(0,1,0) with drift
GAS_LNG_00	-0,104	0,014	1	1	ARIMA(2,1,0)(1,0,0)[4]
GAS_LPG_00	0,102	0,043	1	0	ARIMA(0,1,1)(0,0,1)[4]

Prédiction pour le bateau CO_SML_00 (série 1) :



Les ronds noirs représentent les prédictions fournies par VesselsValue et le trait bleu, nos prédictions. L'halot gris est notre intervalle de confiance à 95%. **Les vagues de la courbe bleue montrent une volatilité élevée dans nos prédictions, ce qui est synonyme de saisonnalité.**

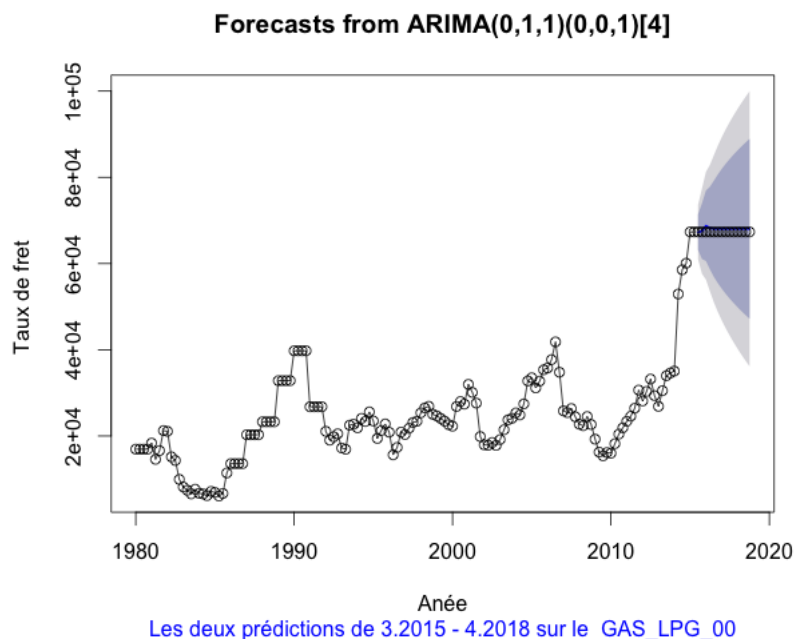
Prédiction pour le bateau TKP_MR_00 (série 76) :



Nous rappelons que les ronds noirs sont les données que nous ont été transmises par VesselsValue et en bleu, nos prédictions avec les intervalles de confiance (95%) en gris.

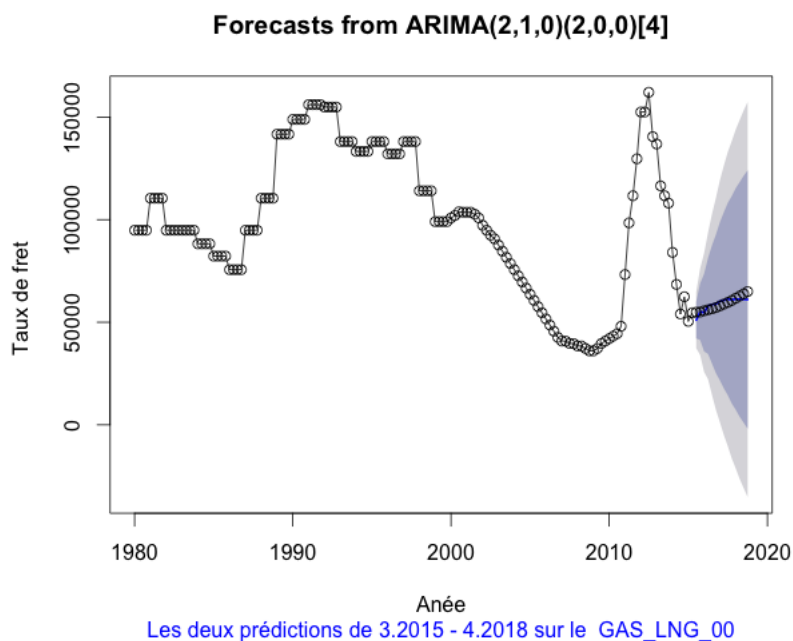
Nous pouvons voir sur ce graphique qu'il apparaît un "drift" : nous avons donc voulu le calculer. Pour cela, nous prenons la moyenne des prédictions et nous comparons la covariance entre celle ci et un vecteur constant : $t = (0, 0.25, 0.5, \dots, 3.25)$, nous prenons ces valeurs car nos données sont trimestrielles. Par conséquent, nous découpons les années en quart (nous avons une annuité qui est donc égale à 1). Nous obtenons alors un drift égal à 2,227.

Prédiction pour le bateau GAS_LNG_00 (série 91) :



Nous obtenons pour ce bateau une prédiction constante, mais c'est aussi ce qui avait été prévu par les estimations fournies par VesselsValue.

Prédiction pour le bateau GAS_LPG_00 (série 96) :



Nous pouvons voir que nos prévisions pour les navires 91 et 96 coïncident avec les prévisions du fichier qui nous a été transmises par VesselsValue. Finalement, nous pouvons faire la conclusion prudente que le modèle SARIMA peut s'adapter aux deux derniers types de bateaux : GAS_LNG et GAS_LPG.

Les tests effectués sur la tendance nous permettent de déduire que, si le modèle prédit bien une hausse lorsqu'il y a une hausse, ou une baisse lorsqu'il y a une baisse, dans ce cas, la valeur vaut "+1", dans le cas contraire "-1".

Sur les 2 premiers trimestres de 2015, pour les bateaux GAS_LPG et GAS_LNG, VesselsValue prédit une baisse alors que le modèle SARIMA donne des prédictions à la hausse. Le même scénario est observé pour les Tankers Product MR (le modèle SARIMA prédit une baisse alors que VesselValue prédit une hausse du taux de fret).

Nous pouvons voir que sur certains types de bateaux, le modèle présente une certaine saisonnalité : c'est le cas pour le premier bateau, le CO_SMI_00, par exemple, mais ce n'est pas le cas pour tous les bateaux. Par conséquent, nous ne pouvons pas généraliser cette saisonnalité.

Les tests donnés par l'erreur sur le maximum des erreurs, nous permet de constater les points suivants : en se basant sur le modèle ARIMA , on trie les bateaux ayant un maximum d'erreurs du premier et du deuxième trimestres inférieur à 15%. Nous constatons que les bateaux suivants peuvent être modélisés par un modèle SARIMA :

Type de Bateau	Modèle
Small Containers	ARIMA(1,1,0)(1,0,2)[4]
GAS LPG	ARIMA(0,1,1)(0,0,1)[4]
GAS LNG	ARIMA(2,1,0)(2,0,1)[4]
Tankers Product MR	ARIMA(0,1,0) with drift

Nous observons des processus SARIMA avec "[4]" en composante saisonnière, donc la tendance est trimestrielle.

Voici un tableau récapitulatif de nos conclusions :

Modèles ARIMA	Modèle SARIMA
BULK PanaMax	Vary Large Containers
BULK SupraMax	Tankers Crude SZX
Bulk Hardy	Tanker Product LR
Bulk Handy Max	Tanker Crude Panamax
Bulk Handy	Tankers Crude Aframax
GAS LPG	Tankers Crude Very Large Containers
Small Containers	Containers Medium
BULKCAP	GAS LNG
Bulk PostPanaMax	
BULK KamsarMax	
Tankers Product MR	
Tankers Product LR	
Large Containers	

2 Le modèle de Vasicek.

Le modèle de Vasicek élaboré en 1977, est un modèle classique de taux qui permet de décrire leurs évolutions. On dit que c'est un modèle à un facteur, car il décrit les mouvements de taux d'intérêt par une seule source de risque de marché. L'objectif de cette partie est de modéliser les spot earnings via le modèle de Vasicek. Pour atteindre cet objectif, nous allons suivre les étapes suivantes :

- Introduction des notions importantes sur le modèle de Vasicek.
- Équivalence entre modèle de Vasicek et modèle AR(1).
- Implémentation.
- Test et validation du modèle.

2.1 Description du modèle.

Vasicek, pour le taux court² r_t considère le processus d'Ornstein Uhlenbeck définit par :

$$dr_t = \alpha(\mu - r_t)dt + \sigma dW_t \quad (2)$$

où,

- r_t le taux à l'instant t .
- α, μ, σ des constantes positives :
 - α : étant la vitesse de retour à la moyenne de long terme.
 - μ : la moyenne de long terme du taux autour de laquelle évolue. le taux court instantané.
 - σ : la volatilité du taux court terme.
- (W_t) un mouvement Brownien.

Après résolution de l'équation on trouve :

$$r_t = r_0 e^{-\alpha t} + \mu(1 - e^{-\alpha t}) + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dW_s \quad (3)$$

avec : r_0 le taux de fret initial.

2.2 Vasicek/ARIMA.

Nous avons vu dans la partie sur les séries temporelles que si $(X_t) \sim AR(p)$ alors :

$$X_t - \phi_1 X_{t-1} + \dots - \phi_p X_{t-p} = \eta_t, \quad \forall t \in \mathbb{Z}$$

2. Le taux court est le taux pratiqué sur les créances à court terme c'est à dire de 0 à 2 ans

Donc :

$$(X_t) \sim AR(1) \iff X_t = \phi_1 X_{t-1} + \eta_t$$

avec (X_t) une solution stationnaire, $|\phi_1| < 1$, et $\eta_t \sim \mathcal{N}(0, \sigma^2)$ un bruit blanc gaussien.

Nous supposons que nous observons un processus à temps discret, et que y_1, \dots, y_n est la série temporelle centrée associée aux spots earnings. Nous avons prouvé dans la partie précédente que l'on a :

$$r_t = r_0 e^{-\alpha t} + \mu(1 - e^{-\alpha t}) + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dW_s \quad (4)$$

$$\Rightarrow r_t - \mu = (r_0 - \mu)e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dW_s \quad (5)$$

Nous allons faire la démonstration de l'équivalence entre le modèle de Vasicek et le modèle AR(1) ci dessous.

Nous posons : $y_t = r_t - \mu = (r_0 - \mu)e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dW_s$.

Par ailleurs, $y_0 = r_0 - \mu$. Nous obtenons alors :

$$y_t = y_0 e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dW_s \quad (6)$$

D'où :

$$y_t = y_0 e^{-\alpha(t-1)} e^{-\alpha} + \sigma e^{-\alpha(t-1)} e^{-\alpha} \int_0^t e^{\alpha s} dW_s \quad (7)$$

Par la linéarité de l'intégrale nous avons :

$$y_t = y_0 e^{-\alpha(t-1)} e^{-\alpha} + \sigma e^{-\alpha(t-1)} e^{-\alpha} \left(\int_0^{t-1} e^{\alpha s} dW_s + \int_{t-1}^t e^{\alpha s} dW_s \right) \quad (8)$$

Alors :

$$y_t = e^{-\alpha} \overbrace{\left(y_0 e^{-\alpha(t-1)} + \sigma e^{-\alpha(t-1)} \int_0^{t-1} e^{\alpha s} dW_s \right)}^{y_{t-1}} + \sigma e^{-\alpha(t-1)} e^{-\alpha} \int_{t-1}^t e^{\alpha s} dW_s \quad (9)$$

Donc :

$$y_t = e^{-\alpha} y_{t-1} + \sigma e^{-\alpha(t-1)} e^{-\alpha} \int_{t-1}^t e^{\alpha s} dW_s \quad (10)$$

$$y_t = e^{-\alpha} y_{t-1} + \sigma e^{-\alpha t} \overbrace{\int_{t-1}^t e^{\alpha s} dW_s}^{\epsilon_t} \quad (11)$$

Nous trouvons donc :

$$y_t = e^{-\alpha}y_{t-1} + \epsilon_t \quad (12)$$

Avec :

- W_t le mouvement Brownien.
- ϵ_t bruit blanc gaussien (car il dépend du mouvement brownien qui est à accroissement indépendants)
- $\epsilon_t \sim \mathcal{N}(0, \sigma^2 \cdot \frac{1-e^{-2\alpha}}{2\alpha})$

En posant $\phi_1 = e^{-\alpha}$ on retrouve :

$$(y_t) \sim AR(1) \Leftrightarrow y_t = \phi_1 y_{t-1} + \epsilon_t \quad (13)$$

Cette équivalence permet de faciliter l'implémentation sous R.

2.3 Implémentation sous R.

Nous avons vu précédemment que l'implémentation du modèle de Vasicek via les séries AR(1) était réalisable sur des séries centrées, ce qui revient à retrancher la moyenne de la série. Après avoir centré les séries, nous pouvons donc rentrer dans le vif du sujet. On note y_t la série centrée. Afin de réaliser cette implémentation, nous devons suivre les étapes suivantes :

2.3.1 Régression linéaire

- La première approche, avec une régression linéaire simple, permet de récupérer les coefficients de la série temporelle. Comme le modèle de Vasicek est similaire à un AR(1) nous avons :

$$y_t = a_i y_{t-1} + W_t$$

où

- Résidus : $W_t \sim \mathcal{N}(0, \sigma^2)$
- Variance résiduelle : σ^2
- Paramètre inconnu : a_i

La régression permet d'obtenir $a_i = e^{-\alpha t}$.

Nous utilisons sous R la fonction `lm` avec :

y_t : la variable à expliquer
 y_{t-1} : la variable explicative

Puis avec `summary(lm)` nous avons un résumé statistique de `lm`. Elle permet aussi de vérifier la véracité du modèle car elle nous donne :

- a_i : le coefficient inconnu.
- Le coefficient de détermination R^2 . Plus il est proche de 1, meilleur est le modèle.

- La seconde approche consiste à calibrer une série chronologique auto régressive (fonction ar sous R) aux données, pour récupérer directement tous les coefficients.

```

a1=ar(yt,order.max=1)
sigma2AR1=a1$var.pred
phi1=a1$ar # expo(-alpha)

```

On calibre la série.
On récupère σ^2 la variance résiduelle
On récupère a_i

2.3.2 Prédiction

Après avoir récupéré les coefficients, nous pouvons donc commencer les prédictions. Nos simulations s'étaleront de 2015 à fin 2018 : il y a donc 16 prédictions à faire (4 données par an comme elles sont trimestrielles). Nous utilisons la fonction "predict" (cette fonction opère le calcul $a_i \cdot y_{t-1}$), pour les 16 données trimestrielles :

```
predi=predict(a1,nhead=16,se.fit=TRUE)
```

Voici les valeurs prédites pour le GAZ LNG 00, l'un de nos quatre bateaux tests :

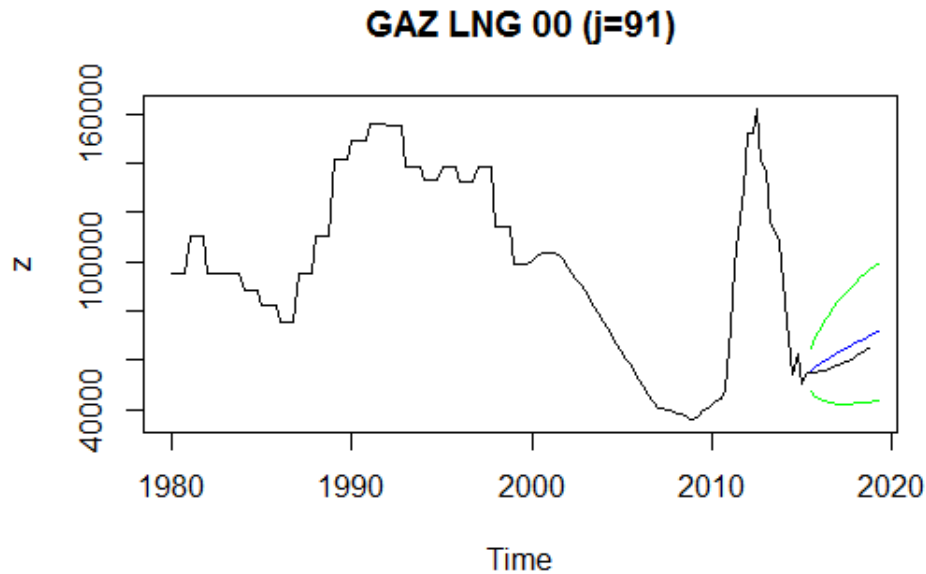
Valeurs réelles (séries 91) GAZ LNG 00				
	Qtr1	Qtr2	Qtr3	Qtr4
2015	50425.20	54627.30	54867.42	54987.48
2016	55587.78	56068.02	56548.26	57148.56
2017	57988.98	58829.40	59669.82	60630.30
2018	61710.84	62791.38	63871.92	64952.46

Série 91: GAZ LNG 00	
Coefficients	Valeurs
a_i	0.76
σ^2	91512364

$$y_t = a_i \cdot y_{t-1} + W_t$$

Valeurs prédites (série 91) GAZ LNG 00				
	Qtr1	Qtr2	Qtr3	Qtr4
2015	55929.18	57193.17	58420.39	59611.90
2016	60768.74	61891.92	62982.41	64041.18
2017	65069.14	66067.19	67036.20	67977.02
2018	68890.46	69777.32	70638.37	71474.38

Voici le graphique que nous obtenons en faisant une prédiction sur un bateau :



2.3.3 Récupération des résidus et test de blancheur.

Si le modèle est bien un AR(1), les résidus sont des bruits blancs gaussiens. Nous allons vérifier si cela est bien le cas.

