

SÉRIES CHRONOLOGIQUES

PROJET

27 février 2017

Axelle Albot et Kilian Fatras



Table des matières

1	Introduction	3
2	Traitement des données et statistiques de bases	4
2.1	Traitement des données	4
2.2	Statistiques de bases	5
3	Étude des données brutes	6
3.1	Étude de la tendance	6
3.2	Étude la saisonnalité	11
4	Lissage exponentiel et prévisions	13
5	Conclusion	19

1 INTRODUCTION

Dans le cadre de la 2^{ème} année du cycle ingénieur de l'ENSTA ParisTech et du cours MAP STA 2, nous étudions plus précisément les séries chronologiques. Nous avons donc un projet sur des données libres que nous devons étudier afin d'appliquer les concepts théoriques et appliqués étudiés en cours.

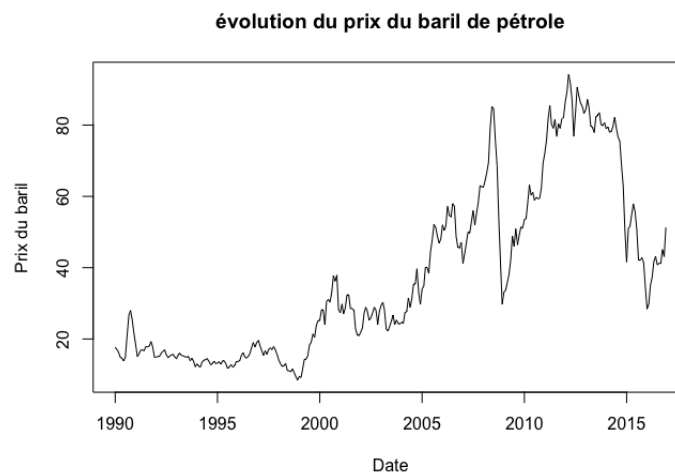
Nous avons choisi le cours du pétrole brut brent (Londres) en euros et par baril. Ce choix est justifié par les récentes conjectures internationales avec le ralentissement de l'économie chinoise qui a baissé la demande, ce qui avait entraîné une chute du prix, puis le choix des pays de l'OPEP de réduire leur production afin de diminuer l'offre et de faire remonter le cours. Nous avons trouvé les données sur le site de l'INSEE.

Dans un premier temps, nous étudierons les différentes statistiques essentielles de la série. Puis nous estimerons sa tendance, sa saisonnalité et sa stationnarité. Enfin nous essaierons d'estimer ses futures valeurs via les techniques de lissages.

Concernant les 3 codes R fournis, le premier à lire est le fichier `cours_petrole.R`, le second `cours_pretrol_linear_smooth.R` (il concerne l'étude l'étude des données modifiées) et le dernier concerne `comparaison_lisse_brut.R` permet de comparer deux prévisions effectuées à partir des données réelles et des données lissées.

2 TRAITEMENT DES DONNÉES ET STATISTIQUES DE BASES

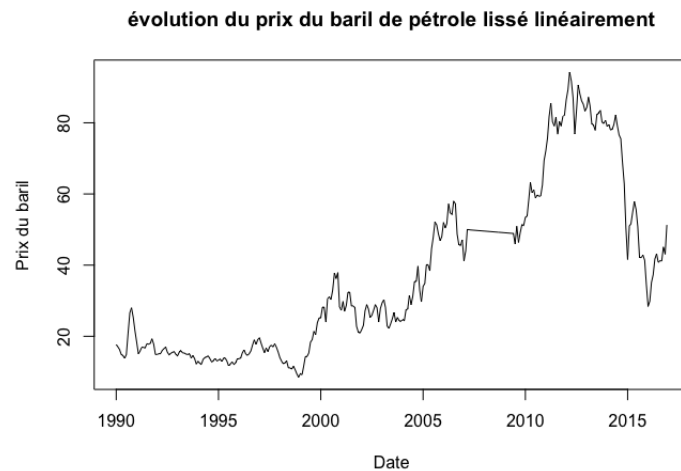
Les données sont des prix mensuelles fournies via un fichier CSV. Elles commencent en décembre 2016 pour finir en janvier 1990. Elles sont donc fournies dans l'ordre décroissantes. Nous avons inclus ces données de manière à les remettre dans l'ordre. Lorsque nous affichons le prix en fonction de la date nous obtenons :



2.1 Traitement des données

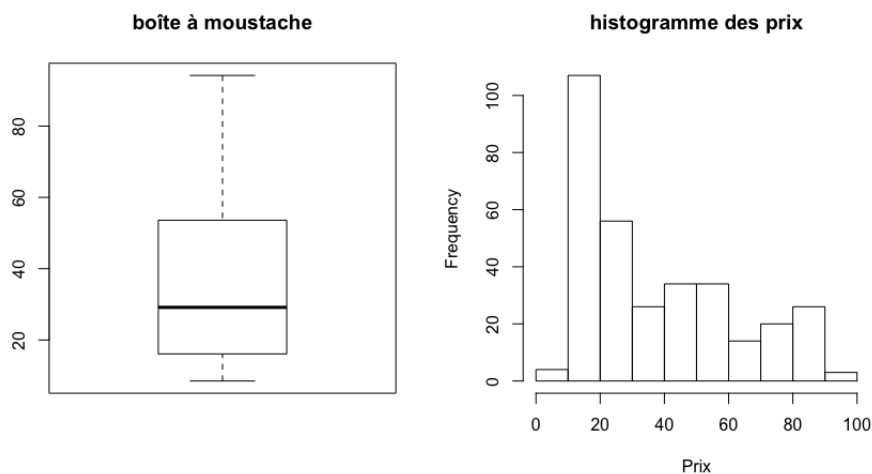
Nous voyons donc un choc lors de l'année 2008. Il est dû à la crise des subprimes. Ce choc est un élément important lors du traitement des données, c'est pourquoi nous étudierons 2 jeux de données différentes, le premier est le jeu brut, le second est un lissage du choc de 2008 par une approche linéaire. Ce second jeu sera étudié dans la partie lissage exponentiel et prévisions.

L'approche linéaire est justifiée par l'étude de la saisonnalité, en effet on observe que les données sont croissantes puis décroissantes, or avant le choc de 2008, on a bien une partie croissante, nous lisons donc de manière décroissante.