Le test standard à utiliser sur les résidus est désigné "test du portemanteau" ("test de blancheur").

Celui ci comporte deux tests différents : celui de Box-Pierce et de Ljung-Box. Ces tests sont implémentés sous R via la fonction `Box.test`.

On teste donc les hypothèses :

$$H_0 = \text{"Residus} \sim BB(0, \sigma^2)\text{"}$$

$$H_1 = \text{"}H_0 \text{ fausse"}$$

Si le test a une p-value inférieure à 5% nous rejetons H_0 .

Voici le code R correspondant à cette opération pour le GAZ LNG 00 :

```
# creation de yt_chap sur les données de 2014 à fin 2014
wt_chap=xc[2:length(xc)]-phi1*xc[1:(length(xc)-1)]
#test de blancheur: les résidus sont des bruits blancs
a=Box.test(wt_chap, lag = 2, type = "Box-Pierce")
b=Box.test(wt_chap, lag = 2, type = "Ljung-Box")
# p-value du test
a$p.value=0.0033
b$p.value=0.0028
# nous acceptons l'hypothèse H0 les résidus ne sont pas des bruits blancs.
```

Nous avons vu pu constater que le navire GAZ LNG 00 ne pouvait être modélisé par un modèle de Vasicek. Nous allons voir si cela est aussi le cas pour les autres bateaux.

2.3.4 Déroulement de l'algorithme sur 100 bateaux.

Nous avons un jeu de données composé de 100 bateaux, l'algorithme utilisé sur les 100 bateaux est donc un peu plus précis que l'algorithme de base utilisé sur un seul type de bateau. Cependant le principe reste le même.

Voici quelques idées d'améliorations apportées à l'algorithme de base :

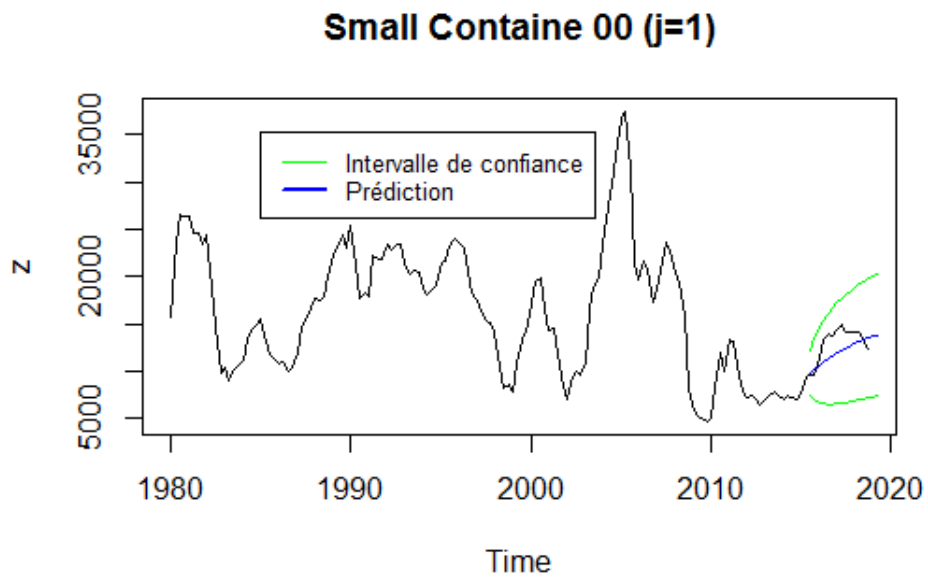
- Création d'une matrice pour stocker les 16 prédictions pour 100 bateaux.
- Création de vecteur pour stocker les p-values du test de blancheur à savoir les p-values des tests de Box-Pierce et de Ljung-Box.
- Nous avons sélectionné les p-values adéquates, c'est-à-dire celles supérieures à 5%. Puis parmi celles ci, nous n'avons gardés que les plus grandes car jugées comme étant les meilleures.
- Création des intervalles de confiance.

Nous avons donc calculé les p-values pour tous les navires. Ci dessous voici les p-values pour nos 4 bateaux tests le CO_SML_00 (série 1),le TKP_MR_00 (série 76), le GAS_LNG_00 (série 91),et la GAS_LPG_00 (série 96) :

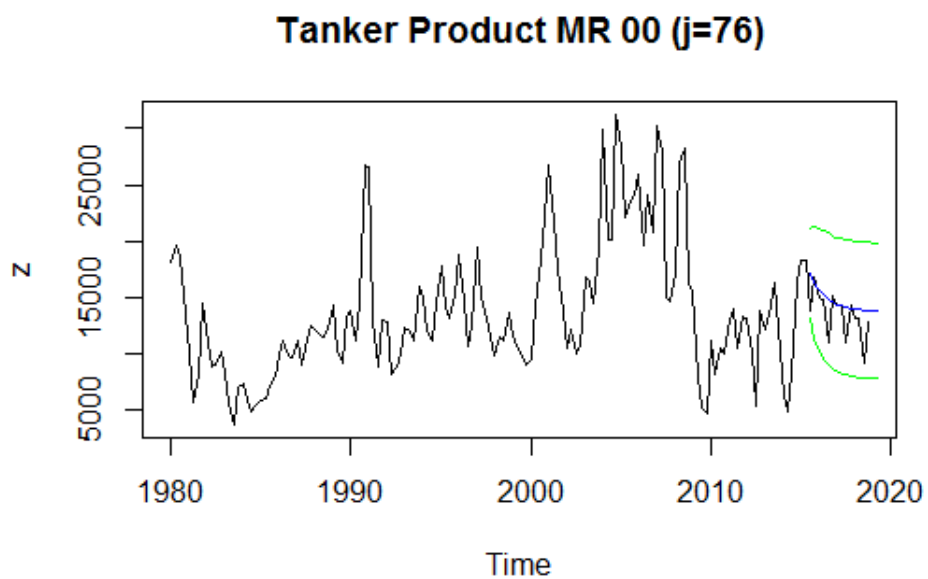
Test	CO SML 00	TKP MR 00	GAS LNG 00	GAS LPG 00
Box-Pierce	1.3031e-07	0.0033	4.5914e-07	8.2033e-05
Ljung-Box	9.1849e-08	0.0028	3.1637e-07	6.6533e-05

Nous pouvons voir que pour ces bateaux, les tests nous indiquent que la modélisation de Vasicek n'est pas adaptée. Cependant, cela n'est pas le cas pour tous les navires, certains d'entre eux peuvent être modélisés par ce processus.

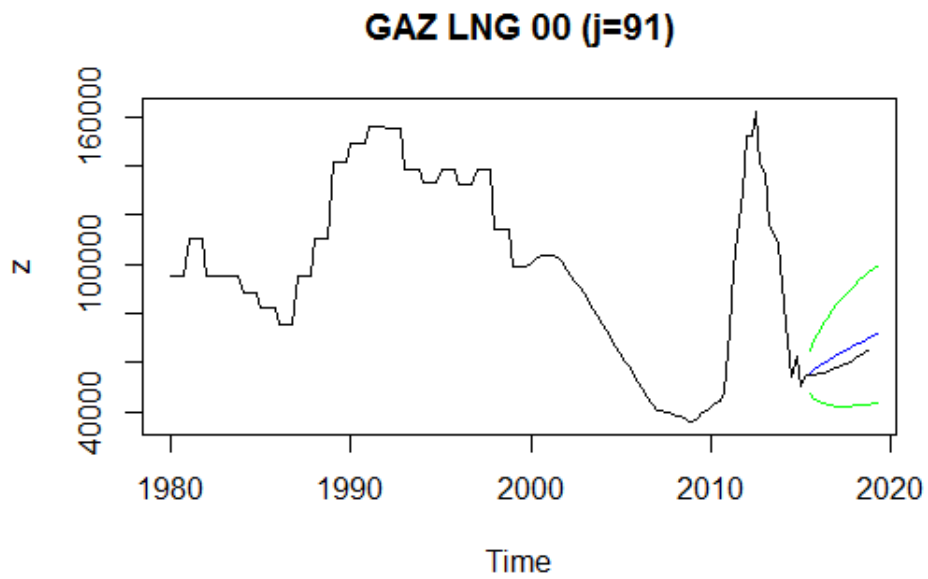
Nous avons tracé les graphiques des prédictions pour les bateaux afin de constater que les prédictions trouvées ne conviennent pas :



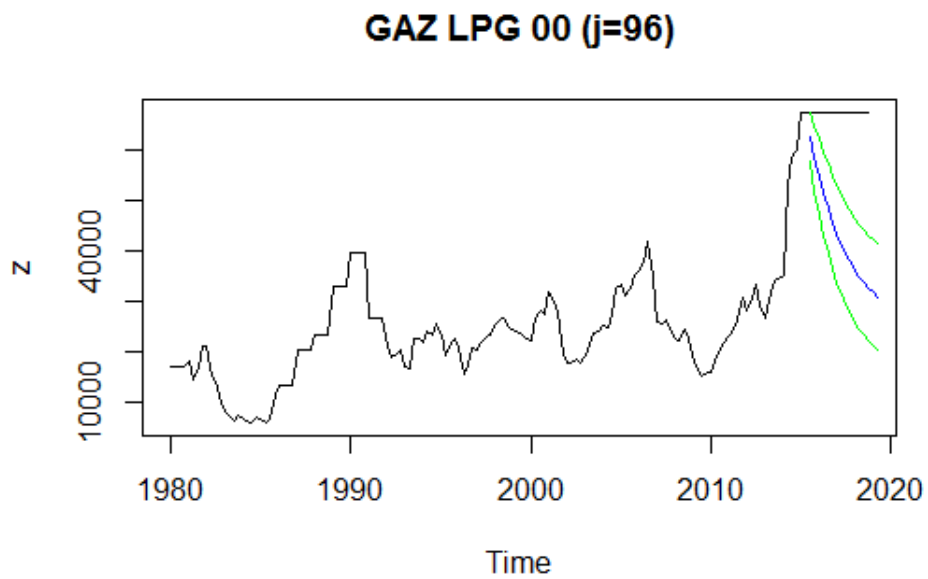
Nous pouvons voir que nos prédictions ne suivent pas celles données par Vessels-Value.



Nous pouvons observer que nous obtenons des intervalles de confiance trop grands. De plus, graphiquement nous pourrions penser que ce modèle est adapté mais nous savons à l'aide des tests de validation du modèle que ce n'est pas le cas.



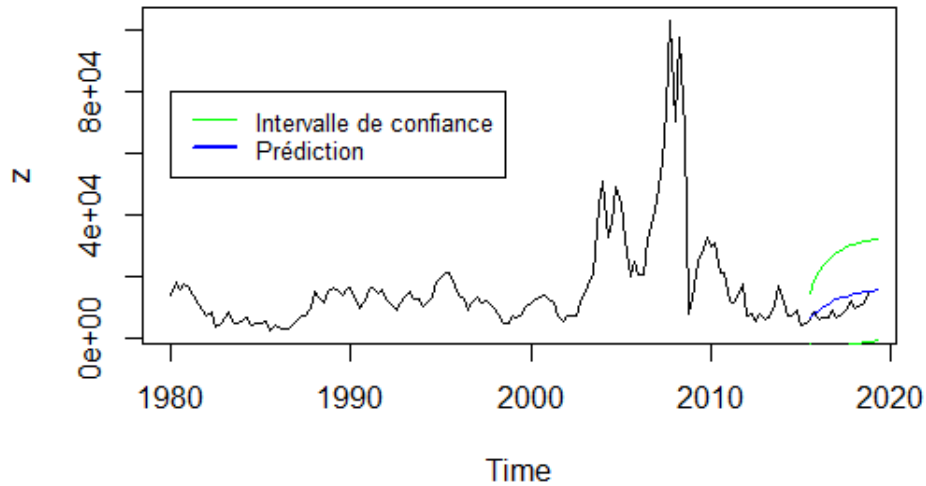
Nous observons ici que notre prédiction envisage une hausse plus élevée que celle de VesselsValue.



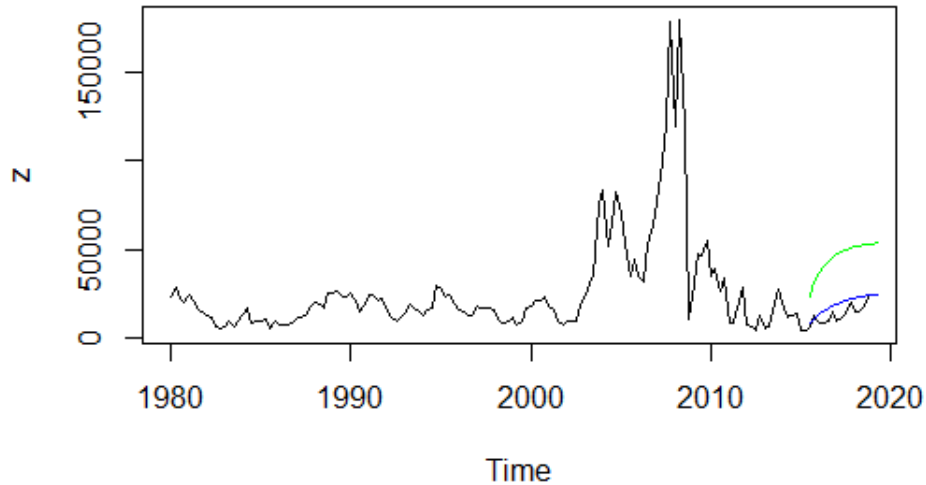
Pour ce dernier, le fait que le modèle n'est pas adapté est flagrant. Nous prédisons une baisse alors que les prédictions de VesselsValue misent sur une stagnation des taux de fret.

Nous décidons de mettre aussi comme exemple les graphiques de deux autres bateaux pour lesquels le modèle de Vasicek est adapté selon les tests de blancheur, le BU CAP 00 et le BU CAP 20 :

BU CAP 00 (j=47)



BU CAP 20 (j=55)



La construction des intervalles de confiance repose sur la somme de la moyenne et plus ou moins l'écart type, ce qui se traduit par la fonction suivante sur R :

$$[\text{mean}(\text{séries})-\text{se}, \text{mean}(\text{séries})+\text{se}]$$

Nous pouvons remarquer que les intervalles de confiance sont assez grands, même pour les navires ayant été sélectionnés par les tests de blancheur. Ce phénomène a été observé sur plusieurs bateaux, malgré le choix de prendre une fluctuation de l'ordre d'un écart-type et non de deux écarts types.

Nous pouvons voir qu'il y a des écarts considérables entre les prévisions et les observations. Afin de mesurer ces écarts, nous avons décidé de calculer l'erreur quadratique moyenne (RMSE : root mean square error en anglais).

Le carré moyen des erreurs ou erreur quadratique moyenne est la moyenne arithmétique des écarts entre les prévisions ($\hat{\theta}$) et les observations (θ). Le RMSE est la racine carrée de ce précédent :

$$RMSE(\hat{\theta}|\theta) = \mathbb{E}((\hat{\theta} - \theta)^2)$$

Sa portée est de 0 à l'infini, un score de 0 étant un score parfait.

Nous avons donc calculé ce coefficient pour ces deux dernières séries et voici ce que nous avons obtenu :

Séries 55				Séries 47		
16 trimestres	série55	predic55	Calcul du RMSE	série47	predic47	Calcul du RMSE
1	4672,0	8398,6	7694,8	4537,4	6538,0	4416,2
2	5078,3	11198,8		4914,2	7939,6	
3	7008,0	13560,5		5890,0	9149,4	
4	12594,1	15552,3		8425,3	10193,6	
5	8531,5	17232,1		6217,5	11094,8	
6	9242,4	18648,8		6469,9	11872,8	
7	10156,5	19843,6		6910,6	12544,2	
8	14930,1	20851,3		9303,0	13123,8	
9	10054,9	21701,2		7020,0	13624,1	
10	11578,4	22417,9		7822,5	14055,9	
11	13711,3	23022,4		8986,7	14428,6	
12	20211,4	23532,2		12369,9	14750,3	
13	15133,2	23962,2		9966,4	15027,9	
14	15641,0	24324,8		10301,7	15267,6	
15	17773,9	24630,7		11398,1	15474,4	
16	23969,3	24888,6		14566,5	15653,0	

Après avoir appliqué l'algorithme, nous avons vu qu'en moyenne, presque la moitié des bateaux pouvaient être modélisés par le modèle de Vasicek. Ci-dessous la liste des bateaux ayant été modélisés via ce dernier :

Vasicek			
jeu de données composé de 5 familles et de 100 bateaux			
Bateaux	Bon modèle: résidus ~BB	Conclusion	
BULK Handy et BULK HandyMax	Oui	50% du jeu de donnée	2 familles sur 5 peuvent être modélisés par vasicek
BULK SupraMax			
BULK Panamax et KamsarMax et Cape			
Tankers Crude Panamax et SuezMax			
Tankers Crude VerylargeCont			

3 Le modèle Cox-Ingersoll-Ross et prédictions du taux de fret.

Le modèle CIR est un modèle permettant de modéliser l'évolution des taux d'intérêts court terme dans la finance. Il a été développé en 1985 par les trois personnes auxquelles il doit son nom : Cox, Ingersoll et Ross.

Afin d'adapter ce modèle à notre jeu de données, nous avons suivi plusieurs étapes :

- Discrétisation de l'intégrale stochastique.
- Détermination des paramètres du modèle.
- Simulation de trajectoires du CIR avec, en paramètres, ceux estimés à l'étape précédente pour effectuer nos prédictions.
- Simulation de trajectoires de CIR, avec un pas de temps Δ fin, avec des paramètres connus à l'avance, afin de faire une étude sur ce modèle, ce qui nous permettra de tester nos fonctions de prédiction.

3.1 Introduction au modèle Cox-Ingersoll-Ross et discrétisation de l'intégrale stochastique.

Afin de modéliser les taux de fret via le modèle CIR, nous avons mis en place le protocole suivant :

- Discrétisation de l'intégrale stochastique.
- Recherche des paramètres α , μ et σ .
- Utilisation du maximum de vraisemblance et optimisation.
- Prédiction du taux de fret.

Nous cherchons donc à discrétiser l'équation différentielle stochastique représentative du modèle CIR :

$$\begin{cases} dr_t = \alpha(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t \\ r_0 = X \end{cases}$$

avec :

- μ la moyenne à long terme.
- α la vitesse à laquelle le processus va converger vers cette moyenne.
- σ la volatilité de ce processus.
- $X \geq 0$.

Comme nous voulons prédire des taux de fret via le modèle CIR, nous allons devoir utiliser un schéma d'Euler.

Afin d'expliquer le déroulement de ce schéma, nous devons d'abord énoncer **le théorème d'Itô**. Les prochaines définitions et démonstrations s'appuieront sur le cours de Mme Pene, "Simulation de Processus" :

Soit $(W_t)_{t \geq 0}$ un mouvement Brownien de dimension d , et $(X_t)_t$ un processus à valeurs dans $\mathcal{C}([0, \infty[\rightarrow \mathbb{R}^n)$, nous posons l'équation différentielle stochastique :

$$\begin{cases} dX_t = \sigma(t, X_t)dW_t + b(t, X_t)dt \\ X_0 = Y \end{cases}$$

avec Y une variable aléatoire dans \mathbb{R}^n et indépendante de W .

Si $(\sigma(t, \cdot))$ et $(b(t, \cdot))$ sont uniformément lipschitziennes et s'il existe une constante $c \in]0, \infty[$ telle que pour tout $t \in [0, T]$, et pour tout $x, y \in \mathbb{R}^n$:

$$|\sigma(t, x)| \leq c(1 + |x|), \quad |b(t, x)| \leq c(1 + |x|)$$

alors pour tout Y , il existe un processus $(X_t)_t$ à valeurs dans \mathbb{R}^n tel que : pour tout $t \in [0, T]$, X_t est $\sigma(Y, \sigma(W_s)_{s \in [0, t]})$ mesurable avec :

$$\int_0^t |\sigma(s, X_s)|^2 ds + \int_0^t |b(s, X_s)| ds < \infty$$

et,

$$X_t = Y + \int_0^t \sigma(s, X_s)dW_s + \int_0^t b(s, X_s)ds$$

De plus, pour tout $p \in \mathbb{N}^*$, il existe $c_p > 0$ tel que pour toute variable aléatoire Y (avec $\mathbb{E}[|Y|^{2p}] < \infty$),

$$\sup_{t \in [0, T]} (|X_t|) \in \mathbb{L}^{2p} \iff \mathbb{E}[\sup_{t \in [0, T]} (|X_t|^{2p})] \leq c_p(1 + \mathbb{E}[|Y|^{2p}])$$

Nous posons à présent le schéma d'Euler³ :

Soit $N \geq 1$ et $h = \frac{T}{N}$. Afin d'approcher la solution de $(X_t)_t$:

$$\begin{cases} dr_t = \alpha(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t \\ r_0 = X \end{cases}$$

Il nous faut d'abord discrétiser l'intervalle $[0, T]$ ainsi : $(0, \frac{T}{N}, \frac{2T}{N}, \dots, T)$ puis, nous pouvons écrire le schéma d'Euler : pour tout $k = 0, \dots, N - 1$, pour tout $t \in [kh, (k + 1)h]$:

$$kh \geq t \geq (k + 1)h, \quad \bar{r}_{(t)}^{(N)} = \bar{r}_{(kh)}^{(N)} + \alpha(\mu - \bar{r}_{(kh)}^{(N)})(t - kh) + \sigma\sqrt{\bar{r}_{(kh)}^{(N)}}(W_t - W_{kh})$$

Mais comme $f(x) = \sqrt{x}$ n'est pas lipschitzienne, la convergence du schéma d'Euler n'est

3. Méthode numérique de résolution des équations différentielles du premier ordre avec une condition initiale par approximation. C'est la plus simple des méthodes de résolution des équations différentielles.

pas assurée par le théorème énoncé précédemment. De plus, comme X_t peut devenir négatif, il y a une incohérence du fait de la racine carrée.

Il faut donc utiliser : $r_t = \max(0, r_t)$. Ce qui nous donne alors avec $kh \leq t \leq (k+1)h$:

$$\begin{cases} \bar{r}_{(t)}^{(N)} = \bar{r}_{(kh)}^{(N)} + \alpha(\mu - \max(\bar{r}_{(kh)}^{(N)}, 0))(t - kh) + \sigma\sqrt{\max(\bar{r}_{(kh)}^{(N)+}, 0)}(W_t - W_{kh}) \\ \bar{r}_{(0)}^{(N)} = X \end{cases}$$

C'est ce dernier schéma que nous avons utilisé afin de rechercher les paramètres inconnus α, μ , et σ .

3.2 Détermination des paramètres $\theta = (\alpha, \mu, \sigma)$.

3.2.1 Détermination des paramètres α et μ via une régression linéaire.

Dans cette partie, la détermination des paramètres va se faire à l'aide de la formule :

$$\bar{r}_{(t)}^{(N)} - \bar{r}_{(kh)}^{(N)} = \alpha(\mu - \max(\bar{r}_{(kh)}^{(N)}, 0))(t - kh) + \sigma\sqrt{\max(\bar{r}_{(kh)}^{(N)+}, 0)}(W_t - W_{kh})$$

Qui peut aussi s'écrire :

$$\bar{r}_{(t)}^{(N)} - \bar{r}_{(kh)}^{(N)} = \alpha(\mu - \max(\bar{r}_{(kh)}^{(N)}, 0))(t - kh) + \sigma\sqrt{\max(\bar{r}_{(kh)}^{(N)+}, 0)}\epsilon_t$$

avec $\epsilon_t = (W_t - W_{kh})$

$$\iff z_t = \alpha\mu\Delta t x_t - \alpha\Delta t y_t + \sigma\epsilon_t \quad (14)$$

en posant :

$$z_t = \frac{\bar{r}_{(t)}^{(N)} - \bar{r}_{(kh)}^{(N)}}{\sqrt{\bar{r}_{(kh)}^{(N)+}}}$$

$$x_t = \frac{1}{\sqrt{\bar{r}_{(kh)}^{(N)+}}}$$

$$y_t = \sqrt{\bar{r}_{(kh)}^{(N)+}}$$

$$\Delta t = (t - kh)$$

Ici, comme les taux que nous étudions sont non nuls et positifs, nous pouvons diviser par la racine de $\bar{r}_{(kh)}^{(N)}$.

Nous pouvons à présent, souligner quelques points importants :

- L'équation (14) nous permettra par la suite, de faire une régression, nous développons la méthodologie ci-dessous, afin d'obtenir les différents paramètres.
- Nous avons 3 paramètres inconnus α, μ et σ .
- Une régression sur les séries z_t, y_t et x_t permettra d'estimer les paramètres α et μ .

- Une méthode calculatoire sera utilisée pour l'estimation de σ .

Nous créons donc le vecteur $\theta = (\alpha, \mu, \sigma)$ contenant les paramètres du modèle :

α : la vitesse de retour à la moyenne du taux.

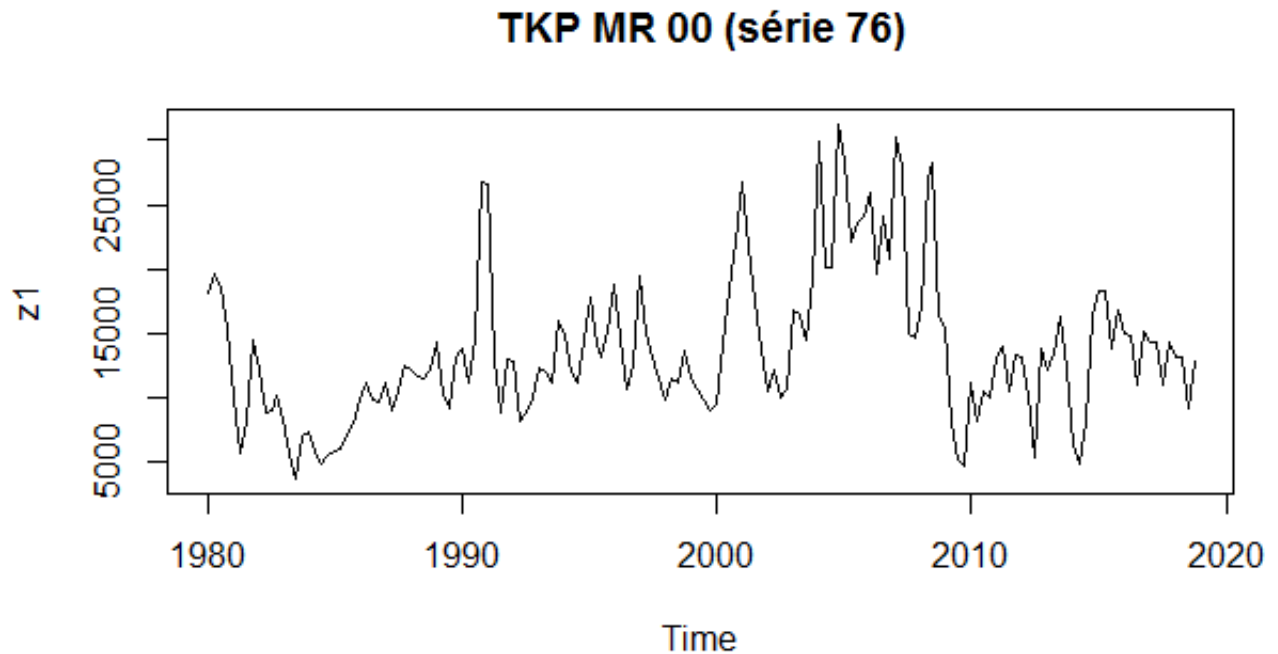
μ : représente la moyenne de long terme du taux court.

σ : représente la volatilité du taux de court terme.

Si α, μ, σ sont tous positifs et que $2\alpha\mu \geq \sigma^2$, alors le processus CIR admet une solution continue, et strictement positive. Nous concentrons nos recherches sur la série 76 (TKP MR 00).

Nous allons à présent expliquer le déroulement de l'algorithme sur la série 76 (TKP MR 00).

Ci dessous le graphique représentant cette série :



Nous fixons un pas de temps $\Delta t = 1$.

Comme nous l'avons précisé ci dessus, nous avons dû faire une régression linéaire afin de déterminer les paramètres α et μ .

Nous rappelons ici le principe de cette méthode. Nous nous appuyons sur le cours de M Ailliot sur les "Modèles linéaires" pour les prochaines définitions. Ce schéma est un modèle de régression d'une variable expliquée sur une ou plusieurs variables explicatives. Nous faisons l'hypothèse que la fonction qui relie les variables explicatives à la variable expliquée est linéaire dans ses paramètres. Elle est représentée à l'aide de la notation simple :

$$y_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_K x_{K,i} + W_i$$

où y_i est la variable à expliquer, $\beta_0, \beta_1, \dots, \beta_K$ sont les paramètres inconnus, $x_{1,i}, \dots, x_{K,i}$

les variables explicatives, (W_1, \dots, W_n) des variables aléatoires iid tel que $\sigma^2 = \text{var}(W_i) < \infty$ et $\mathbb{E}(W_i) = 0$. Le résidu W_i modélise l'écart entre la variable à expliquer Y_i et la combinaison linéaire des variables explicatives.

Ce qui est équivalent matriciellement à :

$$Y = X\beta + W$$

avec :

$$Y = \begin{pmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,K} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 1 & x_{n,1} & \dots & x_{n,K} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \cdot \\ \cdot \\ \cdot \\ \beta_K \end{pmatrix}, W = \begin{pmatrix} W_1 \\ \cdot \\ \cdot \\ \cdot \\ W_n \end{pmatrix}$$

Les coefficients sont déterminés par la méthode des moindres carrés. Nous appelons les estimateurs des moindres carrés de β le vecteur aléatoire :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

et l'estimation des moindres carrés, la quantité :

$$\hat{b} = (X'X)^{-1}X'y$$

avec X' la transposée de X .

Ainsi, nous pouvons en déduire la valeur des paramètres recherchés.

Dans notre cas, pour obtenir les paramètres α et μ , nous calculons d'abord les trois suites x_t , y_t , et z_t pour toutes les valeurs des taux de fret que nous possédons afin d'avoir notre matrice X . Puis, nous utilisons les 3 séries et l'équation linéaire :

$$z_t = \alpha\mu\Delta tx_t - \alpha\Delta ty_t + \sigma\epsilon_t$$

Le schéma d'Euler nous aidera à pallier le problème lié à la racine carrée car nous prendrons le maximum entre 0 et les taux r_t , en sachant que ces taux sont de toute manière non nuls, contenus dans les trois suites : x_t , y_t , et z_t . Nous allons pouvoir faire une régression via le logiciel R sur la formule précédente avec la commande :

$$lm(zt \sim ., data = \text{tauxdefret})$$

Afin d'illustrer nos résultats, nous choisissons de dérouler les prochains calculs sur le bateau TKP_MR_00 (série 76). Nous obtenons alors :

$$\alpha = 0.2219449 \text{ et } \mu = 13749.49$$

3.2.2 Détermination de σ

Afin d'obtenir σ , nous utilisons la formule présentée précédemment :

$$z_t = \alpha\mu\Delta tx_t - \alpha\Delta ty_t + \sigma\epsilon_t$$

Puis nous faisons une approximation pour trouver la valeur de σ en prenant l'écart type des résidus :

$$\sigma = sd(residus) \quad \# \text{ calcul de l'écart type}$$

avec $residus = \sigma\epsilon_t$. Par R, nous obtenons pour le TKP_MR_00 (série 76) :

$$\sigma = 32.4902$$

3.2.3 Logarithme du maximum de vraisemblance.

L'estimation du maximum de vraisemblance est une méthode statistique courante pour déterminer une estimation des paramètres α, μ , et σ dans le modèle CIR.

Les prochains raisonnements se sont appuyés sur des démonstrations faites dans l'article "**Maximum likelihood estimations of the Cox-Ingersoll-Ross process : the matlab implementation**" par Kamil Kladvko. Afin de calculer le logarithme du maximum de vraisemblance, nous devons d'abord calculer pour un taux r_t donné au temps t , la densité de $r_{t+\Delta t}$ au temps $t + \Delta t$, sachant r_t à l'instant t :

$$p(r_{t+\Delta t}|r_t; \theta, \Delta t) = ce^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q(2\sqrt{uv})$$

$$L(\theta) = \prod_{i=1}^{n-1} p(r_{t_{i+1}}|r_{t_i}; \theta, \Delta t)$$

avec $\theta = (\alpha, \mu, \sigma)$.