2.2 Statistiques de bases

Concernant les statistiques de bases, nous avons une moyenne de 37,86€, une valeur minimale de 8,5€, une valeur max de 94,20€, une médiane à 29,15€, le premier quartile à 16,15€ et le troisième quartile à 53,50€. Nous avons par ailleurs la boîte à moustache et l'histogramme suivant :



3 ÉTUDE DES DONNÉES BRUTES

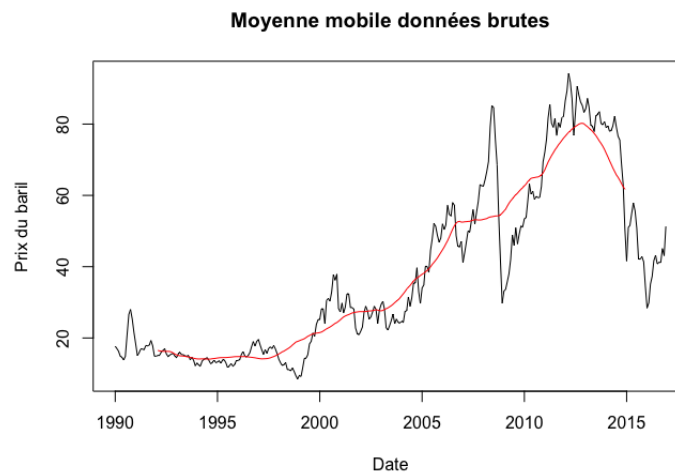
3.1 Étude de la tendance

Plusieurs modélisations possibles de la tendance :

Pour étudier la tendance, nous avons plusieurs outils.

Moyenne mobile :

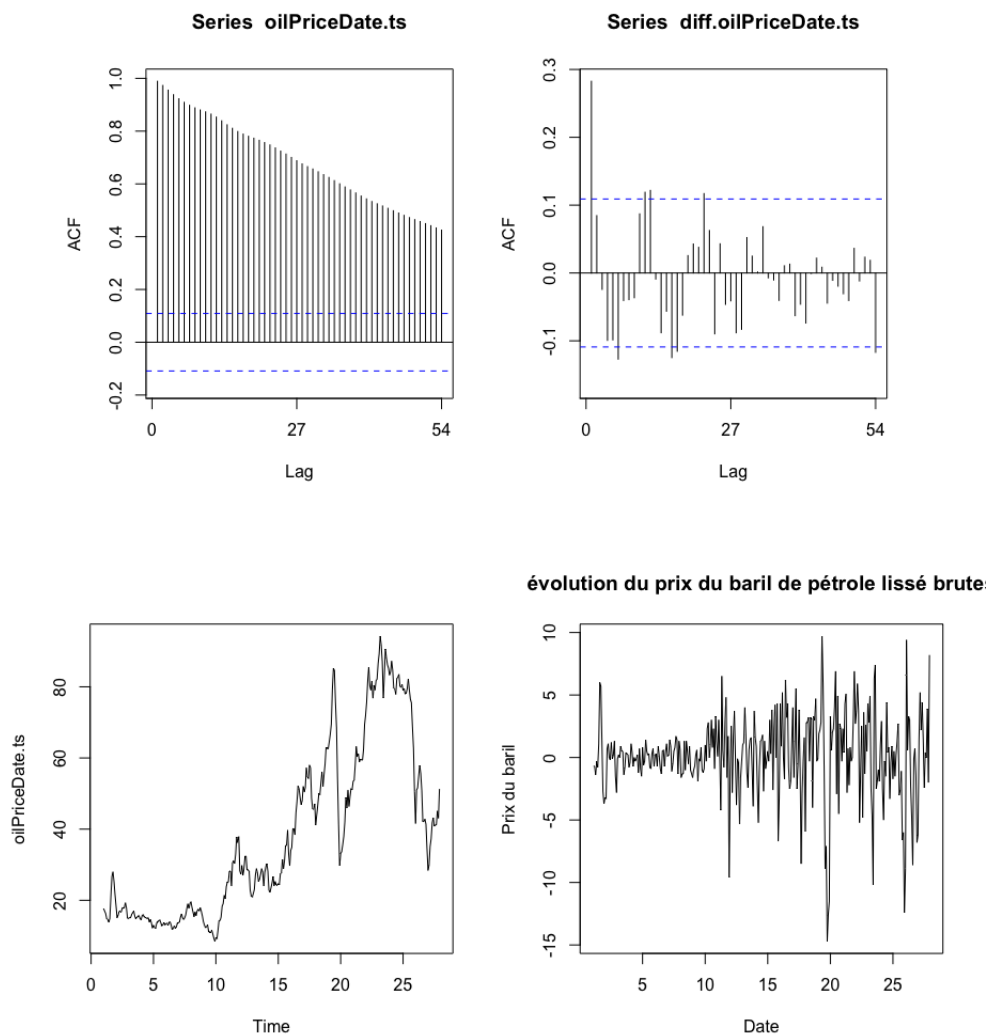
Nous commencerons par utiliser la moyenne mobile, c'est à dire regarder une moyenne locale entre un nombre de données fixés.



Ici nous avons choisit une moyenne sur 50 mois, soit sur un peu plus de 4 ans. Nous observons que la tendance est globalement croissante avec une légère stagnation après chaque augmentation du cours du pétrole.

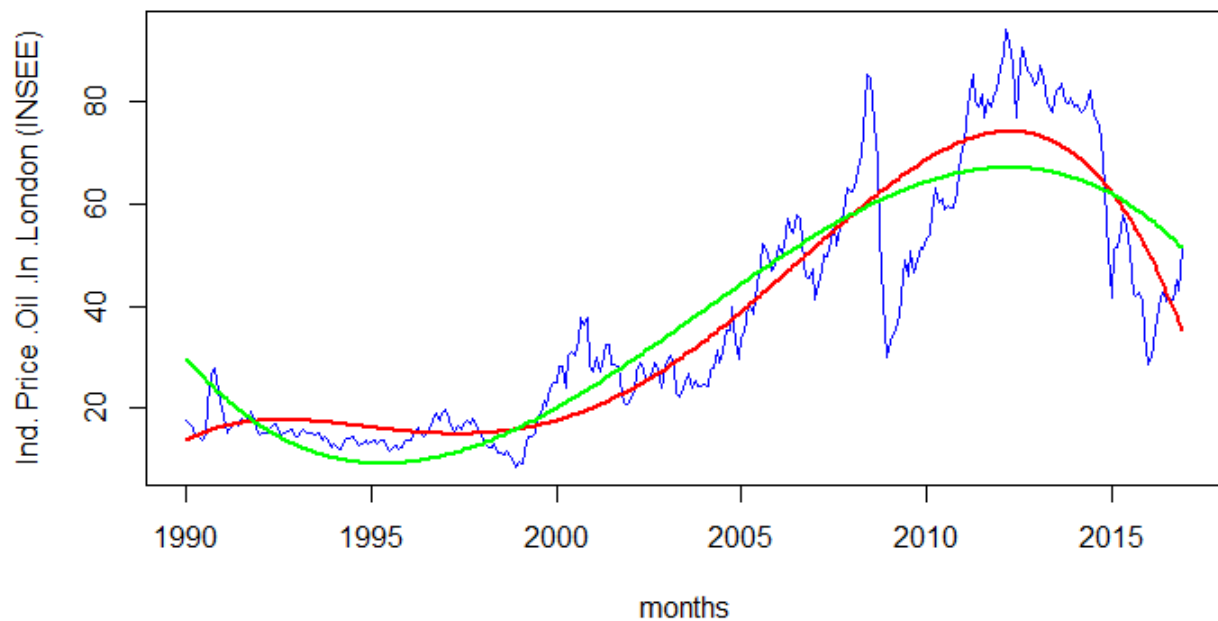
Différenciation :

Nous allons ici utiliser une technique locale. Il s'agit de différencier le point N avec le point N-1. Cela nous permet d'étudier la tendance polynomiale de la tendance de la série. On observe les graphes suivants :



Régression polynomiale :

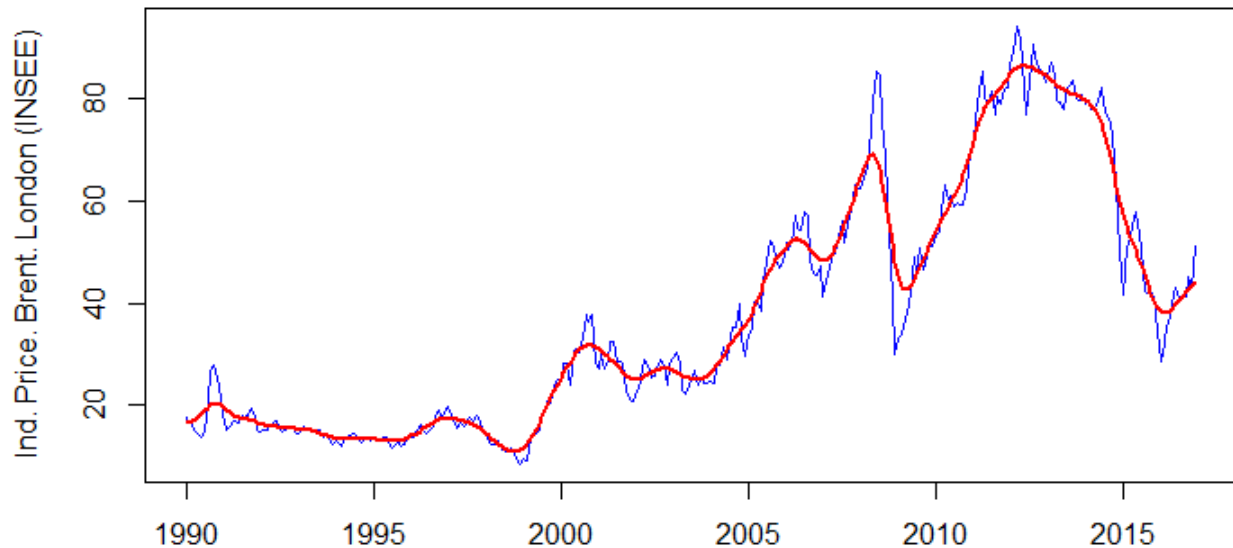
On peut ensuite modéliser la courbe de prix par une courbe polynomiale. En fonction du degré polynomial choisi, la courbe des prix sera plus ou moins bien corrélée à la modélisation. Sur le graphe ci-dessous, nous avons comparé les régressions polynomiales de degré 3 (en vert sur la courbe) et 4 (en rouge). La régression de degré 4 est plus pertinente car plus précise : elle rend mieux compte de la stabilité entre les années 1990 et 2000, puis de l'envolée des prix à partir de 2005. On aurait pu encore augmenter le degré de la régression : nous avons choisi la régression d'ordre 4 qui nous semblait satisfaisante (les polynômes d'ordre supérieur n'apportaient pas de précisions vraiment significatives et il faut aussi prendre en compte le coût des régressions d'ordre important).



Estimateur à noyaux :

L'estimation par noyaux estime la tendance en effectuant des régression fenêtre par fenêtre. Les données de l'estimation seront pondérées différemment selon leur place dans la fenêtre considérées.

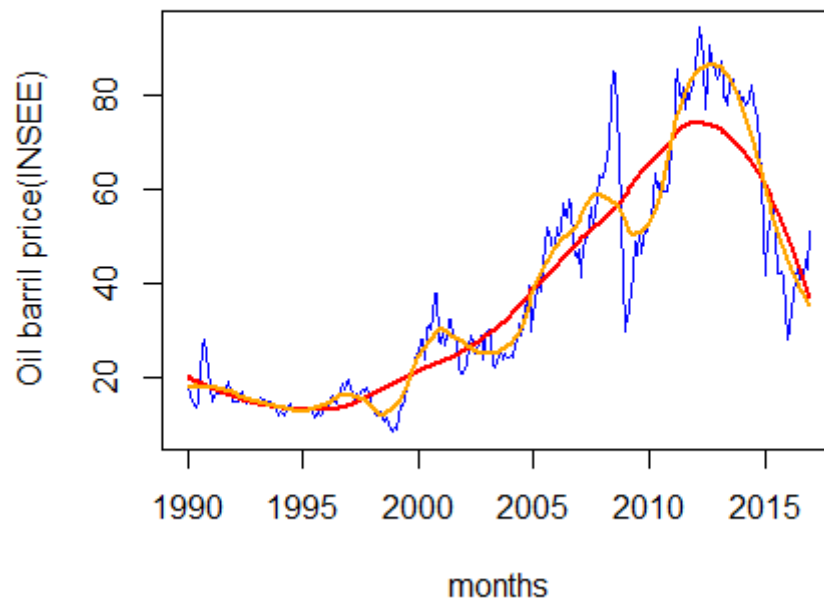
Pour un noyau gaussien de taille 10 mois on obtient l'estimation suivante :