Et lorsque l'on développe cette dernière équation nous trouvons :

$$\ln(L(\theta)) = (n-1)\ln(c) + \sum_{i=1}^{n-1} (-u_{t_i} - v_{t_{i+1}}) + 0.5 \cdot q \cdot \ln\left(\frac{v_{t_{i+1}}}{u_{t_i}}\right) + \ln(I_q(2\sqrt{u_{t_i}v_{t_{i+1}}})) \quad (15)$$

avec : $c = \frac{2\alpha}{\sigma^2(1-e^{-\alpha\Delta})}$, $u = cr_t e^{-\alpha\Delta}$, $v = cr_{t+\Delta t}$, et $q = \frac{2\alpha\mu}{\sigma^2} - 1$

Nous avons aussi $I_q(2\sqrt{uv})$ la fonction modifiée de Bessel du premier genre d'ordre 1. Si nous avons $\alpha, \mu, \sigma \geq 0$ et $2\alpha\mu \geq \sigma^2$ alors le CIR est bien défini et a une distribution marginale stable. La densité marginale a une distribution gamma.

Nous cherchons donc les estimateurs :

$$\hat{\theta} = (\hat{\alpha}, \hat{\mu}, \hat{\sigma}) = \operatorname{argmax}(\ln(L(\theta)))$$

Cela revient à utiliser la fonction "optim(par,fn)" sous R avec :

- par= la valeur initiale du paramètre à optimiser.
- fn= une fonction à minimiser (ou maximiser).

Cette fonction se base sur l'algorithme du simplexe de Nelder-Mead qui est une méthode de résolution des optimisations linéaires, la méthode de quasi Newton qui permet de résoudre des systèmes d'équations non linéaires et des algorithmes du gradient qui servent à minimiser une fonction dérivable.

Nous choisissons $fn = -\ln(L(\theta))$. **Nous pouvons noter que si "fn" est négative la fonction "optim" maximise.**

Après exécution de l'algorithme, nous nous sommes rendus compte qu'il ne fonctionnait pas car nous trouvions un maximum de vraisemblance égal à l'infini. Nous ne savons pas si ce problème provient de nos codes R ou si cela est dû au logiciel R qui ne serait pas assez précis pour traiter les valeurs très petites.

Nous avons donc décidé de mettre en place **deux approches différentes** :

- La première nous permet de faire nos prédictions en ne considérant pas l'étape du max de vraisemblance.
- L'autre nous a permis de vérifier si nos fonctions étaient correctes.

Première approche Nous avons dans un premier temps décidé de faire nos prédictions en omettant le calcul du maximum de vraisemblance. Nous avons fait des prédictions sur l'ensemble du jeu de données ; nous avons décidé afin de les illustrer, de présenter nos résultats pour les navires que nous avons sélectionnés dans les parties précédentes : les bateaux CO_SML_00 (série 1), TKP_MR_00 (série 76), GAS_LNG_00 (série 91), et GAS_LPG_00 (série 96).

A l'aide de cette méthode, nous simulons des trajectoires de CIR avec les paramètres que nous avons estimés précédemment. En prenant la trajectoire moyenne de toutes les simulations, nous obtenons notre prédiction et les simulations les plus éloignées de la moyenne forment notre intervalle de confiance.

Nous avons regroupé les prédictions des quatre bateaux dans le tableau ci-dessous :

Temps	CO_SMI_00	TKP_MR_00	GAS_LNG_00	GAS_LPG_00
2015,2	9 403,20	18 274,53	54 627,30	67 337,54
2015,3	9 516,39	17 170,44	55 249,23	67 647,91
2015,4	9 989,87	16 525,80	56 200,25	68 191,32
2016,1	10 145,39	16 298,93	56 892,37	68 614,70
2016,2	10 432,10	15 849,47	56 746,20	68 578,01
2016,3	10 610,16	15 486,75	56 125,14	69 828,17
2016,4	11 051,68	15 115,43	56 657,27	69 911,91
2017,1	11 361,07	14 906,94	56 727,55	70 711,87
2017,2	11 340,29	14 859,85	56 450,53	70 206,50
2017,3	11 810,30	14 629,88	56 787,06	69 894,66
2017,4	12 070,92	14 383,67	56 310,81	69 781,26
2018,1	12 228,08	14 237,05	57 355,44	68 852,68
2018,2	12 754,25	14 030,56	57 764,03	69 478,21
2018,3	12 905,18	13 516,54	57 568,04	69 968,13
2018,4	13 230,47	13 662,07	57 780,01	70 490,00
2019,1	13 134,64	13 983,90	57 834,39	70 297,11
2019,2	13 259,63	13 480,06	58 041,84	70 308,67
2019,3	13 535,56	13 439,61	59 254,68	70 883,36
2019,4	13 730,16	13 936,89	59 190,81	71 953,05
2020,1	14 072,64	13 894,01	59 974,32	72 343,03

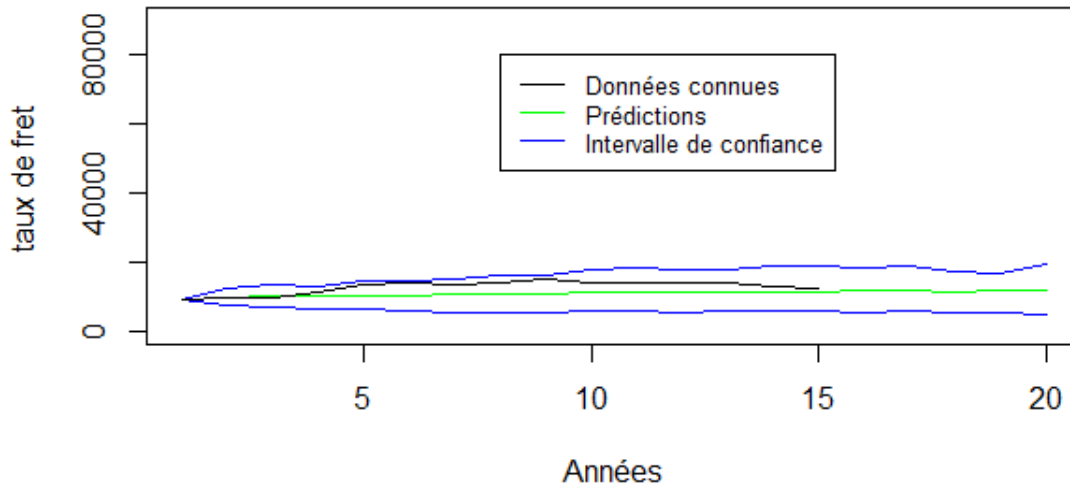
Nous pouvons les comparer aux valeurs fournies par VesselsValue :

Les valeurs fournies par VesselsValue				
Date	CO_SML_00 (série 1)	TKP_MR_00 (série 76)	GAS_LNG_00 (série 91)	GAS_LPG_00 (Série 96)
2015,1	7675,199	18373,313	50425,2	67337,543
2015,2	9403,195	18274,531	54627,3	67337,543
2015,3	9619,029	13829,375	54867,42	67337,543
2015,4	9511,941	16891,594	54987,48	67337,543
2016,1	11347,689	15014,75	55587,78	67337,543
2016,2	13291,187	14915,969	56068,02	67337,543
2016,3	13939,02	11063,5	56548,26	67337,543
2016,4	13723,186	15113,531	57148,56	67337,543
2017,1	14263,6	14323,281	57988,98	67337,543
2017,2	14911,433	14422,063	58829,4	67337,543
2017,3	14047,434	10964,719	59669,82	67337,543
2017,4	14047,434	14422,063	60630,3	67337,543
2018,1	14047,766	13236,688	61710,84	67337,543
2018,2	14047,766	13137,906	62791,38	67337,543
2018,3	13075,022	9285,438	63871,92	67337,543
2018,4	12426,857	12841,563	64952,46	67337,543

Et voici les types de graphiques que nous obtenons avec des intervalles de confiance à 90% :

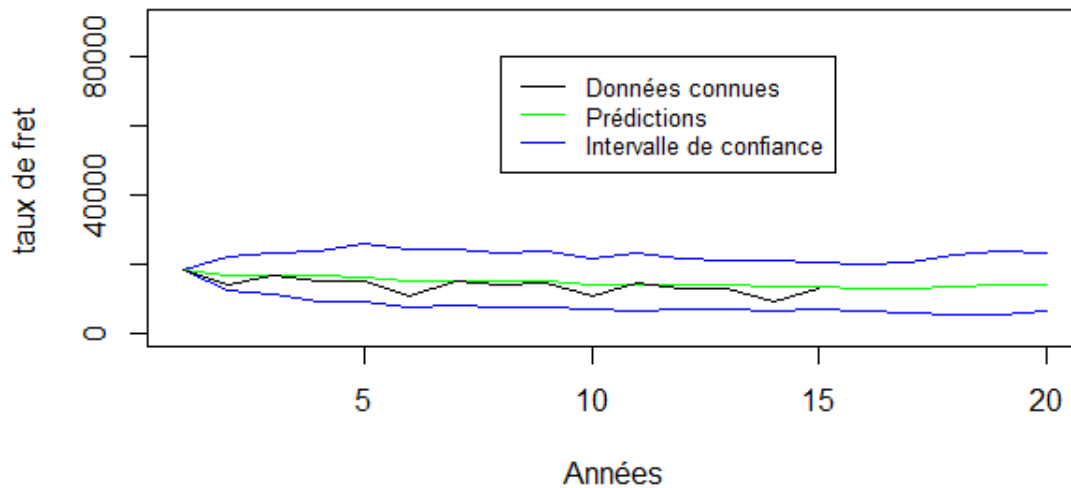
Graphique du Small Container 00 (série 1) :

Prédictions à partir de 2015,3 à 2020,1

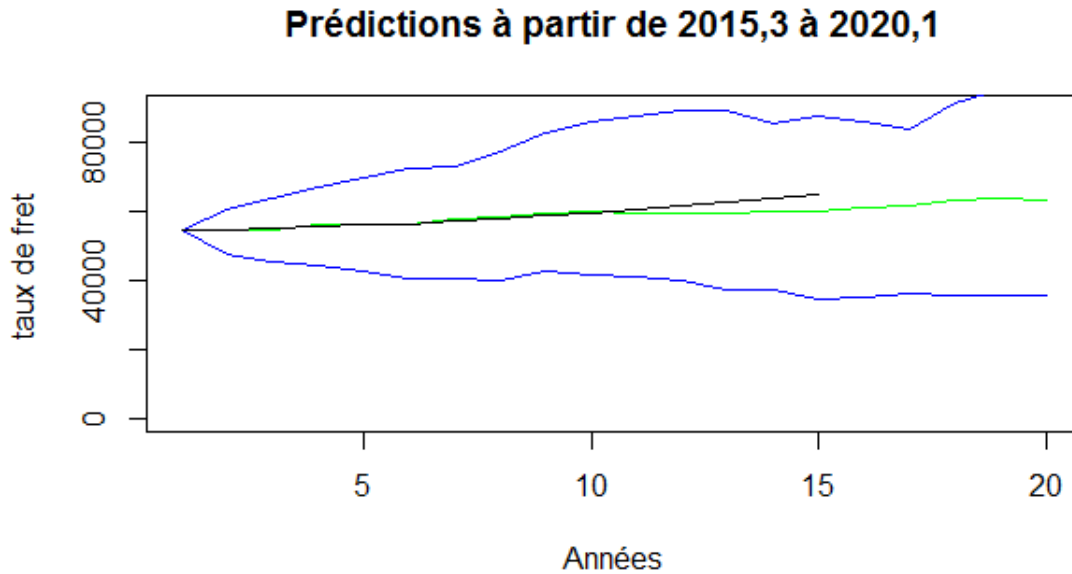


Graphique du TKP_MR_00 (série 76) :

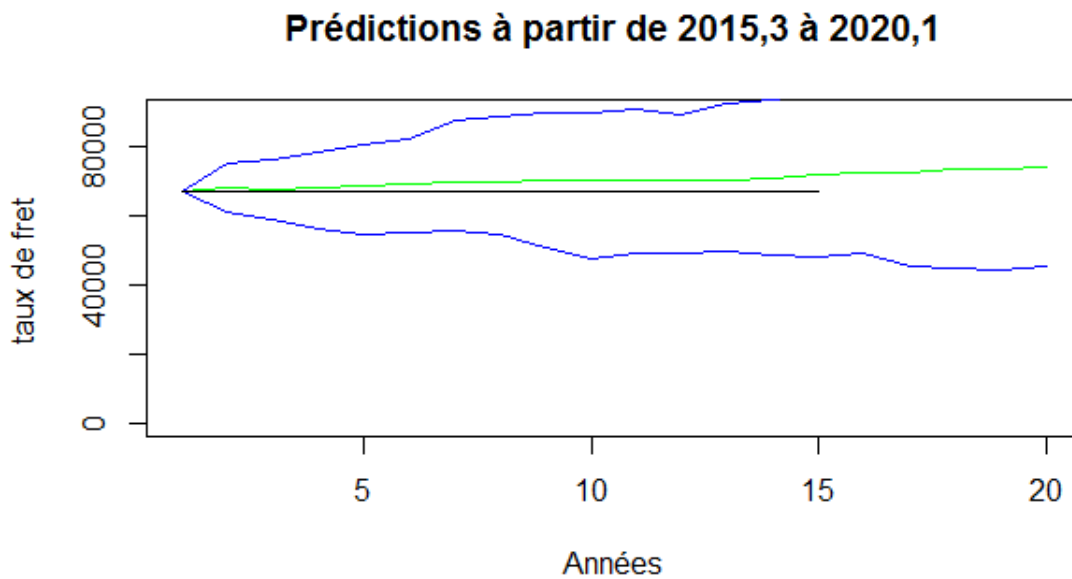
Prédictions à partir de 2015,3 à 2020,1



Graphique du GAZ_LNG_00 (série 91) :



Graphique du GAZ_LPG_00 (série 96) :



Nous pouvons constater sur ces graphiques que les résultats que nous trouvons sont assez proches des prédictions que nous avait fournies VesselsValue. Par ailleurs, elles se situent bien dans l'intervalle de confiance. Nous avons voulu vérifier de manière numérique si nos prédictions étaient acceptables, nous avons donc calculé les RMSE pour les quatre bateaux précédents :

	CO SML 00	TKP MR 00	GAZ LNG 00	GAZ LPG 00
RMSE	7887	8606	7892	8669

Les résultats que nous obtenons sont relativement faibles, ce qui pourrait nous conforter dans le fait que nos prédictions sont correctes. Cependant, nous validerons cette hypothèse lors de nos comparaisons avec les autres modèles dans la dernière partie de ce rapport.

Cependant, nous voulions être certains que nos résultats étaient corrects, c'est dans ce but que nous avons décidé de mettre en place une autre méthode se basant sur la simulation de trajectoires du CIR.

3.3 Simulation de trajectoires du CIR.

Malgré des résultats satisfaisants pour le modèle CIR, nous voulions vérifier s'ils sont optimaux. Nous avons testé nos fonctions sur d'autres modèles de CIR dont les paramètres étaient connus à l'avance. La méthode utilisée pour tester la validité de nos estimations pour α, μ , et σ a été de simuler une trajectoire de CIR avec un pas de temps fin afin de vérifier que les paramètres obtenus avec cette série sont similaires aux paramètres initiaux fixés.

3.3.1 Protocole.

Voici les étapes que nous avons suivies :

- Dans un premier temps nous avons effectué une simulation du CIR (de 10000 valeurs) avec des constantes prédéfinies pour α, μ , et σ .
- Une régression comme nous l'avons vu précédemment en partie 2.2 (x_t, y_t, z_t) nous permet d'estimer les paramètres α, μ , et σ . Parmi les valeurs simulées, nous n'en avons gardé que 200 sur les 10000 précédentes.
- Enfin nous avons appliqué le logarithme du maximum de vraisemblance sur ces dernières et au voisinage des α, μ et σ . Finalement, nous avons effectué une optimisation sur le logarithme du maximum de vraisemblance et nous avons comparé les constantes optimisées aux paramètres initiaux.