Polynômes locaux :

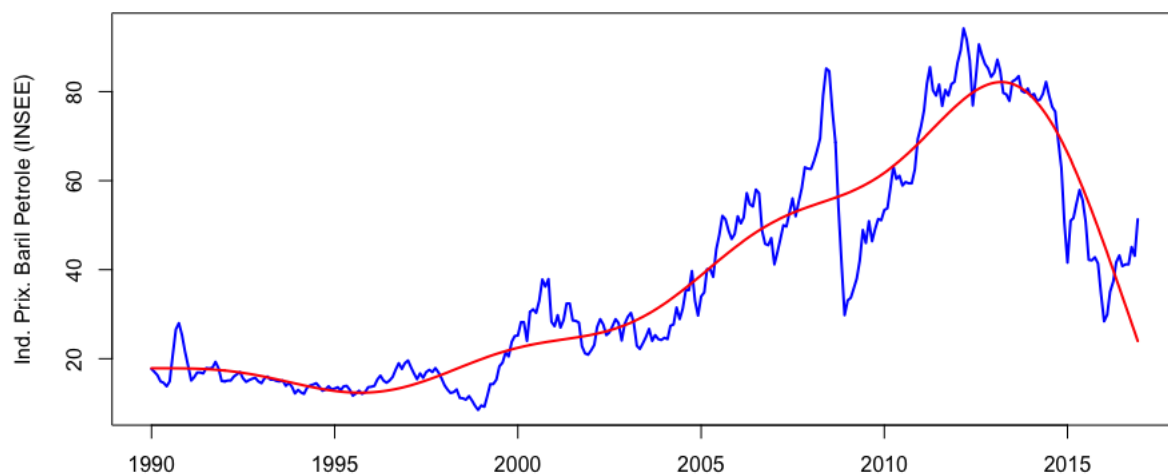
On peut maintenant affiner, on avait modélisé la courbe de prix par un polynôme. On va maintenant la modéliser par plusieurs polynômes : au lieu de modéliser l'ensemble des valeurs par un polynôme (coefficients uniques), on modélise portion par portion. On peut modifier la taille de la fenêtre considérée pour la modélisation. Pour la courbe du rouge, la taille de la fenêtre vaut la moitié de la taille de la série (ie : 162 valeurs sont prises en compte pour chaque fonction polynomiale sur les 324 qui apparaissent réellement dans la série). Pour la courbe orange, 20% des valeurs sont prises en compte. La courbe orange sera donc mieux adaptée à une étude locale, alors que la courbe rouge est plus adaptée à une étude globale.

Ici on a utilisé des polynômes d'ordre 2, on peut modifier ce degré et obtenir des précisions supplémentaires. (Ces précisions ne sont pas forcément utiles à l'étude considérée : dans notre cas et on gardera donc un degré deux).



Estimation semi paramétrique :

Il s'agit ici de projeter la série sur une base de fonctions polynomiales. Nous allons projeter jusqu'à l'ordre 3. Nous obtenons la tendance suivante :



Conclusion sur les différentes tendances :

Après avoir étudié la tendance par différentes méthodes, nous avons pu observer que certaines étaient plus précises que d'autres. Les méthodes locales sont notamment plus précises que les modèles globaux (polynômes locaux/ estimation semi paramétrique/ estimateur à noyaux \geq régression polynomiale). En fonction de la précision attendue et du coût de calcul de tendance, on pourra choisir l'une ou l'autre des méthodes.

Dans le cas du cours des prix du baril de pétrole, l'estimation qui nous semble la plus satisfaisante est celle des estimateur par noyaux. Elle est la seule capable de rendre compte du choc des années 2007-2008 de manière significative.

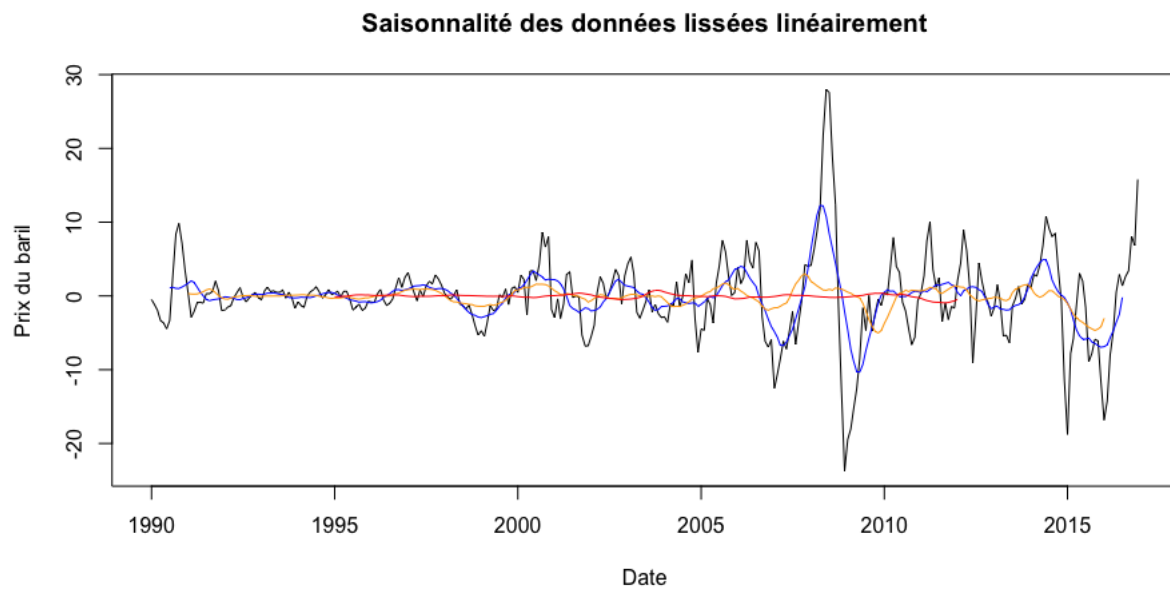
3.2 Étude la saisonnalité

Comme pour l'estimation de la tendance, nous avons plusieurs outils pour déterminer la saisonnalité. Nous utiliserons seulement la moyenne mobile.