3.3.2 Simulation de la trajectoire du CIR.

Comme nous l'avons vu précédemment, la discrétisation de l'équation différentielle stochastique nous donne l'équation :

$$\bar{r}_{(t)}^{(N)} - \bar{r}_{(kh)}^{(N)} = \alpha(\mu - \max(\bar{r}_{(kh)}, 0))(t - kh) + \sigma \sqrt{\max(\bar{r}_{(kh)}^+, 0)} \epsilon_t$$

L'idée est la suivante :

On fixe un vecteur $\theta = (\alpha, \mu, \sigma)$ avec :

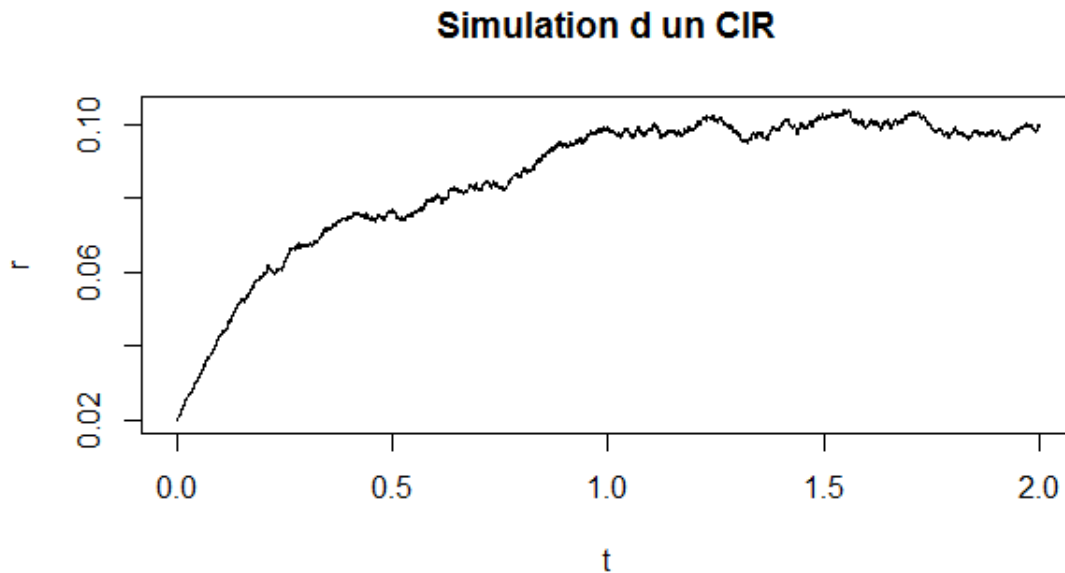
$$\begin{cases} \alpha = 0.3 \\ \mu = 0.1 \\ \sigma = 0.01 \end{cases}$$

Puis nous utilisons l'équation précédente pour simuler notre trajectoire du CIR avec :

$$\begin{cases} r_t = \text{CIR simulé} \\ \Delta t = \text{le pas de temps} \\ \epsilon_t \sim \mathcal{N}(0, \sigma^2) \text{un bruit blanc gaussien} \end{cases}$$

Dans notre implémentation sous R, nous avons pris un pas de temps fin de l'ordre de $\frac{T}{N}$ avec $N = 10000$.

Voici le type de trajectoire que nous obtenons lors d'une simulation :



3.3.3 Calcul du logarithme du maximum de vraisemblance.

Pour effectuer ce calcul, nous sommes passés par des étapes intermédiaires. Nous n'avons pas directement appliqué le logarithme du maximum de vraisemblance sur l'ensemble du CIR simulé, mais seulement sur certaines valeurs que nous avons extraites.

Par exemple, pour un CIR de 10000 valeurs, nous avons retenu 200 valeurs sur lesquelles nous avons appliqué le maximum de vraisemblance. Pour y parvenir nous avons appliqué l'algorithme suivant :

Nous avons d'abord défini : $\Delta = 0.1$.

Puis nous avons choisi un nombre n tel que :

$n = \frac{T}{\Delta}$ qui est le nombre d'observations de la trajectoire extraite.

Une fois la taille du vecteur connue, nous avons appliqué le procédé ci dessous :

- Extraction des indices :

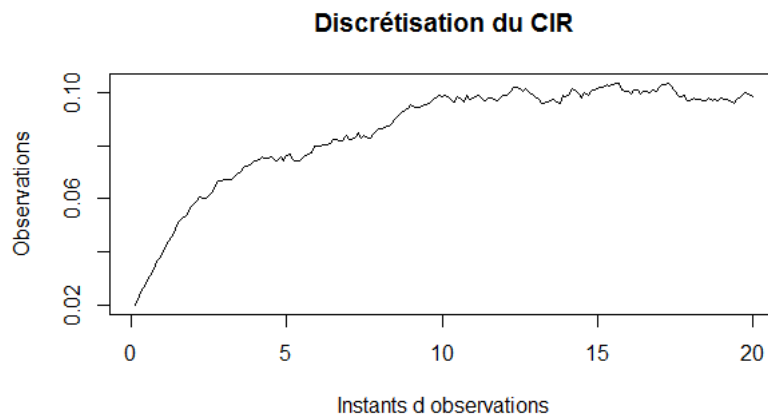


- Extraction des valeurs : Après avoir stocké les indices comme expliqué ci-dessus il suffit de prendre la valeur correspondante à l'indice.

Voici le code R correspondant :

```
##### extraction d'une trajectoire de pas delta #####
n=T/delta
  #nombre d'observations de la trajectoire extraite
t=seq(1,N,by=N/n)
  #indices pour ces observations dans simul
xe=simul[t]
  #extraction des n obs ie discretisation du CIR
temps=seq(delta,T,by=delta) #instants d'observation
plot(temps,xe,type='l')
```

Ci dessous, le graphique du CIR construit sur 200 valeurs :



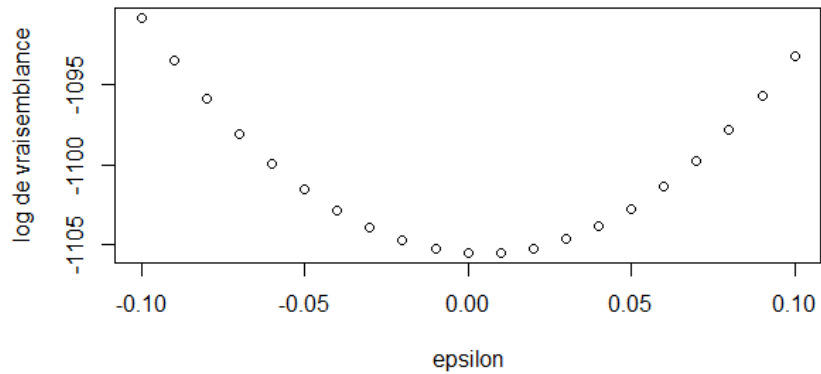
Nous avons ainsi créé un CIR composé de 200 valeurs sur lequel nous appliquerons le logarithme du maximum de vraisemblance. L'étape suivante est de calculer le maximum du log de vraisemblance comme nous l'avons déjà fait dans la partie 2.2.3.

Après cela, nous avons créé une fonction permettant de vérifier que nos estimations pour les paramètres de θ tendaient bien vers les bonnes valeurs. Figures ci dessous les graphiques du logarithme de vraisemblance aux voisinages des trois paramètres ; nous pouvons constater que la valeur minimale pour chacun est :

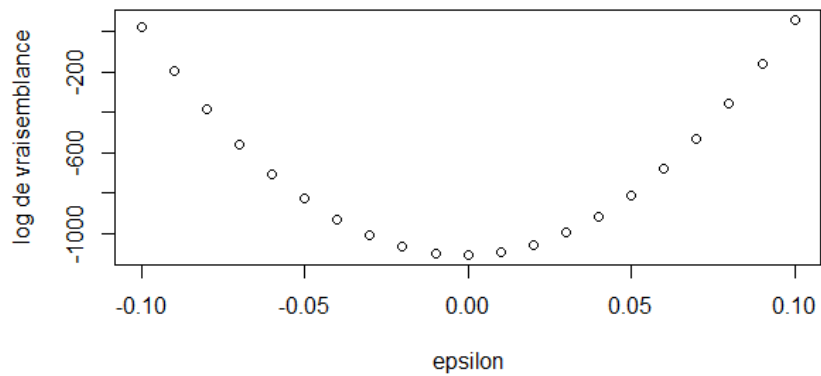
Paramètres	Valeur du paramètre	Valeur du log de vraisemblance
α	0.3	-1102.692
μ	0.1	-1102.692
σ	0.0105	-1103.237

Voici les graphiques associés à ces résultats :

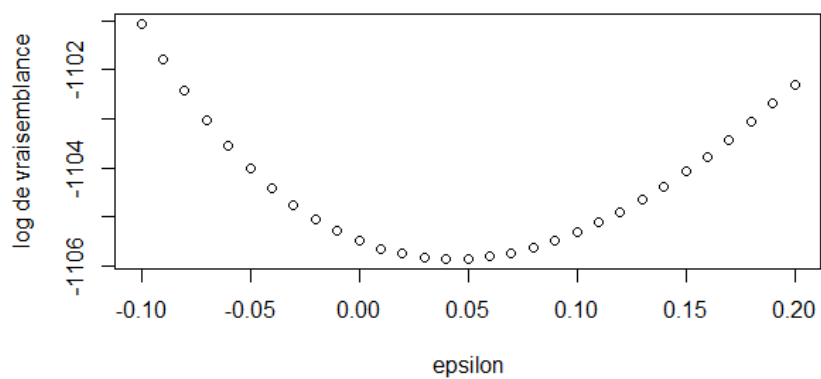
Estimation du log de vraisemblance au voisinage de alpha



Estimation du log de vraisemblance au voisinage de mu



Estimation du log de vraisemblance au voisinage de sigma



Ces trois graphiques ont été tracés avec les paramètres initiaux. Nous pouvons voir que la fonction de vraisemblance fonctionne. Cela nous laisse penser que si nous ne choisissons pas des valeurs très proches des valeurs des paramètres à estimer pour l'optimisation, l'algorithme ne converge pas. C'est peut-être l'une des raisons pour laquelle nous n'avons pas réussi à calculer le maximum de vraisemblance, car il n'y a pas convergence lorsque

nous nous éloignons trop des paramètres. En effet, nous avons fait d'autres tests qui nous laissent penser que si la méthode d'optimisation du maximum de vraisemblance nous renvoyait de mauvais résultats, cela est dû à l'estimation de σ par la méthode de régression linéaire qui n'est pas assez précise.

4 Comparaison des prédictions des différents modèles.

4.1 Comparaison entre les différents modèles.

Nos prédictions nous ont permis de constater que chaque modèle utilisé présentait des avantages et des inconvénients comparés les uns aux autres. Nous allons tenter ici de dégager lequel est le meilleur pour prédire les taux de fret.

4.1.1 Le modèle SARIMA.

Avantages :

- Ce modèle est bien adapté au cas des séries temporelles présentant une décroissance très lente de la fonction d'auto-covariance (présence d'une tendance).
- SARIMA permet de prendre en compte la saisonnalité.

Inconvénients.

- Pour les séries ARMA, ce modèle n'est approprié que si la série chronologique est stationnaire.

4.1.2 Le modèle Vasicek.

Avantages :

- C'est un modèle simple à comprendre d'un point de vue théorique.
- Il fournit des expressions analytiques pour le calcul des taux de fret.
- Il y a peu de paramètre à calculer pour obtenir la prédiction.
- D'un point de vue informatique, il fournit des réponses rapidement.
- C'est un modèle à un facteur.

Inconvénients :

- Les variations des taux sont corrélées entre elles.
- Les taux peuvent devenir négatifs car ils suivent un processus gaussien.

4.1.3 Le modèle Cox-Ingersoll-Ross.

Avantages :

- Ce modèle interdit les taux négatifs.
- Peu de paramètres à calculer.

Problème rencontré :

- La mauvaise estimation de σ par la méthode de régression linéaire empêche la méthode d'optimisation du max de vraisemblance de fonctionner correctement .

4.2 Comparaison des prédictions.

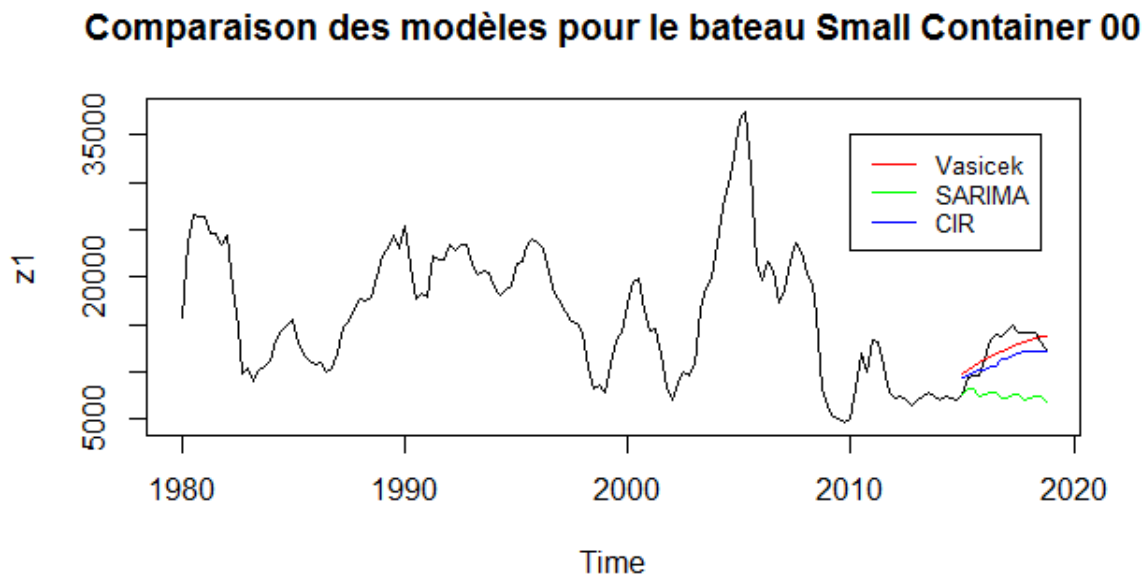
4.2.1 Confrontation des résultats entre les modèles.

Dans les parties précédentes, nous avons comparé à chaque fois les résultats pour les quatre mêmes bateaux, mais dans cette partie nous avons préféré sélectionner un bateau pour chaque grande classe de navires que nous avons : les Containers, les Bulks, les Tankers Crude, les Tanker Product et les Gaz Tanker. Nous avons donc fait nos calculs sur : le CO SML 0, le BU HDY 00, le TKC PNX 00, le TKP MR 00, et le GAZ LNG 00.

Méthode graphique.

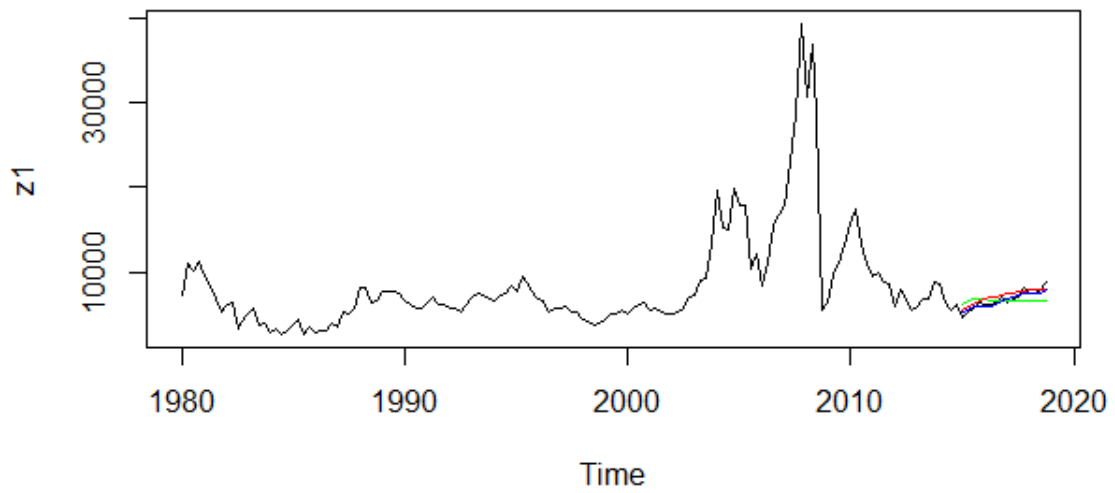
Nous avons, dans un premier temps, tracé les graphiques des prévisions pour chaque modèle pour chaque type de navires afin, de comparer nos différentes prédictions avec les valeurs données par VesselsValue.