Moyenne mobile :

Nous pouvons extraire la saisonnalité de la série avec une moyenne mobile. Pour déterminer la saionnalité, nous enlevons la tendance à la série.

Afin de déterminer la taille de la fenêtre, nous utilisons le fait que la demande de pétrole est toujours plus forte lors de la période hivernale de l'hémisphère nord. En effet, la Chine est un gros demandeur de pétrole ainsi que les pays occidentaux donc il y a une tendance à la hausse à cette période là de l'année. En prenant une fenêtre de 12 mois pour avoir une saisonnalité annuelle, nous obtenons la courbe bleue. Nous observons que la courbe bleue à un mouvement récurrent tous les 2 ans. Nous filtrons de nouveaux cette saisonnalité afin d'obtenir la courbe orange. Enfin, la courbe orange à une saisonnalité de 5 ans, nous faisons un dernier filtrage afin d'obtenir une courbe plate et lisse en rouge. La saisonnalité est donc de 120 mois soit 10 ans.



4 LISSAGE EXPONENTIEL ET PRÉVISIONS

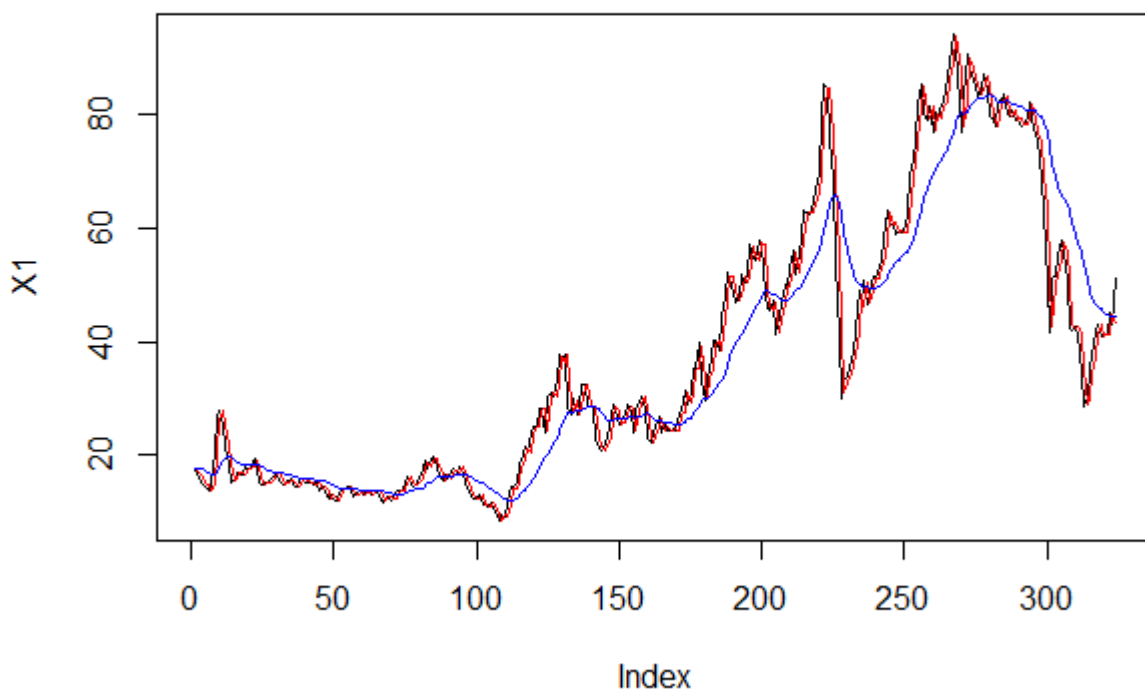
On va lisser exponentiellement la série de données. Le principe du lissage est d'utiliser les valeurs précédentes pour prévoir la ou les valeurs suivantes. L'objectif est la prévision à court terme.

Les données sont pondérées en fonction de leur ancienneté et interviennent ainsi avec plus ou moins de poids dans la détermination de la valeur anticipée.

Lissage simple :

Le lissage simple permet d'effectuer une prévision à l'horizon un mois ($h=1$ pas de temps en général, ce qui correspond à un mois dans notre cas).

On utilise un paramètre α pour pondérer les données. Plus α est proche de 1, moins les données à long terme seront déterminantes dans le calcul de la valeur anticipée. Lorsque que les données d'une série dépendent fortement de leur horizon lointain (forte auto-corrélation à horizon lointain), un coefficient α proche de 0 sera au contraire très approprié.



Ici on a tracé sur un même graphe, la série de données et deux lissages exponentiels simples

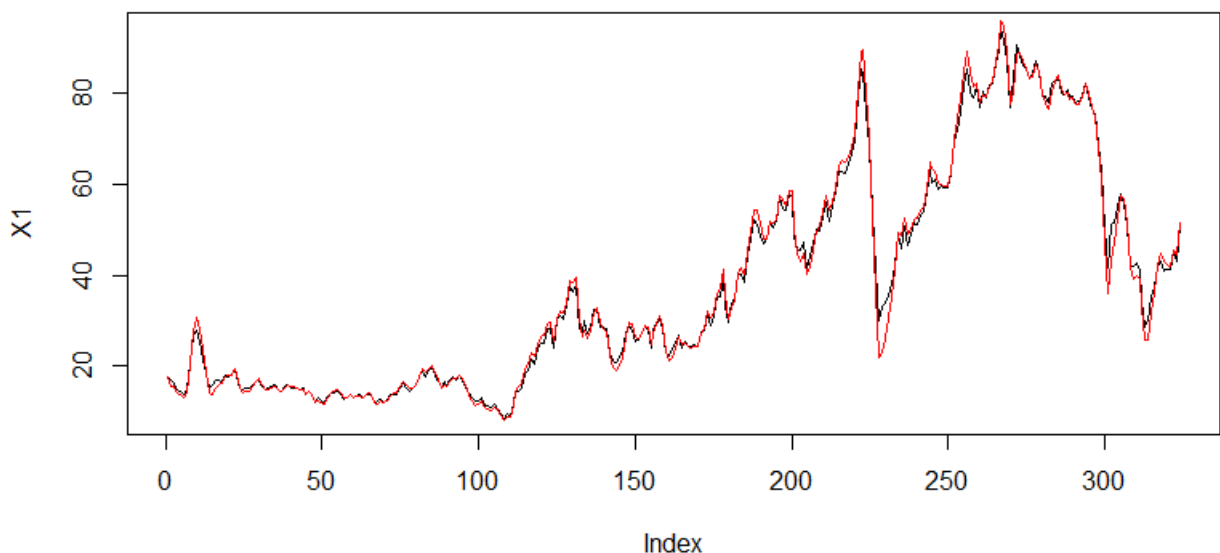
de paramètre $\alpha = 0.9$ (courbe rouge) et $\alpha = 0.1$ (courbe bleue). La courbe rouge est très sensible (et avec un retard quasi nul) aux variations spontanées de la série initiale. La courbe bleue au contraire, prend en compte un large spectre de données à long terme, elle est donc moins sensible à un choc (2007/2008) et réagit avec un retard important.

On a pu calculer le α optimal : il vaut 0.95. Cela signifie que les prix du pétrole peuvent être très variables en très peu de temps. Ce résultat est donc tout à fait cohérent avec la présence de chocs.

Lissage double (de Holt) :

Le lissage de Holt permet d'effectuer des prédictions à horizon $h \geq 2$ (avec h le pas de temps de la série) ; en effet plutôt que d'évaluer un point en fonction des précédents, il permet d'évaluer une droite en fonctions des termes précédents. Ce lissage permet non seulement d'anticiper des valeurs à plus long terme mais il est également plus pertinent sur son anticipation que le lissage simple.

Ci dessous, le lissage de Holt à horizon $h=1$ mois (superposé à la courbe des prix) :



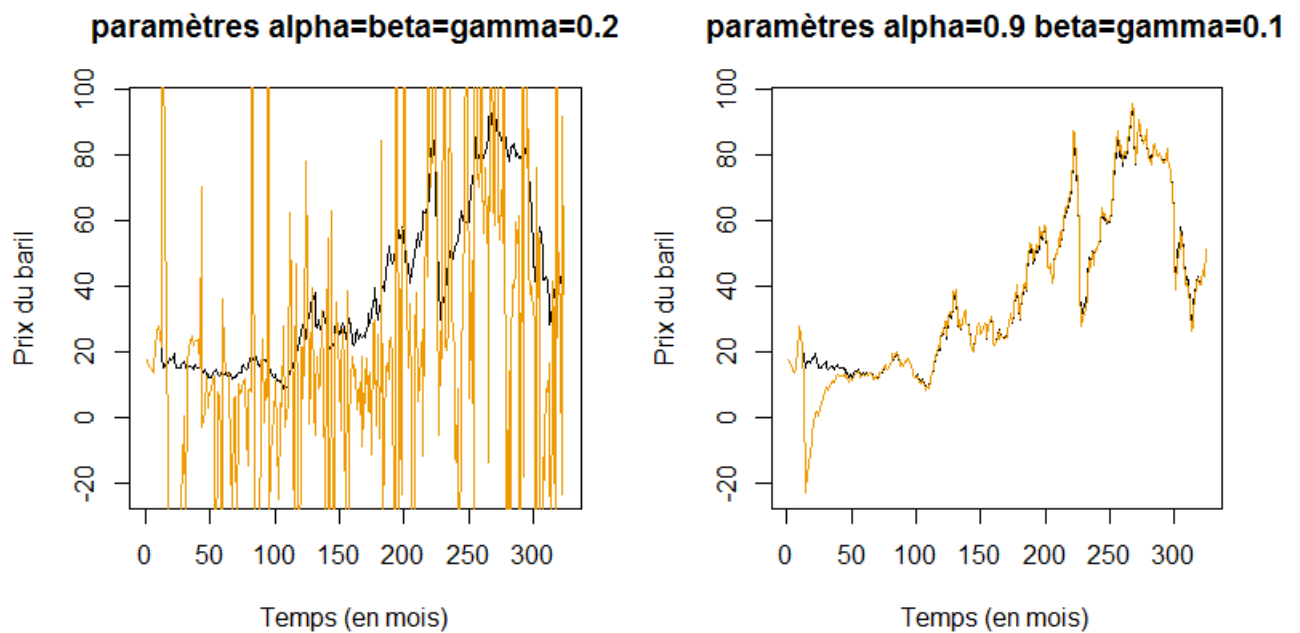
Lissage de Holt-Winters :

Le lissage de Holt-Winter est encore plus exacte : il permet de prendre en compte la saisonnalité (modèle additif ou multiplicatif). Nous avons utilisé deux approches : la première consiste

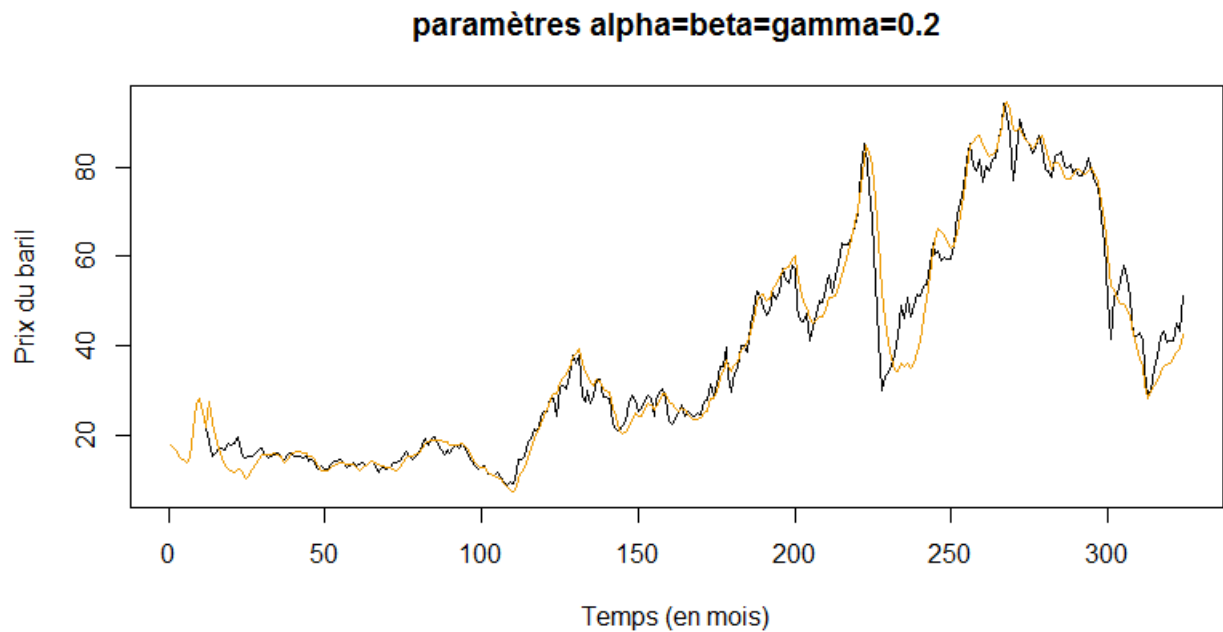
à coder directement des méthodes et la seconde consiste à utiliser la fonction Holt-Winter du logiciel R.