Graphique du Small Container 00 (série 1) :



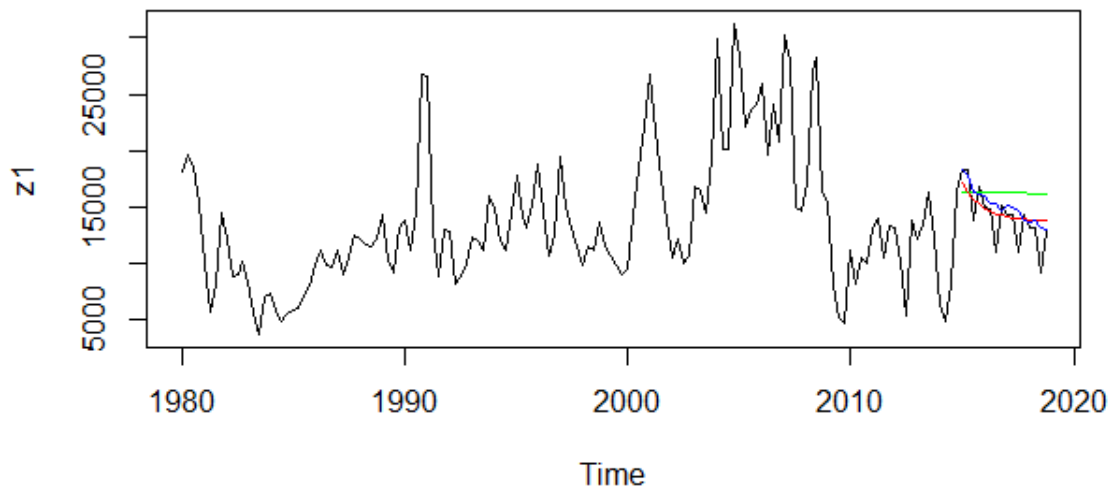
Graphique du BU HDY 00 (série 21) :

Comparaison des modèles pour le bateau BU HDY 00



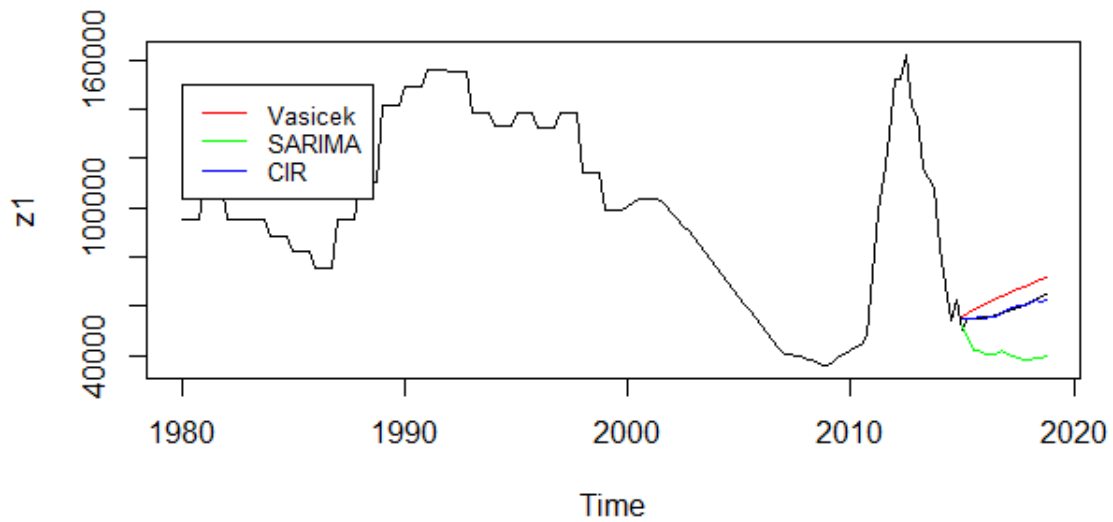
Graphique du Tanker Product MR 00 (série 76) :

Comparaison des modèles pour le bateau Small Container 00



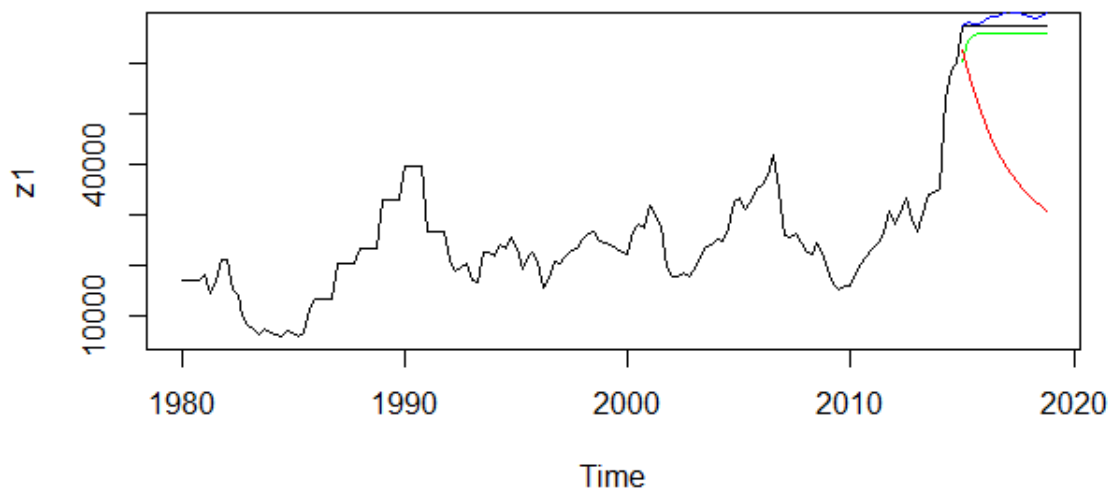
Graphique du Tanker GAZ LNG 00 (série 91) :

Comparaison des modèles pour le bateau GAZ LNG 00



Graphique du TKC PNX 00 (série 97) :

Comparaison des modèles pour le bateau TKC PNX 00



Nous pouvons constater sur ces graphiques que les trois modèles donnent des valeurs relativement proches sauf pour le TKC PNX 00 où le modèle Vasicek prédit une chute des taux. Cependant cela n'est pas suffisant pour conclure, un graphique n'étant pas assez précis. Nous avons donc mis en place une autre méthode.

Calcul des RMSE.

Graphiquement, nous ne pouvons pas constater quel modèle est le plus adapté pour la modélisation des taux de fret ; nous avons donc cherché à départager les différents processus de manière numérique. Pour cela, nous avons décidé de comparer les RMSE des 5 bateaux précédents : le Small Container 00, le BU HDY 00, le Tanker product MR 00, le GAZ LNG 00, et le GAZ LPG 00.

Il existe une fonction sous R ("rmse") avec le package "Metrics" qui permet de faire ce calcul directement :

$$RMSE = rmse(y, \hat{y})$$

avec y les valeurs fournies par VesselsValue et \hat{y} nos prédictions.

Nous avons ensuite comparé les RMSE des différents modèles pour ces navires afin de tenter de déterminer lequel est le plus adapté pour prédire les taux de fret. Nous avons regroupé dans les tableaux ci-dessous ces derniers résultats avec les valeurs des prédictions pour les cinq bateaux sélectionnés précédemment :

RMSE	CO SML 00	BU HDY 00	TKC PNX 00	TKP MR 00	GAZ LNG 00
SARIMA	5508	1059	3191	18149	2306
Vasicek	1412	618	1868	6225	26508
CIR	2009	977	4735	7887	8606

Nous pouvons constater que selon les bateaux, le meilleur modèle varie. En général, c'est Vasicek le modèle qui obtient le meilleur RMSE, c'est à dire le plus faible, sauf pour le GAZ LNG 00, pour lequel vaut mieux utiliser le modèle SARIMA.

Suite à des échanges avec nos professeurs, nous avons réalisé que comparer les prédictions n'étaient peut-être pas le meilleur moyen afin de déterminer le modèle le plus performant. En effet, nous n'avions aucune garantie sur le fait que les prédictions données par VesselsValue étaient correctes.

4.2.2 Confrontation des résultats des modèles.

Afin de déterminer le modèle le plus adapté pour prédire les taux de fret, nous avons décidé de mettre en place une seconde approche.

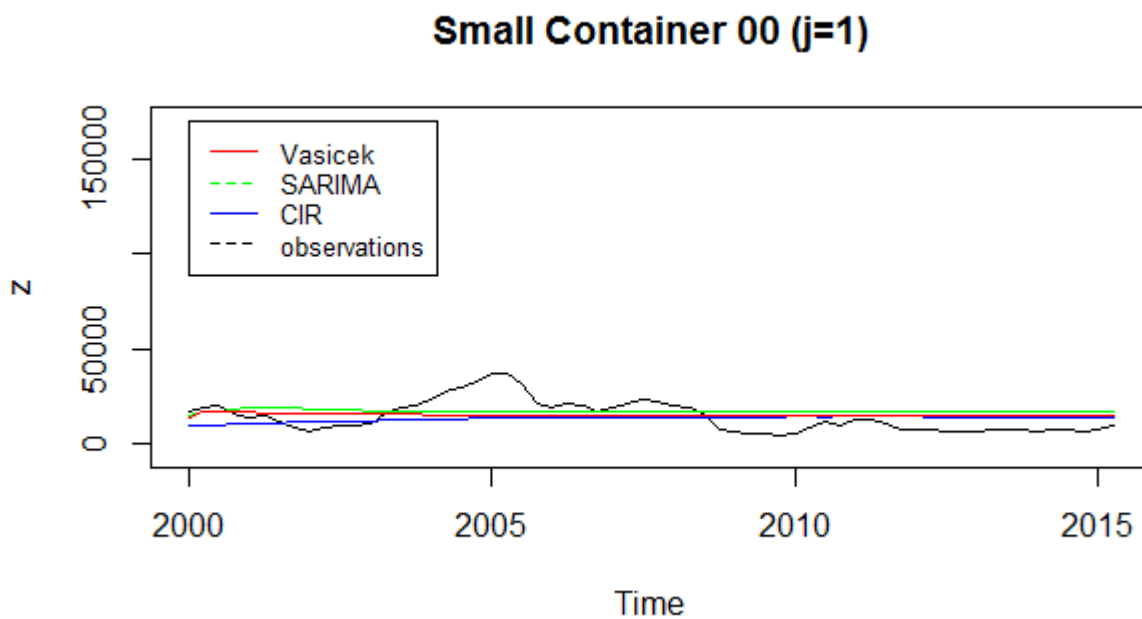
Comme nous disposons des observations des taux de fret jusqu'au second trimestre 2015, nous avons décidé de comparer nos prédictions sur ces périodes avec les vraies valeurs. La comparaison est donc réalisée sur des données connues : les valeurs réalisées de 2000 à 2015.2.

Nous avons donc refait l'étude du paragraphe précédent mais en faisant une comparaison entre les valeurs observées et nos prédictions.

Méthode graphique.

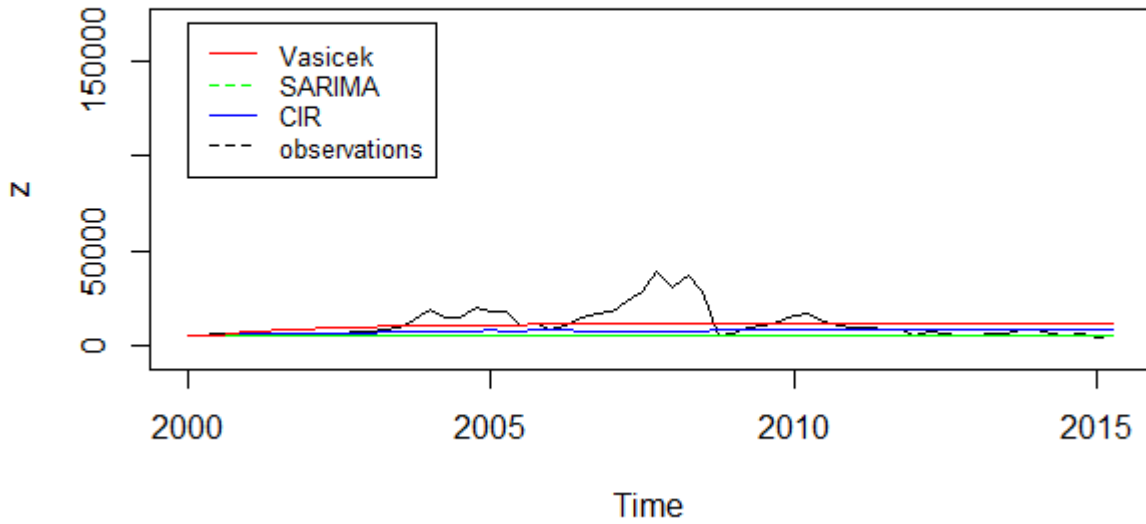
Nous avons d'abord tracé les graphiques des prédictions pour chaque modèle pour chaque type de navires afin, de comparer nos différentes prédictions avec les valeurs données par VesselsValue.

Graphique du Small Container 00 (série 1) :



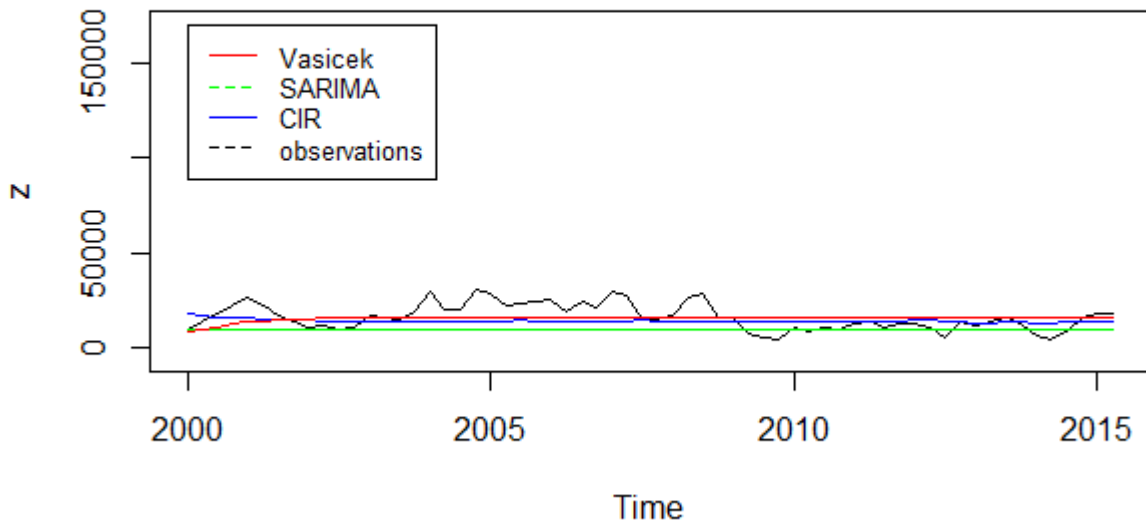
Graphique du BU HDY 00 (série 21) :

BU HDY 00 (j=21)



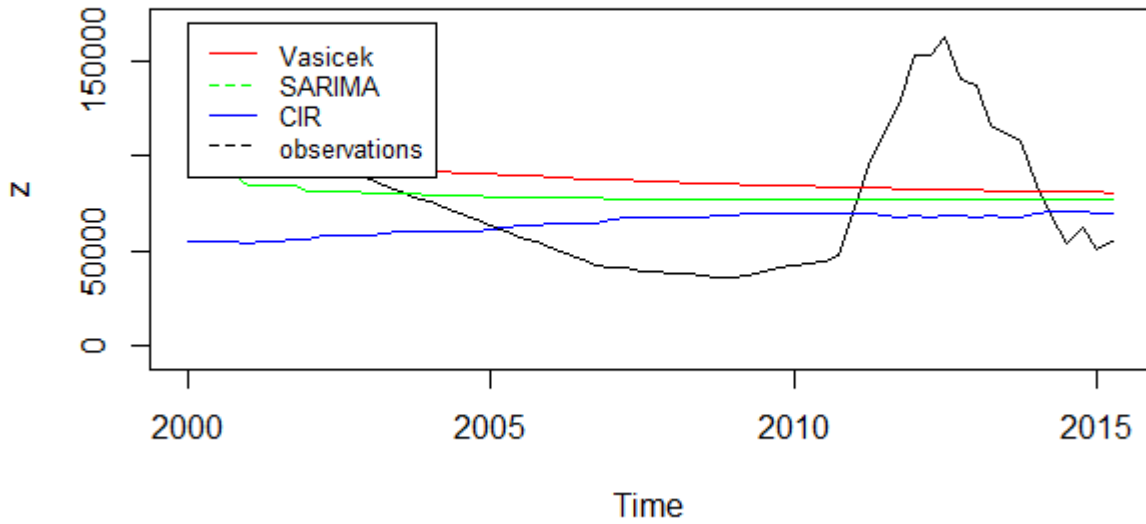
Graphique du Tanker Product MR 00 (série 76) :

Tanker Product MR (j=76)



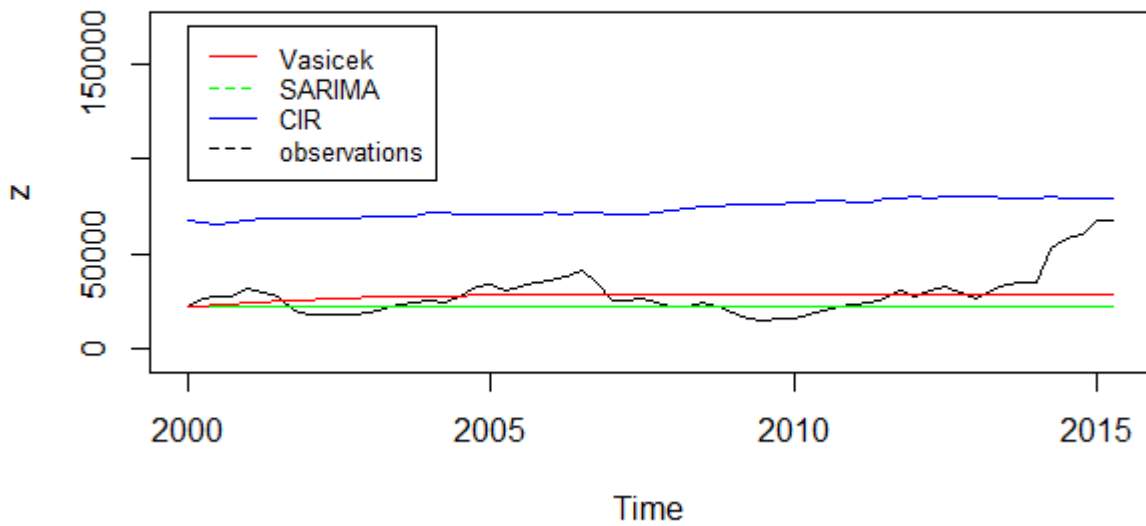
Graphique du Tanker GAZ LNG 00 (série 91) :

Tanker GAZ LNG 00 (j=91)



Graphique du TKC PNX 00 (série 97) :

TKC PNX 00 (j=97)



Nous pouvons constater sur ces graphiques que les trois modèles donnent des valeurs relativement proches des observations même si les modèles ne prédisent pas les sauts des taux.

Par ailleurs, le modèle CIR surestime les variations des taux pour le dernier bateau.

Calcul des RMSE.

Comme nous l'avons vu précédemment, nous ne pouvons pas constater quel modèle est le plus adapté de manière graphique ; nous avons donc appliqué à nouveau la méthode précédente qui consiste à calculer les RMSE des 5 bateaux : le Small Container 00, le BU HDY 00, le Tanker product MR 00, le GAZ LNG 00, et le GAZ LPG 00.

Précédemment, nous avons effectué des comparaisons sur des valeurs prédites (valeurs futures de 2015 à fin 2018), nous pensons qu'il serait plus adéquat de faire une comparaison sur des valeurs connues. Avec les modèles que nous avons implémentés, nous avons fait des prédictions sur des valeurs connues afin de les comparer aux véritables valeurs observées par VesselsValue.

La base de temps choisi pour la comparaison est de 2000 à 2015.2 (arbitraire).

Par ailleurs, le modèle SARIMA a besoin d'un historique, nous allons donc utiliser les données de 1980 à 1999.4 comme base historique afin de faire nos prédictions de 2000 à 2015.2.

Nous avons ensuite comparé les RMSE des différents modèles de navires afin de tenter de déterminer lequel est le plus adapté pour prédire les taux de fret. Nous avons regroupé dans le tableau ci-dessous les valeurs des RMSE obtenues pour les cinq bateaux sélectionnés précédemment :

RMSE	CO SML 00	BU HDY 00	TKC PNX 00	TKP MR 00	GAZ LNG 00
SARIMA	8844	9667	9603	32808	13433
Vasicek	8101	7542	7057	35357	11237
CIR	9013	8896	7343	36538	48088

Comme nous l'avons vu auparavant, c'est le modèle de Vasicek qui semble, en général, le mieux adapté à la modélisation des taux de fret. Cette seconde étude nous permet de voir que pour ces navires c'est le modèle de Vasicek le plus performant. Cependant, nous ne pouvons pas généraliser que ce modèle est le plus adapté pour tous les navires. Nous avons fait des calculs supplémentaires qui nous ont permis de déterminer quel modèle était le plus adapté selon le type de bateau, les résultats sont regroupés sur la page suivante.

A propos du calcul du RMSE, nous pouvons en déduire qu'il n'existe pas un modèle plus performant que les autres mais plutôt qu'ils se complètent. Nous avons donc attribué à chaque type de bateau le modèle qui semble le mieux le modéliser.

Nous avons donc regroupé dans un tableau les modèles qui peuvent être utilisés pour les prévisions concernant les différents navires :

jeu de données composé de 5 familles et de 100 bateaux	
Vasicek	
Bateaux	Bon modèle: résidus ~BB
BULK Handy et BULK HandyMax	Oui
BULK SupraMax	
BULK Panamax et KamsarMax et Cape	
Tankers Crude Panamax et SuezMax	
Tankers Crude VerylargeCont	
SARIMA	
Type de bateaux	Bon modèle: 3 critères
GAZ LNG	Oui
GAZ LPG	
Containers small	
Tankers MR	
C.I.R	
Type de bateaux	Bon modèle:residus ~ BB
Bulk Supramax	Oui
Tankers Crude Suezmax	

Nous pouvons notamment remarquer que tous les bateaux pouvant être modélisés par le CIR le sont aussi par le modèle de Vasicek ce qui est assez logique.

Conclusion

Le but de ce bureau d'étude : "Modélisation stochastique de taux de fret." est de réussir à l'aide des différents instruments mathématiques et informatiques, à prédire des taux de fret pour différents types de bateaux. C'est dans cette optique que nous nous sommes dans un premier temps tournés vers les séries temporelles.

Après avoir transformé nos données en série chronologique, nous avons appliqué la fonction d'auto-corrélation partielle (ACF) sous R sur 2 types de bateaux. Les résultats obtenus par cette fonction étaient semblables à celui d'un modèle AR(2). A l'aide de l'équivalence : $AR(2) \iff ARIMA(2,0,0)$, nous avons implémenté ce modèle sur nos données et nous avons effectué le test de Box-Pierce. Ce test évalue la corrélation existant entre les résidus. Il teste l'hypothèse : "Est-ce que les résidus sont de bruit blanc fort ?", et donne une p-value. Nous avons vu précédemment que cette p-value était inférieure à 5%, nous avons dû rejeter cette hypothèse. Le modèle AR(2) n'est donc pas adapté.

Le fait que nos données présentaient une certaine saisonnalité nous a incités à nous tourner vers les séries SARIMA. Après avoir fait des prédictions sur tous les navires à l'aide de ces dernières, nous avons créé des tests afin de vérifier que nos estimations étaient correctes et de voir quels navires pouvaient être modélisés par ces séries. Lorsque nous avons tracé les graphiques, nous avons pu voir qu'il y avait bien une saisonnalité qui apparaissait.

Voici les bateaux qui peuvent donc être modélisés par ce modèle :

SARIMA	
Type de bateaux	Bon modèle: 3 critères
GAZ LNG	Oui
GAZ LPG	
Containers small	
Tankers MR	

La seconde méthode étudiée est le modèle de Vasicek car c'est un modèle courant dans la modélisation de taux d'intérêts. Dans l'étude de ce modèle, nous avons d'abord présenté celui ci et démontré qu'il est comparable à un modèle AR(1), ce qui nous a permis de faciliter l'implémentation sous R. Ensuite, nous avons calculé nos prédictions à l'aide de l'équivalence entre le modèle AR(1) et Vasicek. Après avoir effectué des tests sur les résidus, nous avons dégagé une liste de bateaux modélisables par le modèle de Vasicek. Enfin, nous avons effectué le calcul de l'erreur quadratique moyenne pour mesurer l'écart entre nos prédictions et celles données par VesselsValue.

Nous obtenons donc que les bateaux suivants peuvent être modélisés par le modèle de Vasicek :

Vasicek	
Bateaux	Bon modèle: résidus ~BB
BULK Handy et BULK HandyMax	Oui
BULK SupraMax	
BULK Panamax et KamsarMax et Cape	
Tankers Crude Panamax et SuezMax	
Tankers Crude VerylargeCont	

Ensuite, nous nous sommes intéressés au modèle de Cox Ingersoll Ross qui est aussi un modèle classique de description des taux. D'autre part, celui-ci interdit les taux négatifs ce qui est un avantage pour le calcul de nos taux de fret. Afin de réaliser nos prédictions nous sommes passés par les étapes de discrétisation de l'intégrale stochastique, puis nous avons dû estimer les différents paramètres qui composent ce processus. Nous avons effectué une régression linéaire sur l'équation discrétisée, pour déterminer deux des trois paramètres du modèle : α , et μ . Pour déterminer la dernière inconnue, σ , nous avons utilisé une méthode calculatoire. Ensuite nous avons appliqué le log du max de vraisemblance sur le vecteur $\theta = (\alpha, \sigma, \mu)$ que nous avons optimisé par la suite. Malheureusement, l'algorithme d'optimisation n'était pas adéquat, nous avons donc voulu tester si nos fonctions de calcul des paramètres étaient correctes. Nous simulions des trajectoires en prenant une discrétisation plus fine avec en entrée des paramètres que nous avons fixés et nous vérifions si nos fonctions retrouvaient bien les mêmes paramètres.

Le modèle CIR nous donne donc les modélisations des navires suivants :

C.I.R	
Type de bateaux	Bon modèle:residus ~ BB
Bulk Supramax	Oui
Tankers Crude Suezmax	

Dans la dernière partie de ce rapport, nous faisons un comparatif entre les méthodes présentées auparavant. Pour cela, nous opposons d'abord graphiquement les résultats pour un bateau de chaque famille, soit les bateaux suivants : le Small Container 00, le Tanker product MR 00, le GAZ LNG 00, le GAZ LPG 00, et le TKC PNX 00. Puis, dans un souci de rigueur, nous utilisons l'erreur quadratique moyenne pour mettre en balance les résultats pour les prédictions des trois modèles afin de tenter de dégager quel processus est le plus performant. Le résultat de notre étude est qu'il n'y a pas un modèle qui surpasse les autres mais plutôt que ces modèles se complètent dans la modélisation des taux fret : certains processus donnent de meilleures prédictions pour un type de bateau donné.

Pour conclure voici comment pourrait être partagée la modélisation des différents navires par modèle :

jeu de données composé de 5 familles et de 100 bateaux	
Vasicek	
Bateaux	Bon modèle: résidus ~BB
BULK Handy et BULK HandyMax	Oui
BULK SupraMax	
BULK Panamax et KamsarMax et Cape	
Tankers Crude Panamax et SuezMax	
Tankers Crude VerylargeCont	
SARIMA	
Type de bateaux	Bon modèle: 3 critères
GAZ LNG	Oui
GAZ LPG	
Containers small	
Tankers MR	
C.I.R	
Type de bateaux	Bon modèle:residus ~ BB
Bulk Supramax	Oui
Tankers Crude Suezmax	

Concernant le modèle de Cox ingersoll Ross, il serait intéressant de se pencher sur la simulation des incréments du processus sans l'utilisation des approximations. En effet, ces incréments suivent une loi du χ_2 décentrée qui peut être simulée exactement. Néanmoins, la simulation exacte est bien plus coûteuse en temps de calcul et les schémas de discrétisation s'avèrent être plus efficaces pour générer des trajectoires. Par ailleurs, la convergence avec le schéma de discrétisation de Milstein est bien meilleure que celle du schéma d'Euler utilisée dans notre étude.

Bibliographie.

Livres :

- ALIZADED Amir, NOMIKOS Nikos. "Shippin Derivatives and Risk Management".Palgrave Macmillan. Edition 2009.
- KARAKITSOS Elias, VARNAVIDES Lambros. "Maritime Economics" : A macroeconomic approach. Palgrave Macmillan. Edition 2014.
- STOPFORD Martin. "Maritime Economics". Routledge. Edition 2009.
- ALIZADED Amir, NOMIKOS Nikos. "Dynamics of the Term Structure and Volatility of Shipping Freight Rates" Edition 2011

Mémoires :

- SMAILI Amine, RGUIBI El Mehdi, WANG Chengming. Bureau d'étude EURIA : "Conception de scénarios de Stress Tests" (2013).
- MAGIROU Envangelos, PSARAFTIS Harilaos, CHRISTODOULAKIS Nikoloas. " Quantitative methods in shipping :a survey of current use and future trends. "(1992).
- KOEKEBAKKER Steen, ADLAND Roar. "Value based trading of real assets in shipping under stochastic freight rates."(2006).
- ADLAND ROAR. "The stochastic behavior of spot freight rates and the risk premium in bulk shipping"(1972).
- KAVUSSANOS Manolis, ALIZADEH Amir. "Seasonality patterns in dry bulk shipping spot and time charter freight rates."(2001).
- AMIN H.H.N. "Calibration of Different Interest Rate Models for a Good Fit of Yield Curves"(2012).

Documents divers :

- BARTLETT Roger. "The Valuation of Ships Art and Science" (présentation du 1er Novembre 2012 au sixième forum sur la finance maritime coréenne à Busan).
- ADAMOU Alex. "Behind the Screens : the Mathematics of Vessel Valuation" (document interne de VesselsValue 06 Février 2014).
- KARATZAS Basil. "Small details that make a big difference for FMV in shipping" (article internet provenant de "bmkaratzas.wordpress.com" du 05 Septembre 2013).

Annexes.

Codes R.

Code pour les séries temporelles.

```
library(forecast)
y=ts(taux2[,t],start=c(1980,1),end=c(2015,2),frequency=4)
y1=ts(taux2[,t],start=c(1980,1),end=c(2014,4),frequency=4)
a1=auto.arima(y1)
z=forecast(a1,h=2)
plot(z)
points(y)
title(sub =paste("Prédictions du 1re &
2ème trimestre de 2015 sur le ", names(taux2)[t]),
      xlab = "Année",ylab = "Value de Bateau ",col.sub="blue")

y=ts(taux2[,t],start=c(1980,1),end=c(2018,4),frequency=4)
y1=ts(taux2[,t],start=c(1980,1),end=c(2015,2),frequency=4)
a1=auto.arima(y1)
z=forecast(a1,h=14)
plot(z)
points(y)
title(sub =paste("Les deux prédictions
de 3.2015 - 4.2018 sur le ", names(taux2)[t]),
      xlab = "Année",ylab = "Valeur du bateau ",col.sub="blue")
return(z)}
```

Code pour le modèle CIR.

```
# simulation de la trajectoire d'un CIR partant de x, de paramètres a,b,s
CIR1=function(T,N,a,b,s,x)
{ h=T/N
X=rep(x,N)
for (i in 1:(N-1))
{y=X[i]
X[i+1]=y+a*(b-max(y,0))*h+s*sqrt(h*max(y,0))*rnorm(1)}
return(X)}
```

```
Trajectoire_CIR=function(s){
# calibration CIR avec méthode regression
```



```

taux2=read.csv2("D:\\Users\\Proprietaire\\Documents\\Euria\\BE\\codes\\tauxdefret2.

#s=29
#y=ts(taux2[,j],start=c(1980,1),end=c(2014,4),frequency=4)
y=ts(taux2[,s],start=c(1980,1),end=c(2015,2),frequency=4)

n=length(y)
z=(y[2:n]-y[1:(n-1)])/sqrt(y[1:(n-1)])
x1=1/sqrt(y[1:(n-1)])
x2=-sqrt(y[1:(n-1)])

reg=lm(z~0+x1+x2)
c=reg$coefficients
ae=c[2]
be=c[1]/ae
se=sd(reg$residuals)
d=c(ae,be,se)

N=1e4
T=20
x0=y[n]

# simulation d'une trajectoire d'un CIR avec les parametres estimes
r=CIR1(T,N,ae,be,se,x0)
K=seq(1,N, by=N/T)
r=r[K]
#plot(1:T,r,type='l')

# simulation de M trajectoires
M=100
R=matrix(rep(0,M*T),ncol=T)
for (i in 1:M)
{r=CIR1(T,N,ae,be,se,x0)
R[i,]=r[K]}
#R

# A partir des M trajectoires, on determine un intervalle de confiance empirique
# instant futur) et une prevision (moyenne empirique des M trajectoires)
mini=rep(0,T)