Codage direct : nous avons d'abord testé directement les modèles de saisonnalité multiplicative et additive. Après évaluation des résultats nous avons opté pour le modèle de saisonnalité additive. En effet comme on le voit sur les graphiques suivants, le lissage exponentiel est soit de très mauvaise qualité, soit très proche de la courbe originale (la saisonnalité n'est pas supprimée). Au contraire un modèle de saisonnalité additive permet bien de supprimer la saisonnalité annuelle, tout en restant proche des fluctuations tendancielle de la courbe des prix.

Lissage exponentiel avec modèle de saisonnalité multiplicative pour différents paramètres :

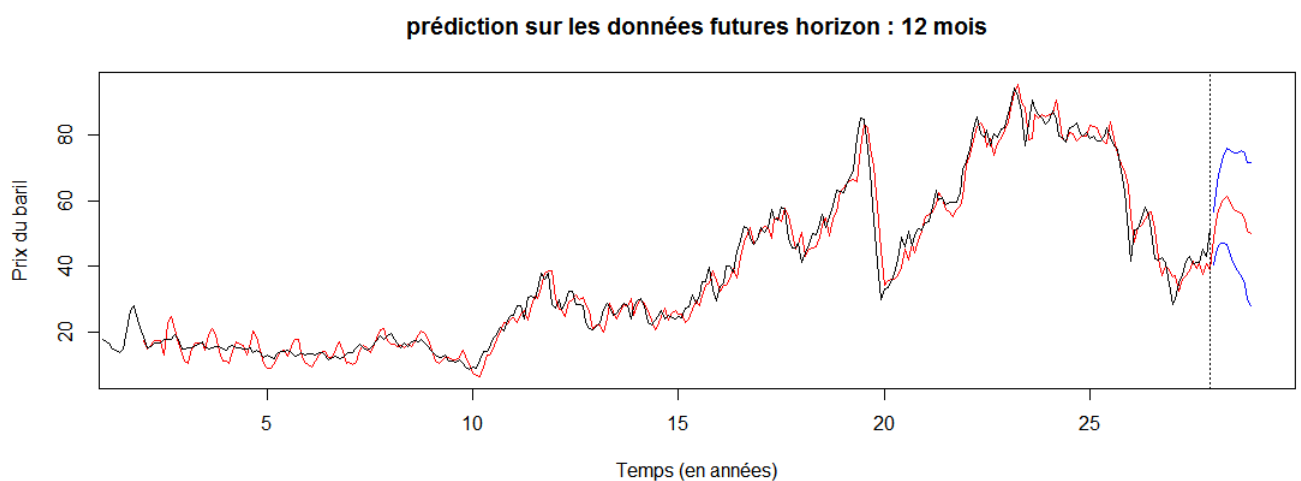


Lissage exponentiel avec modèle de saisonnalité additive :



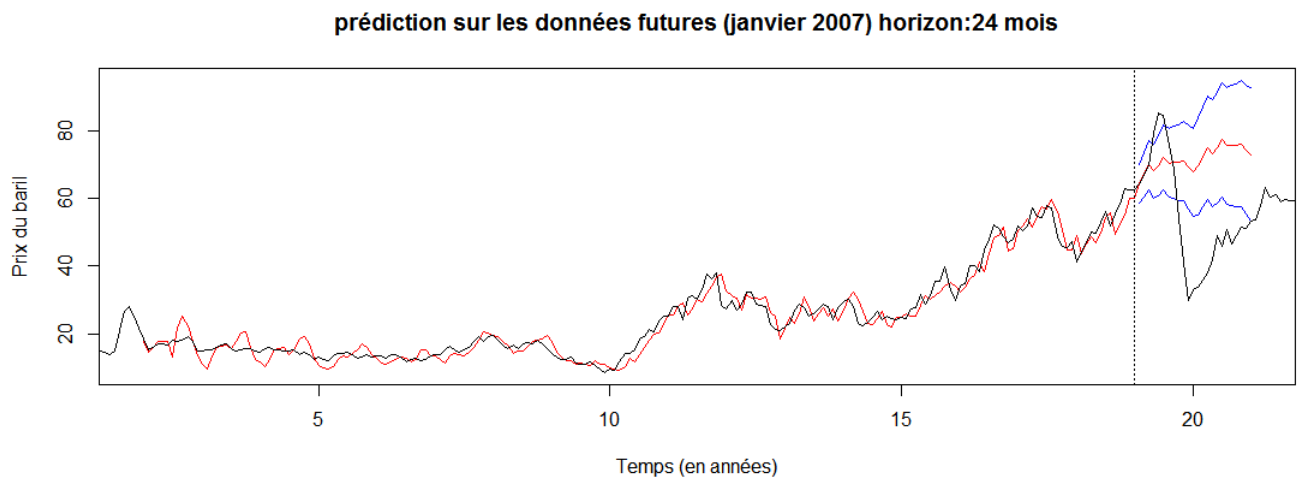
Par la fonction R : Holt-Winter : en appliquant la fonction Holt-Winter on peut obtenir une prévision sur les données futures.

Sur le graphe suivant on a représenté (en rouge) le lissage puis la prévision sur 12 mois, et en bleu les bornes de sûreté à 95%. On remarque que l'intervalle de sûreté s'élargit avec l'horizon (cela est encore plus visible sur le graphique de la prévision à horizon 24 mois)



Il convient cependant d'être prudent : il s'agit de prévisions dépendant uniquement des données précédentes. Ces données mathématiques ne prévoient en aucun cas la conjoncture

économique (dans notre cas) ni les éventuels phénomènes uniques (qui ne se sont jamais produits auparavant). Un statisticien qui aurait effectué une prévision en janvier 2007 à horizon 2 ans et avec un niveau de sûreté 95% aurait obtenu les résultats suivants.



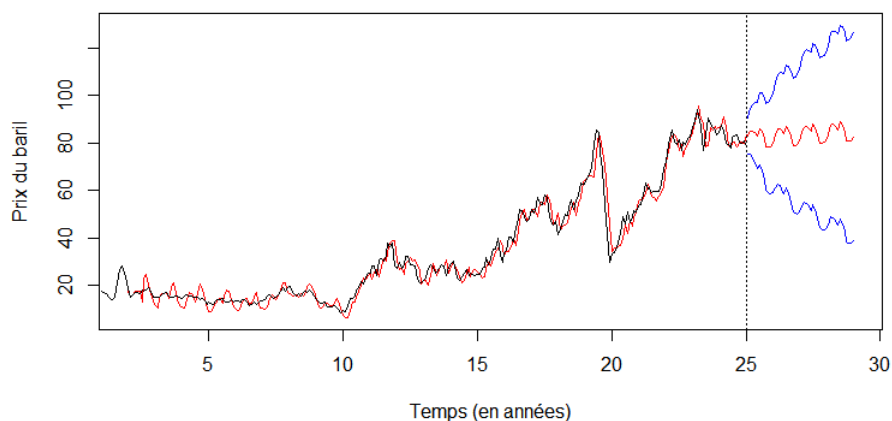
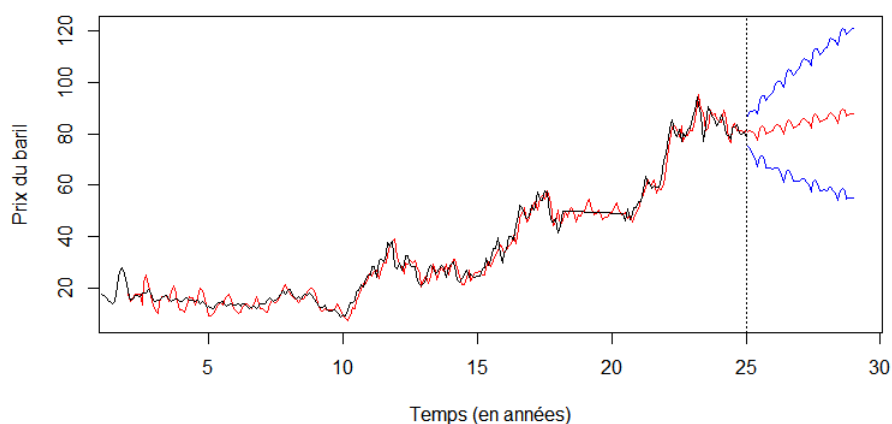
La courbe réelle des prix n'est pas comprise dans l'intervalle de confiance. Le test ne peut pas prévoir un choc ou un changement brutal de comportement dans le cas où celui-ci ne dépend pas (ou ne dépend pas uniquement) des données précédentes.

Traitement des données préalables aux prévisions :

Si on tente d'effectuer un lissage à horizon lointain, l'intervalle de sûreté à 95% est très important. Or on désire faire la prédiction la plus précise possible et donc réduire la taille de l'intervalle de sûreté.

Une méthode est de lisser les chocs (celui lié à la crise de 2007/2008 en particulier). Pour la prévision, on ne va travailler que sur les données entre 1990 et 2013 (afin de ne pas prendre en compte le second choc qui est difficilement lissable car les prix ne sont pas encore réstabilisés).

On effectue donc un lissage linéaire du choc de 2007. Nous avons étudié les séries obtenues intégralement (saisonnalité, régression polynomiale, estimation à noyaux...) mais nous ne présenterons ici que les résultats obtenus pour la prévision à 48 mois. (L'étude intégrale de la série lissée est disponible sur le fichier `cours_petrol_linear_smooth.R`).

prédiction sur les données futures (données brutess) horizon : 48 mois**prédiction sur les données futures (données lissées) horizon : 48 mois**

Si les bornes hautes des intervalles de confiance sont semblables pour les deux graphiques, la borne basse de l'intervalle est plus élevée pour les données lissées : la taille cet intervalle est donc inférieure à celui obtenu avec des données brutes. Le fait que que les bornes hautes des intervalles soient équivalentes n'est pas incohérent. En effet, malgré que le choc en lui même ait été lissé il reste une augmentation brutale du cours entre 2010 et 2013. On pourrait faire le choix de lisser également cette augmentation mais nous avons considéré que les données seraient alors trop modifiées pour que les prévisions soient réalistes.

5 CONCLUSION

L'étude des données statistiques temporelles du prix du baril de pétrole nous a permis de dégager plusieurs informations clefs. Elle a permis de dégager une tendance plus lisible que les données brutes et rendant comptes des chocs, augmentations et baisses des prix. Elle a également permis de confirmer plusieurs saisonnalités : annuelle mais aussi sur 2 et 5 ans. La connaissance de ces saisonnalités peut être utile aux économistes, acheteurs et vendeurs de pétrole qui désirent connaître le meilleur moment pour acheter ou vendre des barils par exemple. Si la saisonnalité annuelle peut être lue graphiquement, ce n'est pas le cas des deux autres saisonnalités. De plus il convient de confirmer toute supposition faite à l'œil nu par une étude mathématique.

Enfin, l'étude du lissage exponentiel nous a permis de prévoir statistiquement une évolution possible des cours futurs du prix du pétrole. Ces prévisions ne sont cependant pas à considérées comme certaines et ne peuvent pas anticiper les phénomènes complètement nouveaux (chocs). Les intervalles de confiances donnés sont aussi très importants : cela signifie que la prédiction est très imprécise.

Il est possible d'effectuer des lissages sur les chocs ou données qui rendent compte d'un phénomène unique pour améliorer la précision de la prédiction. Attention encore une fois : cette amélioration est fictive puisqu'elle résulte d'une série de données falsifiée.