```

```

maxi=rep(0,T)
med=rep(0,T)
moy=rep(0,T)
for (j in 1:T)
{a=sort(R[,j])
mini[j]=a[M*0.1]
maxi[j]=a[M*0.9]
med[j]=a[M/2] # la médiane
moy[j]=mean(a)} # la moyenne

plot(mini,xlab = "Années",ylab = "taux de fret ",col="blue",type='l',main="Prédicti
lines(maxi,col='blue')
#lines(med,col='green')
lines(moy,col='green')
lines(taux2[142:156,s],col='black')
#title(sub =paste("Prédictions à partir de 2015,3 à 2020,1 ", names(taux2)[t]),
# xlab = "Année",ylab = "Value de Bateau ",col.sub="blue")
#legend(8,80000,legend=c("Données connues","Prédictions","Intervalle de confiance")
return(moy)
}
Trajectoire_CIR(96)
library(Metrics)
j=96
write.table(data.frame(tableau), "tableau.csv",sep=";")
tableau=cbind(Trajectoire_CIR(1),Trajectoire_CIR(76),Trajectoire_CIR(91),Trajectoire_C
#tableau=cbind.data.frame(Trajectoire_CIR(21),Trajectoire_CIR(71),Trajectoire_CIR(70)
#RMSE_cir=rmse(tableau[,1],taux2[137:156,j])
#RMSE_cir

length(tableau[,1])
length(taux2[141:156,j])
# pour avoir le nom des bateaux
names(taux2)[71]
names(taux2)

```

Code pour la simulation des trajectoires du CIR.

```

set.seed(400821)

##### Simulation de la trajectoire du CIR#####

```

```

CIR_simulation=function(T,N,a,b,s,x){
  # simulation d'un CIR
  # T : temps final
  # N : nombre de points de discrétisation
  h=T/N # pas de discrétisation
  X=rep(x,N)
  for (i in 1:(N-1)){
    y=X[i]
    X[i+1]=y+a*(b-max(y,0))*h+s*sqrt(h*max(y,0))*rnorm(1)
  }
  return(X)
}

```

fin simulation de la trajectoire du CIR

```

parametre=function(simul,delta){

  #calibration de la série par régression
  pour retrouver les paramètres de  $\theta$ 
  n=length(simul)
  y=simul
  z=(y[2:n]-y[1:(n-1)])/sqrt(y[1:(n-1)])
  x1=delta/sqrt(y[1:(n-1)])
  x2=-sqrt(y[1:(n-1)])*delta
  fit=lm(z~0+x1+x2)

  #récupération des coefficients du C.I.R
  vecteur=fit$coefficients
  thetae=c(vecteur[2],vecteur[1]/vecteur[2],sd(fit$residuals))
  #theta=c(estimateur de a, estim de b, estim de sigma)
  return(thetae)
}

```

```

log_vraisembl=function(T,N,delta,simul,thetae)
{

```

```

u<-c();v<-c();q<-c();
som=0;

#### Extraction d'une trajectoire de pas delta ####
n=T/delta
  #nombre d'observations de la trajectoire extraite
t=seq(1,N,by=N/n)
  #indices pour ces observations dans simul
xe=simul[t]
  #extraction des n obs ie discrétisation du CIR
temps=seq(delta,T,by=delta) #instants d'observation
#plot(temps,xe,type='l')

##### Max du log de vraisemblance #####
long=length(xe)

#construction de u v c et q
#theta=(estimateur de a, estim de b, estim de sigma)
c=(2*thetae[1])/((1-exp(-thetae[1]*delta))*thetae[3]^2)
q=((2*thetae[1]*thetae[2])/(thetae[3]^2))-1
for(i in 1: long){
  u[i]=c*xe[i]*exp(-xe[i]*delta)
  v[i]=c*xe[i+1]
}
u=u[1:(long-1)]
v=v[1:(long-1)]

som=(long-1)*log(c)+sum(-u-v+0.5*q*log(v/u)
+log(besselI(2*sqrt(u*v),q,expon.scaled=TRUE))
+2*sqrt(u*v))
  #som represente le calcul du log du max de vraisemblance
return(-som)
}

voisinage=function(thetae,simul)
{
  lva=c();lvb=c();lvs=c()
  a=thetae[1];b=thetae[2];s=thetae[3];

```

```

eps=seq(-0.1,0.1,by=0.01)
eps2=seq(-0.1,0.2,by=0.01)

for (ep in eps)
{
  #voisinage de a
  ya=c(a+ep,b,s)
  lva=c(lva,log_vraisembl(T,N,delta,simul,ya))

  #####voisinage de b #####
  yb=c(a,b+ep,s)
  lvb=c(lvb,log_vraisembl(T,N,delta,simul,yb))
}

for (ep in eps2)
{
  #voisinage de s
  ys=c(a,b,s*(1+ep))
  lvs=c(lvs,log_vraisembl(T,N,delta,simul,ys))
}

plot(eps,lva,main='voisinage de a')
plot(eps,lvb,main='voisinage de b')
plot(eps2,lvs,main='voisinage de s')

}

#x=c(ae,be,se)

#test du programme
N=1e4
T=20
a=0.3
b=0.1
s=0.01
x0=0.02
delta=0.1
theta=c(a,b,s)

```

```

#execution
simul=CIR_simulation(T,N,a,b,s,x0)
thetae=parametre(simul,delta) # donne les estimateurs du vecteur thetae
log_max=log_vraisembl(T,N,delta,simul,thetae)

#graphe et voisinage
voisinage(theta,simul)

vec_theta=theta
log_vraisembl(T,N,delta,simul,theta)
op1=optim(par=vec_theta,log_vraisembl(T,N,delta,simul,vec_theta))
vec_theta #valeur initial
op$var

vec_theta=theta*(1+0.05)
log_vraisembl(T,N,delta,simul,vec_theta)
op2=optim(par=vec_theta,log_vraisembl(T,N,delta,simul,vec_theta))
vec_theta #valeur initiale *(1+0.05)
op$var

vec_theta=thetae
log_vraisembl(T,N,delta,simul,vec_theta)
op3=optim(par=vec_theta,log_vraisembl(T,N,delta,simul,vec_theta))
vec_theta # estimation
op$var

```

Code pour le modèle de Vasicek.

```

x=ts(taux2[,j],start=c(1980,1),end=c(2015,2),frequency=4)
mu=mean(x) # estimation de mu
xc=x-mu # series des taux, centrée
a1=ar(xc,order.max=1)
predi=predict(a1, n.ahead = 16, se.fit = TRUE)

```

Code du calcul du RMSE pour tous les modèles et tous les bateaux.

```

#Vasicek
library(Metrics)
RMSE_Vas=NULL
for(j in 1:100){
  x=ts(taux2[,j],start=c(1980,1),end=c(2015,2),frequency=4)
  mu=mean(x) # estimation de mu
  xc=x-mu # series des taux, centrée
  a1=ar(xc,order.max=1)
  predi=predict(a1, n.ahead = 16, se.fit = TRUE)
  RMSE_Vas=c(RMSE_Vas,rmse(predi$pred+mu,taux2[141:156,j]))
  #calcul du RMSE pour Vasicek
}

#ARIMA
n=100
library(forecast)
prediction=matrix(0,nrow=16,ncol=n)
for (j in 1:n)
{y=ts(taux2[,j],start=c(1980,1),end=c(2015,2),frequency=4)
y1=ts(taux2[,j],start=c(1980,1),end=c(2014,4),frequency=4)

a1=auto.arima(y1)
z=forecast(a1,h=16)
pred=z$mean # predictions ponctuelles
prediction[,j]=pred
}
RMSE_ARIMA=NULL
for(j in 1:100){
  RMSE_ARIMA=c(RMSE_ARIMA,rmse(prediction[,j],taux2[141:156,j]))
  #calcul du RMSE pour ARIMA
}

```

Tableau récapitulatif du calcul des erreurs pour tous les bateaux.

Nom	Erreur1	Erreur2	ErreurMax	ErreurTend	Modèle	Colonne1
CO_SML_00	0,013	0,135	1	1	ARIMA(1,1,0)(1,0,2)[4]	
CO_SML_05	0,013	0,135	1	1	ARIMA(1,1,0)(1,0,2)[4]	
CO_SML_10	0,013	0,135	1	1	ARIMA(1,1,0)(1,0,2)[4]	
CO_SML_15	0,013	0,135	1	1	ARIMA(1,1,0)(1,0,2)[4]	
CO_SML_20	0,013	0,135	1	1	ARIMA(1,1,0)(1,0,2)[4]	
CO_MED_00	-0,119	-0,194	0	1	ARIMA(2,0,0)(1,0,1)[4] with non-zero mean	
CO_MED_05	-0,119	-0,194	0	1	ARIMA(2,0,0)(1,0,1)[4] with non-zero mean	
CO_MED_10	-0,119	-0,194	0	1	ARIMA(2,0,0)(1,0,1)[4] with non-zero mean	
CO_MED_15	-0,119	-0,194	0	1	ARIMA(2,0,0)(1,0,1)[4] with non-zero mean	
CO_MED_20	-0,119	-0,194	0	1	ARIMA(2,0,0)(1,0,1)[4] with non-zero mean	
CO_LGE_00	0,198	0,101	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_LGE_05	0,198	0,101	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_LGE_10	0,198	0,101	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_LGE_15	0,198	0,101	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_LGE_20	0,198	0,101	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_VLGE_00	0,228	0,143	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_VLGE_05	0,228	0,143	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_VLGE_10	0,228	0,143	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_VLGE_15	0,228	0,143	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
CO_VLGE_20	0,228	0,143	0	0	ARIMA(2,0,0)(1,0,2)[4] with non-zero mean	
BU_HDY_00	-0,354	-0,285	0	0	ARIMA(0,1,0)(1,0,0)[4]	
BU_HDY_05	-0,354	-0,285	0	0	ARIMA(0,1,0)(1,0,0)[4]	
BU_HDY_10	-0,354	-0,285	0	0	ARIMA(0,1,0)(1,0,0)[4]	
BU_HDY_15	-0,354	-0,285	0	0	ARIMA(0,1,0)(1,0,0)[4]	
BU_HDY_20	-0,354	-0,285	0	0	ARIMA(0,1,0)(1,0,0)[4]	
BU_HDYM_00	-0,578	-0,493	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_HDYM_05	-0,578	-0,493	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_HDYM_10	-0,578	-0,493	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_HDYM_15	-0,578	-0,493	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_HDYM_20	-0,578	-0,493	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_SUP_00	-0,609	-0,535	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_SUP_05	-0,609	-0,535	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_SUP_10	-0,609	-0,535	0	0	ARIMA(0,1,0)(0,0,1)[4]	
BU_SUP_15	-0,609	-0,535	0	0	ARIMA(0,1,0)(0,0,1)[4]	

BU_SUP_20	-0,609	-0,535	0	0 ARIMA(0,1,0)(0,0,1)[4]
BU_PAN_00	-0,593	-0,665	0	1 ARIMA(2,1,2)(0,0,2)[4]
BU_PAN_05	-0,593	-0,665	0	1 ARIMA(2,1,2)(0,0,2)[4]
BU_PAN_10	-0,593	-0,665	0	1 ARIMA(2,1,2)(0,0,2)[4]
BU_PAN_15	-0,593	-0,665	0	1 ARIMA(2,1,2)(0,0,2)[4]
BU_PAN_20	-0,593	-0,665	0	1 ARIMA(2,1,2)(0,0,2)[4]
BU_KAM_00	-0,823	-0,678	0	1 ARIMA(0,1,0) with drift
BU_KAM_05	-0,823	-0,678	0	1 ARIMA(0,1,0) with drift
BU_KAM_10	-0,823	-0,678	0	1 ARIMA(0,1,0) with drift
BU_KAM_15	-0,823	-0,678	0	1 ARIMA(0,1,0) with drift
BU_KAM_20	-0,823	-0,678	0	1 ARIMA(0,1,0) with drift
BU_PPAN_00	-1,032	-0,87	0	1 ARIMA(0,1,0) with drift
BU_PPAN_05	-1,032	-0,87	0	1 ARIMA(0,1,0) with drift
BU_PPAN_10	-1,032	-0,87	0	1 ARIMA(0,1,0) with drift
BU_PPAN_15	-1,032	-0,87	0	1 ARIMA(0,1,0) with drift
BU_PPAN_20	-1,032	-0,87	0	1 ARIMA(0,1,0) with drift
BU_CAP_00	-2,029	-1,773	0	1 ARIMA(0,1,0) with drift
BU_CAP_05	-2,029	-1,773	0	1 ARIMA(0,1,0) with drift
BU_CAP_10	-2,029	-1,773	0	1 ARIMA(0,1,0) with drift
BU_CAP_15	-2,029	-1,773	0	1 ARIMA(0,1,0) with drift
BU_CAP_20	-2,029	-1,773	0	1 ARIMA(0,1,0) with drift
TKC_PNX_00	0,328	0,405	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_PNX_05	0,328	0,405	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_PNX_10	0,328	0,405	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_PNX_15	0,328	0,405	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_PNX_20	0,328	0,405	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_AMX_00	0,365	0,417	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_AMX_05	0,365	0,417	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_AMX_10	0,365	0,417	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_AMX_15	0,365	0,417	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_AMX_20	0,365	0,417	0	0 ARIMA(0,1,2)(0,0,1)[4]
TKC_SZX_00	0,299	0,255	0	0 ARIMA(0,1,3)(0,0,1)[4]
TKC_SZX_05	0,299	0,255	0	0 ARIMA(0,1,3)(0,0,1)[4]
TKC_SZX_10	0,299	0,255	0	0 ARIMA(0,1,3)(0,0,1)[4]
TKC_SZX_15	0,299	0,255	0	0 ARIMA(0,1,3)(0,0,1)[4]

TKC_SZX_20	0,299	0,255	0	0	ARIMA(0,1,3)(0,0,1)[4]
TKC_VLCC_04	0,297	0,419	0	0	ARIMA(2,1,3)(0,0,1)[4]
TKC_VLCC_04	0,297	0,419	0	0	ARIMA(2,1,3)(0,0,1)[4]
TKC_VLCC_14	0,297	0,419	0	0	ARIMA(2,1,3)(0,0,1)[4]
TKC_VLCC_14	0,297	0,419	0	0	ARIMA(2,1,3)(0,0,1)[4]
TKC_VLCC_24	0,297	0,419	0	0	ARIMA(2,1,3)(0,0,1)[4]
TKP_MR_00	0,108	0,104	1	0	ARIMA(0,1,0) with drift
TKP_MR_05	0,108	0,104	1	0	ARIMA(0,1,0) with drift
TKP_MR_10	0,108	0,104	1	0	ARIMA(0,1,0) with drift
TKP_MR_15	0,108	0,104	1	0	ARIMA(0,1,0) with drift
TKP_MR_20	0,108	0,104	1	0	ARIMA(0,1,0) with drift
TKP_LR1_00	0,125	0,265	0	0	ARIMA(2,1,3)
TKP_LR1_05	0,125	0,265	0	0	ARIMA(2,1,3)
TKP_LR1_10	0,125	0,265	0	0	ARIMA(2,1,3)
TKP_LR1_15	0,125	0,265	0	0	ARIMA(2,1,3)
TKP_LR1_20	0,125	0,265	0	0	ARIMA(2,1,3)
TKP_LR2_00	0,364	0,403	0	0	ARIMA(0,1,2)(0,0,2)[4]
TKP_LR2_05	0,364	0,403	0	0	ARIMA(0,1,2)(0,0,2)[4]
TKP_LR2_10	0,364	0,403	0	0	ARIMA(0,1,2)(0,0,2)[4]
TKP_LR2_15	0,364	0,403	0	0	ARIMA(0,1,2)(0,0,2)[4]
TKP_LR2_20	0,364	0,403	0	0	ARIMA(0,1,2)(0,0,2)[4]
GAS_LNG_00	-0,104	0,014	1	1	ARIMA(2,1,0)(1,0,0)[4]
GAS_LNG_05	-0,104	0,014	1	1	ARIMA(2,1,0)(1,0,0)[4]
GAS_LNG_10	-0,104	0,014	1	1	ARIMA(2,1,0)(1,0,0)[4]
GAS_LNG_15	-0,104	0,014	1	1	ARIMA(2,1,0)(1,0,0)[4]
GAS_LNG_20	-0,104	0,014	1	1	ARIMA(2,1,0)(1,0,0)[4]
GAS_LPG_00	0,102	0,043	1	0	ARIMA(0,1,1)(0,0,1)[4]
GAS_LPG_05	0,102	0,043	1	0	ARIMA(0,1,1)(0,0,1)[4]
GAS_LPG_10	0,102	0,043	1	0	ARIMA(0,1,1)(0,0,1)[4]
GAS_LPG_15	0,102	0,043	1	0	ARIMA(0,1,1)(0,0,1)[4]
GAS_LPG_20	0,102	0,043	1	0	ARIMA(0,1,1)(0,0,1)[4]